Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

Article

Comparison of Lake Area Extraction Algorithms in Qinghai-Tibet Plateau Leveraging Google Earth Engine and Landsat-9 data

Xusheng Li [1], Donghui Zhang [2], Chenchen Jiang [3], Yingjun Zhao [1,*], Hu Li [4], Donghua Lu [1], Kai Qin [1], Donghua Chen [5], Yufeng Liu [5], Yu Sun [1] and Saisai Liu[5]

[1] National Key Laboratory of Remote Sensing Information and Imagery Analyzing Technology, Beijing Research Institute of Uranium Geology, Beijing 100029, China; saintlxs@foxmail.com (L.X.);zhaoyingjun@briug.cn (Z.Y.);ludonghua@briug.cn (L.D);qinkai@briug.cn (Q.K.);sunyu@briug.cn (S.Y.)
[2] National Engineering Laboratory for Satellite Remote Sensing Application (NELRS), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zhangdonghui@aircas.ac.cn (Z.D.)
[3] Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University, Beijing 100871, China; jiangchenchen@stu.pku.edu.cn (J.C.)
[4] School of Geography and Tourism, Anhui Normal University, Wuhu, Anhui 241000, China; lihu2881@ahnu.edu.cn (L.H.)
[5] School of Computer and Information Engineering, Chuzhou University, Chuzhou, Anhui 239000, China; chendonghua@chzu.edu.cn (C.D.); Liuyufeng@chzu.edu.cn (L.Y.);lss@chzu.edu.cn (L.S.)
* Correspondence: zhaoyingjun@briug.cn; Tel.: 010-64964719

Abstract: The monitoring of lake waterbody area in the Qinghai Tibet Plateau (QTP) is of great significance to deal with global climate change. As the latest generation of Landsat series satellites, Landsat-9 data not only have higher radiometric resolution, but also cooperate with other Landsat satellites to greatly improve the temporal resolution. It has great application potential in lake waterbody area monitoring. In order to explore the performance of different algorithms for extracting waterbody and lake waterbody area in Landsat-9 data under large-scale QTP regions, this study relies on Google Earth Engine (GEE) platform and selects 10 waterbody extraction algorithms as the basis to realize the quantitative evaluation of QTP lake waterbody area extraction results. The results show that the Random Forest (RF) algorithm performs best in all models. The overall accuracy of waterbody extraction is 95.84%, and the average error of lake waterbody area extraction is 1.505%. Among the traditional threshold segmentation waterbody extraction algorithms, the overall accuracy of the NDWI waterbody extraction method is 89.89%, and the average error of lake waterbody area extraction is 3.501%, which is the highest performance model in this kind of algorithms. This study proves that Landsat-9 data can effectively classify QTP waterbodies. With the development of cloud computing technologies such as Gee, more complex models such as RF can be selected to improve the extraction accuracy of water body and Lake area in large-scale research.

Keywords: Landsat-9 data; Qinghai-Tibet Plateau; Lake Waterbody; GEE; Algorithms comparison

1. Introduction

The Qinghai-Tibet Plateau (QTP), referred to as 'water tower of Asia', is the birthplace of many major rivers in Asia, such as the Yangtze River, the Yellow River. It is about 2.6 million km^2 in area and most of it lies at an evaluation of more than 4 km above sea level [1]. The unique alpine terrain of the QTP has blocked and raised the warm and humid South Asian monsoon [2], forming rich water resources reserves in the region. The QTP contains approximately 1400 lakes of area larger than 1 km^2, with a total area of about 50000 km^2, which is one of the most densely distributed areas of lakes in the world [3–5]. Studies have shown that the QTP is one of the most sensitive regions to global climate change, and the lakes, which are located in closed inner flow areas, are less affected by human activities and are outposts of the cryosphere and climate change [6,7]. Accurately

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

depicting the lake area of the whole QTP is one of the concerns of the Intergovernmental Panel on Climate Change (IPCC, https: //www.ipcc.ch/srocc/), which helps to enhance the understanding of climate change under the background of global warming through the temporal and spatial changes of lake water bodies in the plateau. It provides scientific support for the protection and development of ecological resources in the QTP lake area and the adoption of climate change countermeasures.

However, the traditional method of extracting the lake area by directly measuring water level and shoreline has a high cost, poor timeliness, and spatial accessibility. Due to the way monitoring data is acquired and stored varies from site to site, it is difficult to meet the requirements of current global change research on the integrity of hydrological monitoring data [8,9]. Because of its advantages of wide spatial coverage, strong cyclicality and low cost, remote sensing has become an important means of lake waterbody area extraction, and has also greatly promoted the dynamic monitoring of lake area on the QTP. From the perspective of data, satellite remote sensing provides a large number of multi-source data with long-term series and high-spatial resolution. The data sources commonly used in the research mainly include MODIS [10,11], Landsat [12–14], SPOT [15,16], ALOSE [17], ASTER [18,19], HJ-1A/1B [20,21], WorldView [22,23], QuickBird [24], IKONOS [25], GaoFen [26,27], ZY-3 [28,29], SAR [30] and hyperspectral data [31,32]. In addition, optical remote sensing imagery is the primary source of data for the extraction of water features, in particular, Landsat-9 data with high-spatial resolution and long-time series, which is freely available from the U.S. Geological Survey (USGS, http://glovis.usgs.gov/) and is the first choice data of most researches [33]. Landsat-9 data, which is accessible by the USGS on February 10, 2022, is basically identical in band setting to the Landsat 8 sensors, and include higher radiometric resolution (14-bit quantization increased from 12-bits for Landsat 8), which means that Landsat-9 is more theoretically sensitive and more probing. Since Landsat-9 is matched with Landsat 8 in track design, if the two data are combined for analysis, the temporal resolution could be improved from 16 days to 8 days [34]. However, at present, there is no comprehensive research on water area extraction in QTP based on landsat-9 data. The research on water extraction based on landsat-9 data is of great significance to the protection and development of ecological resources in QTP lake area.

At present, many scholars have proposed various lake water feature extraction models for different application scenarios with multiple data, which can be mainly divided into the single-band threshold method, spectral relation method, water index method, machine learning and multiple model assisted synthesis method [35]. The single-band threshold method mainly uses the difference in reflectance between waterbodies and other ground objects in the infrared band, and selects water features with thresholds, which is widely used when early remote sensing data is scarce. For example, Frazier et al. [36] used the threshold method to extract the waterbody of Wagga Lake with the TM4, TM5 and TM7 bands, and the results showed that the extraction method with TM5 band was better. The spectral relation method mainly extracts the waterbodies by analyzing spectral curves of the characteristic band on multiple ground objects and constructing logical classification rules. For example, Du et al. [37] and Yang et al. [38] extracted the waterbody information by building a logical relation based on the characteristics of the gray value of the waterbody, and the difference in reflectivity degradation between the waterbody and shadow in the blue-green band, respectively. The spectral relation method can better distinguish between waterbodies and mountain shadows, but the anti-noise interference ability is poor, and it is susceptible to the influence of non-water noise such as vegetation and buildings, which is the same as the single-band threshold method. The water index method uses the bands with water reflection difference to construct the ratio operation, which suppresses the vegetation information, weakens the influence of soil, buildings and shadows, and highlights the water information. Since Mcfeeters [39] proposed NDWI (Normalized Difference Water Index), many scholars have proposed a variety of improved water index according to different background features and water

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

characteristics, such as MNDWI (Modified Normalized Difference Water Index) [40], EWI (Enhanced Water Index) [41], AWEI (Automated Water Extraction Index) [13], SWI (Shadow Water Index) [42] and LSWI (Land Surface Water Index) [43] and have driven the rapid development of water information extraction research [44–46]. Although the water index method are simple to operate and can eliminate shadows such as mountains, buildings, and vegetation, recognition errors of small water bodies and the boundaries between water and land are large. The machine learning method regard water bodies as a category, using specific classification rules to classify them to obtain water features. Machine learning methods such as Minimum Distance (MD) [47], Decision Tree (DT) [48], Support Vector Machine (SVM) [49], Neural Network (NN) [50], Random Forest (RF) [51] and Deep Learning (DL) [52] are widely used. Machine learning algorithms realize the effective use of spatial and texture information of high-spatial resolution images, and has a good effect on the extraction of small water bodies, but the process is relatively complex. For example, Sui et al. [53] integrated three modules of embryo water extraction, iterative segmentation and change detection with the help of GIS technology to realize the automation of water extraction process; and Qiao et al. [54] proposed an adaptive extraction method of "whole-local" progressive spatial scale transformation based on NDWI, and combined with spectral feature fitting method and iterative algorithm to accurately extract the lake range. Multiple model assisted synthesis method is to synthesize a variety of models and methods to solve the problem of water extraction in large-scale and complex background scenes, but the process is complex and the generalization ability of the model is poor. By expounding and analyzing the principles, advantages and disadvantages of different water extraction methods, it is not difficult to see that the water extraction algorithm has no absolute advantages and disadvantages, and the consistency with the data source and applications will also affect the accuracy of the algorithm. Previous studies have shown that the traditional NDWI method has a good performance in water extraction of the QTP region [55], but the relevant conclusions are not clear for Landsat-9 imagery under the background of large-scale plateau region. Therefore, the comparative study of large-scale QTP lake extraction algorithms based on Landsat-9 images has certain practical significance, and can provide a useful reference for the future researches.

To sum up, the research on remote sensing extraction of spatial distribution information of lakes in the QTP has important scientific significance for coping with global climate change. Although water extraction algorithms based on different resolutions and types of remote sensing data combined with different principles have been widely used in water information extraction, there has not been a comparative study of Plateau Lake area extraction algorithms based on landsat-9 images. Therefore, this paper uses Google Earth Engine （https://developers.google.cn/earth-engine ） [57, 58, 59], which has significant advantages in high spatial-temporal resolution and large-scale water extraction, to carry out a comparative study of landsat-9 lake area extraction in combination with 10 widely used water extraction algorithms (including a single band threshold extraction algorithm, two spectral correlation algorithms, four water index algorithms and two machine learning algorithms). Through comparative research, quantitatively evaluate the consistency between different algorithms and landsat-9 data in the plateau lake extraction scene, so as to determine the algorithm for large-scale plateau lake area extraction suitable for landsat-9, and provide some reference and suggestions for further research in subsequent related fields.

2. Materials

2.1. Study region

The QTP region as shown in Figure 1, is our study area, which locates between 67°40′37″E~104°40′57″E and 25°59′30″N~40°1′0″N. The total area of the region is over 3million km2, with an average altitude of about 4320 m, and it spans nine countries including

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

China, India, Pakistan and Afghanistan[62]. This region is an important water resource reserve area in China. The annual outbound water volume of the rivers in Southwest China, which mainly originate here, accounts for about 95% of China's total annual water consumption (2020). In addition, there are many lakes in the region, with a total area of about 50000 km2, accounting for more than half of the lake area in China [3,5].

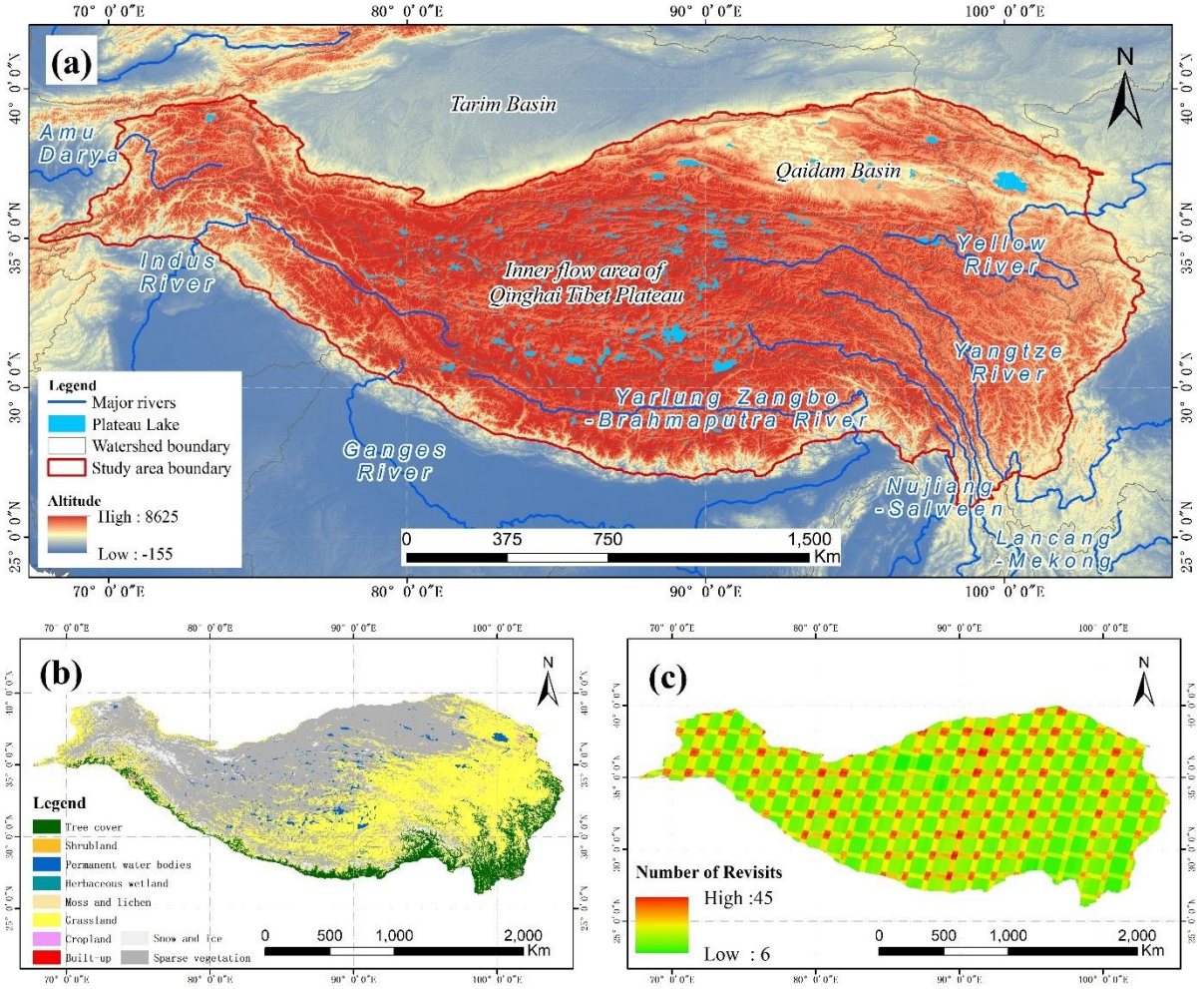


Figure 1. Study region and dataset. (a) Locations of the study area and the Terrain characterize. Study area boundary data were mainly based on TPBoundary_2021. The major river data were collected from HydroRIVERS. The lake data were based on TPLA_V3. The topographic data were produced by NASA-provided SRTM DEM [59]; (b) The Landcover data source was ESA_WorldCover; (c) Landsat-9 images coverage in the study area were based on L9C2_SR.

2.2. Dataset

Landsat-9 satellite was successfully launched on September 27, 2021, and Landsat 9 data was publicly available on February 10, 2022. Landsat-9 satellite carries the Opera-tional Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). The radiometric resolution of the sensor is improved to 14-bit quantization. Landsat-9 orbit has time interval with Landsat-8 and Sentinel-2 orbit, so multi-source data analysis can be carried out to improve time resolution [34]. This study uses the Landsat 9 Collection 2 Surface Reflectance (L9C2_SR) which has been geometrical and radiometric corrected by USGS and downloaded on the GEE platform. L9C2_SR provides data for eight spectral bands with a ground sampling distance (GSD) of 30-meters(m), includes ultra-blue (0.435-0.451 μm)、blue (0.452-0.512 μm)、green (0.533-0.590 μm)、red (0.636-0.673 μm), near infrared (0.851-0.879 μm)、shortwave infrared 1 (1.566-1.651 μm), shortwave infrared 2 (2.107-2.294 μm)and surface temperature (10.60-11.19 μm) bands. In this study, 1211 Landsat-9

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

images with a time span from January 1, 2022 to April 30, 2022 were mosaic to obtain high-quality images covering the whole study area. The images coverage are shown in Figure 1 (c). The whole study area has image coverage, with a minimum of 6 times and a maximum of 45 times.

In addition, the following datasets were also used in this study: the QTP lake dataset (V3.0) (TPLA_V3, http://data.tpdc.ac.cn) [60,61], QTP boundary data (TPBoundary_2021, http://www.geodoi.ac.cn) [62], WorldCover data with 10-m spatial resolution published by European Space Agency (European Space Agency WorldCover 10 m 2020 product, ESA_WorldCover, https://zenodo.org/record/5571936), Global River Water data released by WWF Conservation Science Program and USGE in conjunction with a number of scientific institutions (vectorized line network of rivers, HydroRIVERS) [63]. Among them, TPLA_V3 data was used as the ground-truth value to generate lake samples for threshold determination and accuracy validation; TPBoundary_2021 data was used to determine the research regions; ESA_WorldCove was used to generate samples for machine learning algorithm classification; HydroRIVERS combined ESA_WorldCover to construct non-lake waterbodies for post-classification processing.

3. Methods

The step-wise progression of waterbody area extraction is illustrated in Fig. 2 and encompasses four steps: data collection and data processing, classification, and evaluation.

Step 1: Landsat-9 images, the lake data, the river data and landcover data were collected as input data.

Step 2: The Landsat-9 images were processed by removing clouds and shadows, and mosaicked into a de-cloud map. The remaining data were used to generate lake samples, landcover samples and non-lake water regions, and finally formed the data pre-processing results.

Step 3: The model parameters and the thresholds were adjusted according to the lake and landcover samples. Based on specific model parameters and thresholds combined with multiple models, waterbody extraction in the QTP study area was realized. Further, the waterbody extraction results were excluded by using non-lake waterbody data, and the lake waterbody extraction results in the study area were obtained.

Step 4: Finally, we evaluated the performance of algorithms with accuracy metrics based on the extraction information derived from step 3.

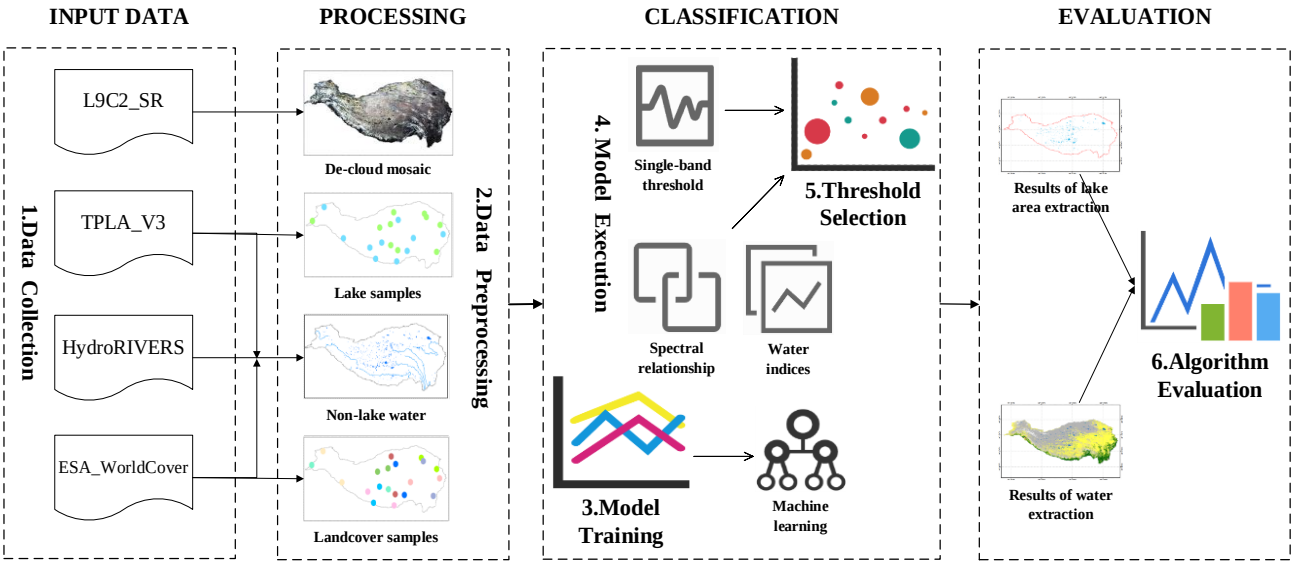


Figure 2. The general workflow of the experiment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

3.1. Data preprocessing

Data preprocessing includes three parts: image de-cloud mosaic, classification sample construction, and non-lake water area extraction. In order to obtain high-quality cloud-free images covering the study area, cloud removal becomes the significant preprocessing step and has a great impact on waterbody extraction results. This study used the "CLOUD_COVER" attribute and the "QA_PIXEL" band included in the L9C2_SR to remove clouds and cloud shadows [64]. Due to Landsat-9 level 2 products (L9C2_SR) have limited images in the time interval from January 1, 2022 to April 30, 2022, and the image cloud coverage in January and February is large. Therefore, by directly setting the attributes of "CLOUD_COVER", "QA_PIXEL" or "IMAGE_QUALITY_OLI", simply using functions such as median and mean to mosaic cloud-free images, it will lead to poor cloud shadow removal results or the mosaic image cannot cover the study area. Through the analysis of dataset images and many comparative experiments, we found that after removing clouds and cloud shadows by using "QA_PIXEL" band, without "CLOUD_COVER" attribute setting and we can get mosaic images with good cloud removal effect and complete coverage of the study area through the following piecewise mosaic steps: adopted the minimum value in the area with revisit times less than 12; Taken the quartile in the area where the number of revisits is greater than or equal to 12 and less than 30; The median value of the area with the number of revisits greater than or equal to 26 is taken for image mosaic, and the results are shown in Figure 3.

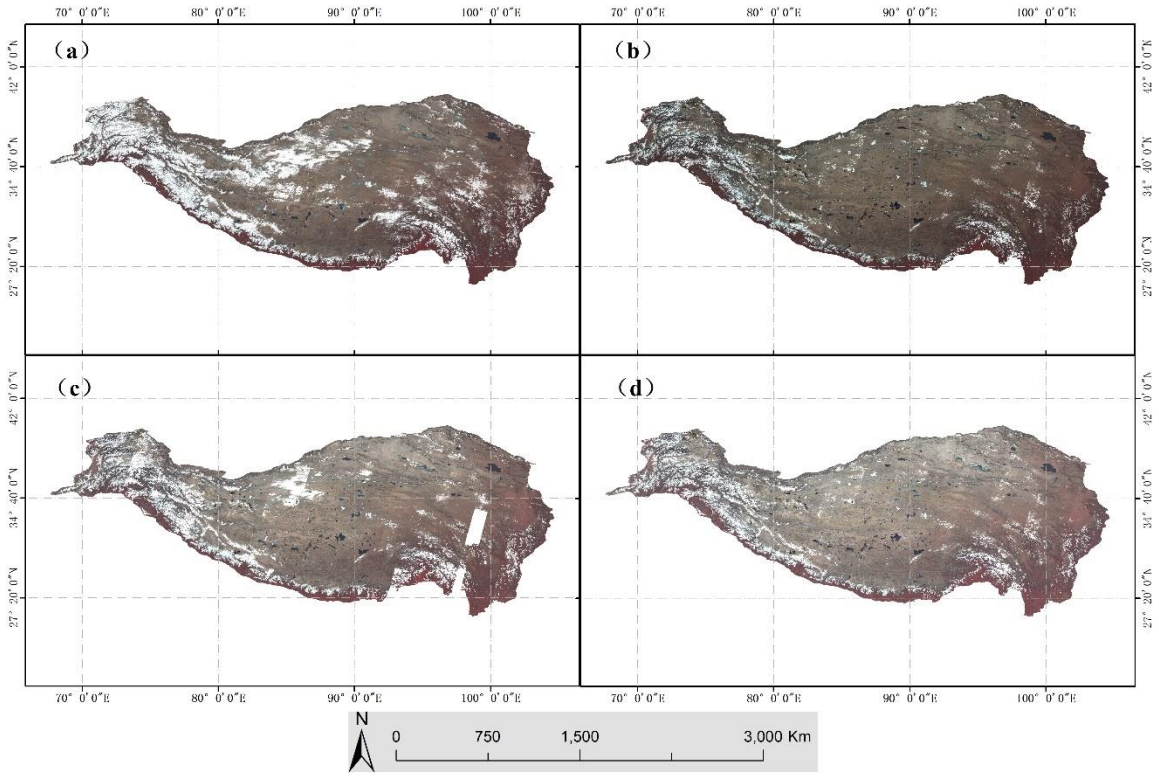


Figure 3. Mosaic images under removing cloud were synthesized in false color in bands 5, 4 and 3, and stretched with 2.5 standard deviations. (a) The cloud removal mosaic image with QA_PIXEL band, IMAGE_QUALITY_OLI wat set to 9, and median function. But there are still a lot of clouds in the northwest and southeast regions. (b) The cloud removal mosaic image with QA_PIXEL band, IMAGE_QUALITY_OLI wat set to 9, and min function. The result shows that the cloud removal effect is better, but the shadow effect is enhanced. (c) he cloud removal mosaic image with QA_PIXEL band, CLOUD_COVER wat set to 20, and min function. Mosaic result cannot cover the study area and there are still some clouds obscuring the northwest. (d) The cloud removal mosaic image with QA_PIXEL band and piecewise mosaic. The results show that the cloud removal effect is relatively obvious, the shadow enhancement effect is relatively small, and the coverage of the study area is complete.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

Referring to the algorithm proposed by Deng et al. [56], the ESA_WorldCover (global land cover dataset) and the TPLA_V3 QTP (lake dataset) were used for hierarchical automatic random sampling [57]. The details of sample construction are shown in Table 2. ESA_ Worldcover divides the land cover into 11 categories. After removing the mangroves that do not exist in the Qinghai Tibet Plateau, combined with TPLA_ V3 Lake data, generate stratified random samples. According to Landsat-9 images of Google Earth cloud removal and Sentinel-2 images of the same period with higher spatial resolution, the generated random samples were manually revised to eliminate wrong samples. The samples were randomly divided into test set, training set and validation set according to the ratio of 3: 4: 3. The test set was used for the search and optimization of model hyper-parameters, the training set was used for model training, and the validation set was used for model accuracy verification. In contrast, for non-machine learning models, the test set and training set were used to determine the threshold, and the validation set was used for accuracy evaluation.

Table 1. Landcover samples.

Landcover class	Landcover ID	Number	Source
Tree cover	1	11127	ESA_WorldCover
Shrubland	2	4345	ESA_WorldCover
Grassland	3	13210	ESA_WorldCover
Cropland	4	7376	ESA_WorldCover
Build-up	5	1767	ESA_WorldCover
Bare/sparse vegetation	6	15972	ESA_WorldCover
Snow and Ice	7	8409	ESA_WorldCover
Water bodies	8	41654	TPLA_V3 & ESA_WorldCover
Herbaceous	9	2437	ESA_WorldCover
Moss and lichen	10	10726	ESA_WorldCover

Previous study combined the reservoir, dam database, river database and other non-lake waterbodies to extract the lake waterbody area based on the extraction results of waterbodies [65]. Therefore, this study combined with ESA_WorldCover、TPLA_V3 and HydroRIVERS data to establish a non-lake waterbody dataset for post-processing of waterbody extraction. Non-lake permanent waterbodies were obtained by a geometric union of waterbodies between TPLA_V3 and ESA_WorldCover after erasing the intersection waterbodies. The buffer regions were built according to the inter-annual variations of the lake [55], and the buffer regions were used to erase the HydroRIVERS data to obtain the river dataset. Finally, we formed the non-lake waterbody dataset for post-processing.

3.2. Waterbody extraction algorithms

At present, waterbody extraction algorithms such as the single-band threshold method, spectral relation method, water index method, and machine learning are widely used, especially the water index method and machine learning [48]. This study selects four water index models, two machine learning models, two spectral relationship models, and a single-band threshold model to explore the extraction results of waterbody and lake waterbody in the QTP based on Landsat-9 data and to determine the best performance of waterbody area extraction algorithms under the background of the large-scale study region. The detailed information of each model is shown in Table 2. Where ρ_i represents the band i, which refers to section 2.2. N_1 and N_2 are the experience thresholds. X, C, $gamma$, nt, mf represent the feature sets, regularization parameter (known also as penalty factor), kernel width, the number of decision trees, and the number of input features used to split the nodes, respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

Table 2. Details of waterbody extraction algorithms.

Model type	Name	Formulas/Parameters	References
Single-band threshold method	SBT	$\rho_{SWIR1} < N_1$	Frazier & Page [36]
Spectral relationship method	SR	$\rho_{\text{green}} + \rho_{red} > \rho_{NIR} + \rho_{SWIR1}$	Du & Zhou [37]
	mSR	$(\rho_{\text{green}} + \rho_{red}) - (\rho_{NIR} + \rho_{SWIR1}) - (\rho_{\text{blue}} - \rho_{green}) > N_2$	Yang, et al. [38]
Water indices model	NDWI	$NDWI = (\rho_{\text{green}} - \rho_{NIR}) / (\rho_{\text{green}} + \rho_{NIR})$	Mcfeeters [39]
	mNDWI	$mNDWI = (\rho_{\text{green}} - \rho_{SWIR1}) / (\rho_{\text{green}} + \rho_{SWIR1})$	Xu [40]
	AWEI	$AWEI_{nsh} = 4*(\rho_{\text{green}} - \rho_{SWIR1}) - (0.25*\rho_{NIR} + 2.75*\rho_{SWIR2})$ $AWEI_{sh} = \rho_{blue} + 2.5*\rho_{\text{green}} - 1.5*(\rho_{NIR} + \rho_{SWIR1}) - 0.25*\rho_{SWIR2}$	Feyisa , et al. [13]
	mAWEI	$mAWEI = 5*(\rho_{\text{green}} - \rho_{NIR}) + (\rho_{blue} + \rho_{red} - 4*\rho_{SWIR2})$	Nie, et al. [66]
Machine learning model	SVM	X , C , $gamma$	Razaque, et al. [49]
	RF	X , nt , mf	Ko, et al. [51]

3.3. Model parameters and thershoholds

Previous studies have shown that the classification accuracy of machine learning models is largely dependent on model hyper-parameters [67]. In order to effectively adjust the hyper-parameters and optimize the classification accuracy of the model, drawing on the ideas proposed by Porwal et al. [68], this study used the sklearn package which is the open-source machine learning toolkit, and the RandomizedSearchCV to search for an optimal solution in large-scale range, and then used GridSearchCV to search for a certain floating fine-tuning the hyper-parameters within the small-scale range. Further, the 5-fold cross-validation method was used to verify the classification performance of the test set on the model, and the average accuracy was regarded as the estimated value to fine-tuning the parameter optimization.

In Table 2 of the section 3.2, SVM model uses Radial Basis Function (RBF) kernel, and needs to search and optimize the parameters of C and $gamma$. Random Forests (RF) model uses Classification And Regression Trees (CART) as the basic algorithm, and nt and md parameters need to be adjusted and optimized.

Appropriate thresholds are critical for water extraction models based on threshold segmentation. The selection of thresholds has a certain randomness, which varies with the type of index, the identification scenario, and the subpixel water/non-water components [69]. The OTSU method [70] and its improved algorithms [71,72] are commonly used for water index threshold automatic extraction, which use waterbodies and near-water land to present a bimodal distribution in the frequency domain of the index, so as to set the optimal threshold to split the image into "foreground" and "background" to achieve the classification of waterbodies and non-water bodies [73]. However, for the extraction of waterbodies and lake waterbodies in the QTP region, there are two problems with this threshold algorithm. First, there is a huge difference in the proportion of waterbodies and non-water bodies area in the region, which makes it impossible to effectively classify waterbodies in the frequency domain when they are covered by other landcover classes [71]. Second, the area of the study region exceeds 3 million km², and the use of such threshold segmentation methods requires extremely high computing power. Therefore, this study uses lake samples to manually adjust the threshold.

3.4. Evaluation metrics

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

In this study, the extraction accuracy of water and lakes is verified by the confusion matrix with verification set, and the performance of the above water extraction model is quantitatively evaluated through four evaluation metrics, including overall accuracy, Kappa, map-level accuracy and user's accuracy. Among them, the overall accuracy reflects the overall effect of the algorithm. Kappa indicates the degree of consistency between the ground-truth data and the predicted value. Map-level accuracy represents the probability that the validation sample is correctly classified. The user's accuracy represents the ratio of the inspection points falling on Category i on the classification diagram to be correctly classified as Category i [72]. In addition, the error analysis formula is introduced to evaluate the accuracy of area extraction.

$$\delta = \frac{|A_r - A_e|}{A_r} \tag{1}$$

Where δ is the error result. A_r represents the ground truth area and A_e is the area extracted by the algorithm.

4. Results and discussion

4.1. Parametesr and thresholds selection results

4.1.1. Optimization parameters in machine learning methods

For the machine learning methods, we evaluated the importance of the sample features involved in classification, reduced the redundancy between sample features, to improve the efficiency and accuracy of the algorithms. In the sample construction stage, we used the test dataset constructed in Section 3.1 to construct the waterbody and non-water body samples, and referred to [74] to evaluate the importance of 8 bands and 7 water indexes that participated in the calculation based on the normalized Gini index. According to the importance score, we successively put the features into the classifier for cross accuracy verification, so as to determine the optimal feature parameters for water extraction.

The evaluation and selection results of characteristic parameters are shown in Figure 4. Among the features represented by the horizontal axis of Figure 4, the importance score of features gradually decreases from left to right, and the features with the highest and lowest importance scores are SR_B5 and AEWIns, respectively. The low importance scores of AEWIs and AEWIns may be due to the large correlation between the indexes and other characteristics [74]. Further, according to the overall accuracy of cross validation, it can be seen that when the total number of features reaches 6, the overall accuracy reaches 93.80%, and the remaining features make less cumulative contribution to the improvement of accuracy. Based on the importance scores and overall accuracy results, the feature sets for water extraction methods (RF and SVM classifiers) are defined as {SR_B5, NDWI, SR_B7, mNDWI, mAEWI, SR_B6}.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

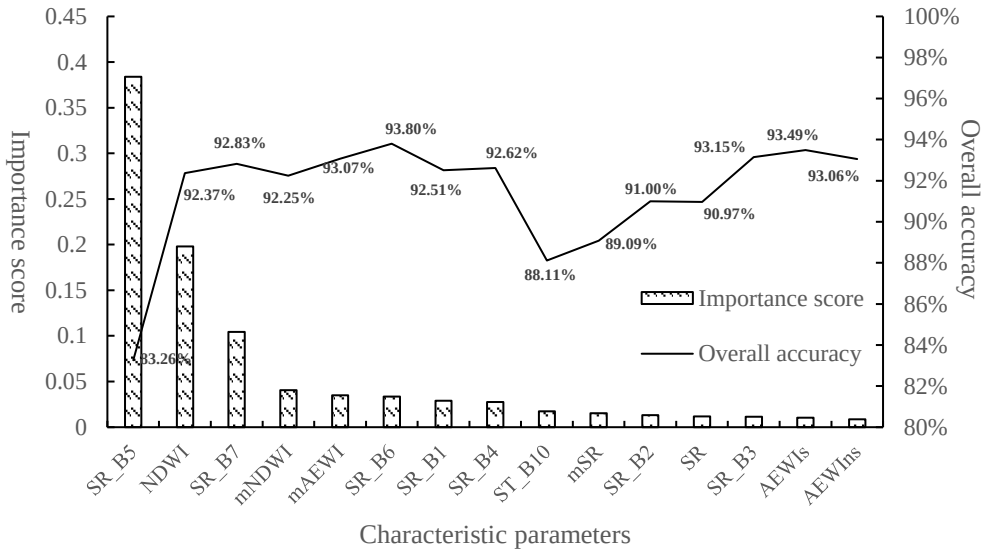


Figure 4. Evaluation and selection results of characteristic parameters.

In order to improve the accuracy of machine learning models, we optimized the key parameters of SVM and RF algorithms based on the above constructed feature sets. The key parameters of SVM algorithm include C regularization parameter, kernel function types and their related kernel parameters. Common kernels with SVM classifiers include polynomial kernel, Radial Basis Function (RBF) kernel, linear kernel, and more. Previous studies have shown that RBF kernel had better performance for image recognition with obvious non-linear characteristics [75]. Therefore, we selected the RBF kernel as the kernel function and optimized the parameters of C and $gamma$. C is the relaxation vector parameter in SVM classifier. When the C value is small, the interface is smooth. On the contrary, when it becomes large, the complexity of the model will increase. Kernel parameter $gamma$ defines the magnitude of the effect of a training sample, which is the reciprocal of the width of the RBF kernel. When the $gamma$ value is larger, the influence on the radius is smaller, and over fitting is easy to occur. On the contrary, under fitting is easy to occur [76]. The effective range of C value and $gamma$ value is 10^{-8}~10^{8}, but in practical application, the possible optimal values are generally in the range of (0.1,100) and (0.0001,10) [77]. Therefore, we used RandomizedSearchCV function to search for the optimal parameters of C and $gamma$ values in the interval of (0.1,100) and (0.0001, 10) with the multiple of 10 as the step size, and determine that the optimal values appear around 10~100 and 0.01~0.1, respectively. Then, we used the GridSearchCV function to realize grid search in steps of 10 and 0.01 between (10,100) and (0.01, 0.1). Figure 5(a) shows that the overall accuracy reaches the maximum value of 95.79% when $C = 40$ and $gamma = 0.04$. Therefore, the SVM classifier will be trained with these parameters in the experiment.

As an ensemble learning classifier, the random forest (RF) classifier is widely used in remote sensing information extraction. Randomness in RF is mainly reflected in the random selection of datasets and features used in each tree. The corresponding parameters are the number of decision trees (nt) and the maximum number of features (mf) to be selected for the node split when growing the trees [78]. When nt is larger, more decision trees are involved and the algorithm is more complex. The mf parameter allows each tree to be trained to use only some features at random, reducing the overall operation while also allowing each tree to focus on the features it chooses. The researches show that larger nt and mf will reduce the randomness and operation efficiency of the RF model and contribute less to the improvement of accuracy. Generally, the values of nt

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

and mf will not exceed 1000 and 50, respectively [78,79]. As mentioned above, two parameters need to be set in order to produce the forest trees in this study. According to optimization methods of the parameters in SVM, we determine that the optimal values of nt and mf appear around from 80 to 100, and from 5 to 15, respectively. We performed grid search on the parameters in steps of 2 and 1, respectively. Figure 5(b) shows that the overall accuracy reaches the maximum value of 94.97% when $nt = 90$ and $mf = 6$ or $nt = 100$ and $mf = 10$. Considering that larger nt and mf will increase the complexity of the RF model and reduce the generalization ability of the model, $nt = 90$ and $mf = 6$ are selected as RF parameters.

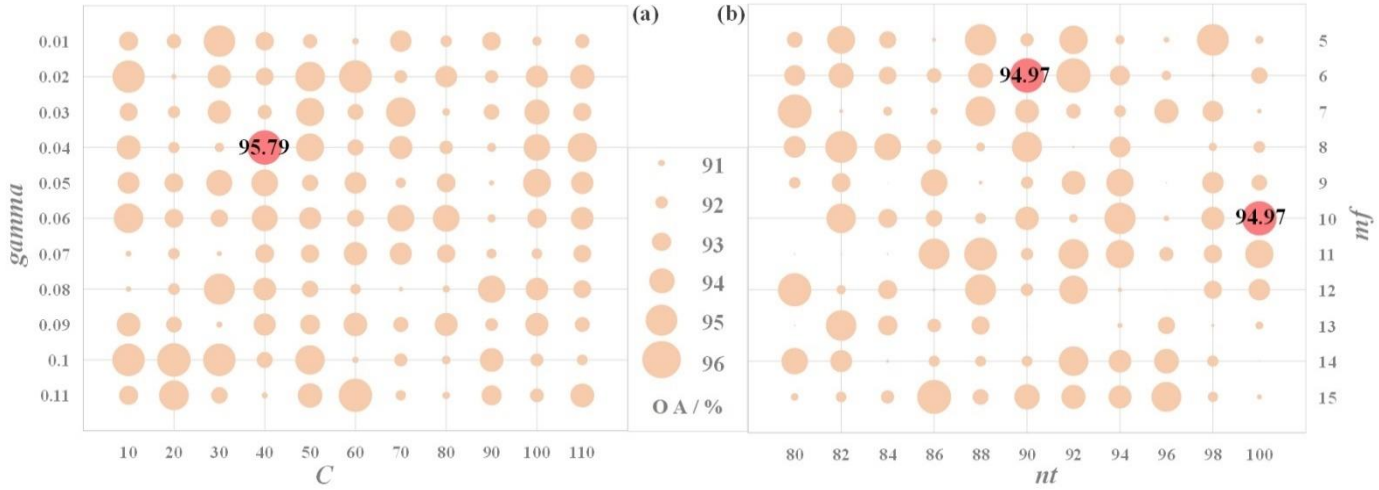


Figure 5. Optimization of characteristic parameters based on grid search.

4.1.2. Selection of waterbody extraction thresholds

For the samples of the test set and training set constructed in Section 3, we extracted the corresponding waterbody index information. We combined the sample labels and waterbody index information to generate confusion matrix, quantitatively verify the accuracy of waterbody extraction with different thresholds, and explore the best segmentation threshold for different waterbody indexes to improve the accuracy of waterbody extraction and reduce the subjectivity and contingency of the artificially threshold. In order to divide the threshold, we rewrite the formula of the two spectral relationship algorithms into the equation form: $SR = (\rho_{\text{green}} + \rho_{red}) - (\rho_{NIR} + \rho_{SWIR1})$ and $mSR = (\rho_{\text{green}} + \rho_{red}) - (\rho_{NIR} + \rho_{SWIR1}) - (\rho_{\text{blue}} - \rho_{green})$. Among the 8 waterbody extraction models based on threshold segmentation, except for the single band threshold (SBT) extraction algorithm, which defines the area less than the threshold as waterbodies, the other models define the area greater than the threshold as waterbodies The variation of waterbody extraction accuracy of each algorithm model with the thresholds is shown in Figure 6.

In the overall accuracy of waterbody extraction results based on the threshold segmentation models, it can be found that except for the SBT model, the best overall accuracy of other threshold segmentation models for water extraction has reached more than 70%. The NDWI threshold segmentation model has the highest overall accuracy for waterbody extraction at 83.62%, followed by mAEWI (82.02), AEWIs (76.89%), AEWIns (75.21%), mNDWI (74.90%), SR(74.51) and mSR (71.32%). The ranking of the overall accuracy of waterbody extraction results is highly consistent with the conclusions obtained in the previous study [55]. The SBT model relies on the low reflection of water at the 1.5~1.6μm spectrum to extract waterbody, but 1.5~1.6μm spectrum is also at the absorption peak of vegetation [37]. As shown in Figure 6(a), it can be seen that the threshold selection of -0.0137 cannot effectively distinguish water from landcover class 1 (Tree) and class 3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

(Grassland), so its overall accuracy is only 58.09%. In addition, when the threshold value is -0.06, the overall accuracy of SBT extraction result reaches the highest, at 63.77%. However, there are a large number of true negative (TN) samples in this waterbody extraction result, and all samples are classified as non-water bodies, so the threshold segmentation results fail to reach the requirements of waterbody extraction.

Additionally, the optimal segmentation thresholds selected by the overall accuracy can effectively distinguish waterbodies from other landcover classes except snow and ice, which can be mutually confirmed with some conclusions of reaches [55,80]. At the same time, it also shows that the threshold selection method based on the overall accuracy has high accuracy. By analyzing the selection of threshold values of each model, it can be seen that the thresholds of mAEWI, SR, mSR, AEWIs and AEWIns models are close to zero, which are -0.0085, 0.0086, 0.00189, 0.0397 and 0.0233, respectively, with strong anti-interference, consistent with the principle of each algorithm design [13,37,38,66].

Finally, comparing the blue boxes in Figure 6(e) and (f), it can be seen that when the thresholds of NDWI and mNDWI models are close to -1 and 1, the overall accuracy is significantly distorted. This is because there are a large number of abnormal values higher than 1 and some lower than -1 in the waterbody samples when using the surface reflection data for the division band operation. During the abnormal value processing, the values greater than 1 and less than -1 are reclassified as 1 and -1, resulting in the accumulation of a large number of waterbody samples at the thresholds of 1 and -1. The occurrence of a large number of outliers may be due to the use of min function in major areas during band cloud removal synthesis, which expands the shadow effect and leads to the increase of abnormal low values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

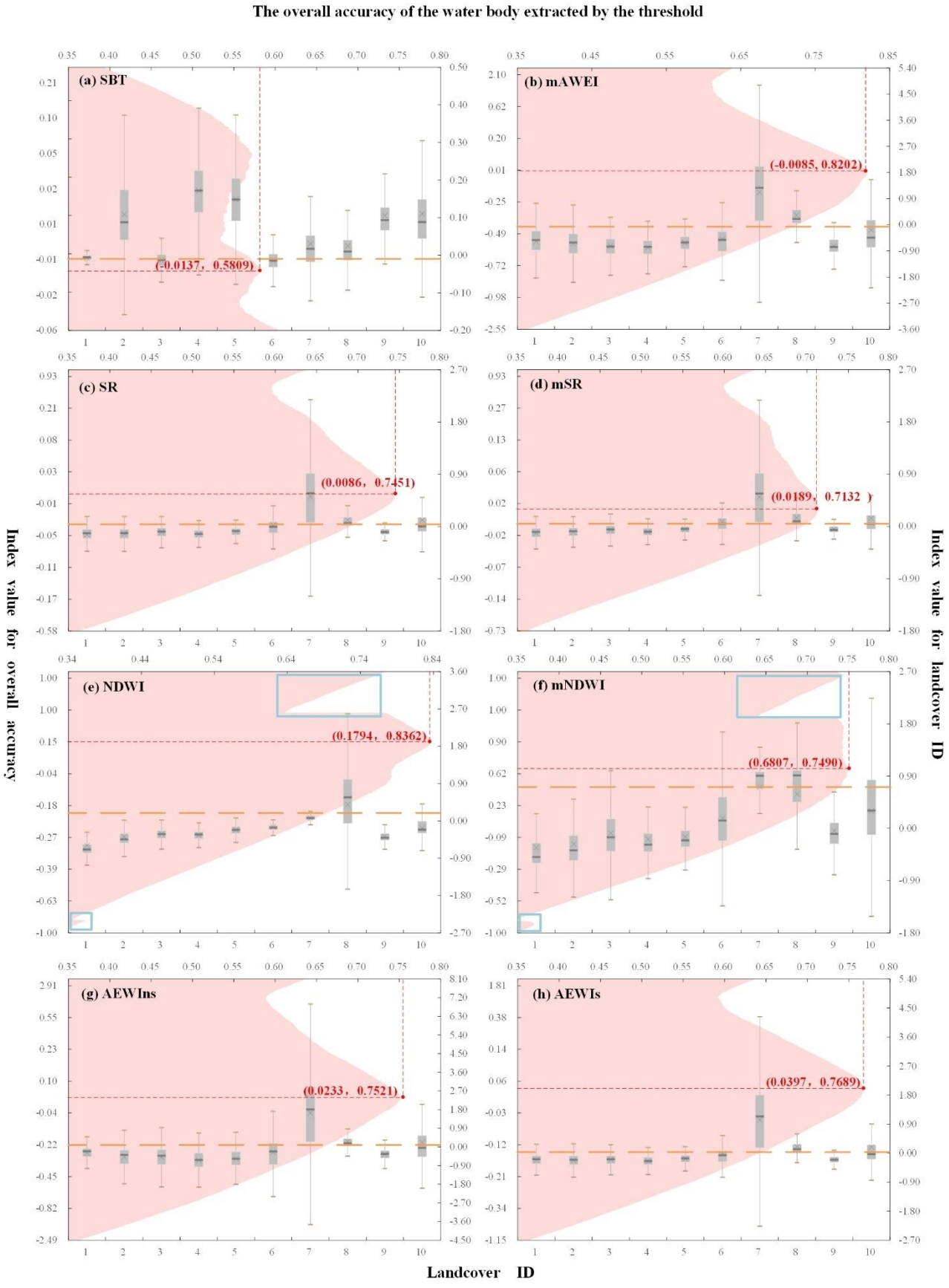


Figure 6. The overall accuracy of the waterbody extracted by the threshold segmentation. The bottom X-axis represents the Landcover code (as shown in Table 1), corresponding to the Y-axis on the right and the box diagram in the figure, reflecting the distribution of samples of different landcover

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

classes. In these box diagrams, boxes are interquartile ranges. Whiskers indicate the 1.5 times interquartile range. The horizontal line and cross sign in boxes represent the median and mean respectively. The top X-axis is matched with the left Y-axis to reflect the overall accuracy of segmentation results using different thresholds. The red dotted line intersection coordinates represent the best threshold and its overall accuracy of the model, and the yellow dotted line represents the position of the best segmentation threshold on the box diagram.

4.2. Analysis of waterbody extraction results

Based on the above optimization results of machine learning model parameters, we used the training samples to train and classify the SVM classifier and RF model, reclassified the results into waterbodies and non-water bodies, and calculated the confusion matrix on the validation set to evaluate the accuracy of waterbody extraction. Based on the above selected waterbody segmentation thresholds, we used each threshold segmentation model to extract waterbody, and used the validation set to verify the accuracy of the extracted results.

The accuracy validation results of the confusion matrix are shown in Table 3. By analyzing the overall accuracy of different algorithms, it can be seen that the accuracy validation results are highly consistent with the results of parameter optimization and threshold selection. On the one hand, RF and SVM algorithms have the highest classification accuracy, achieving an overall accuracy of over 95%, which is significantly better than other threshold segmentation waterbody extraction algorithms which are essentially linear segmentation algorithms in the spectral domain, while the RF and SVM with RBF kernel can realize the non-linear segmentation extraction in the spectral domain by using ensemble learning and non-linear kernels, which have stronger adaptability and generalization to the non-linear spectral domain features of different landcover classes. On the other hand, the order of the overall accuracy of each threshold segmentation waterbody extraction algorithm is consistent with the threshold selection results, and generally has an improvement of about 5%, and the overall accuracy of SBT model has increased by about 10%. This is due to the fact that in the threshold segmentation waterbody extraction algorithms, the landcover classes of Ice and snow, Moss and lichen, Tree cover (as shown in Figure 6), which are greatly confused with the waterbody threshold, account for a large proportion in the samples. When selecting thresholds, we selected 70% of the samples to calculate the overall accuracy. The sample imbalance is large, and the overall accuracy is low. The data set used for validation accounts for 30% of the samples, and the sample imbalance is reduced, so the overall accuracy is improved.

Additionally, analyzing the user's accuracy and producer's accuracy of different models, it can be seen that almost all models have relatively high producer's accuracy and low user's accuracy of waterbody extraction, while the user accuracy of non-water bodies is high and the producer's accuracy is low. This shows that in large-scale regions of the QTP, the omission error of waterbody extraction is not high, and the commission error is mainly concentrated in the misclassification of a large number of non-water bodies (Ice, Snow, Vegetation and shadows) into waterbodies. Comparing the kappa coefficients of different models, it can be seen that the prediction results of machine learning models are highly consistent with the actual results, indicating that the models are relatively stable. In the threshold segmentation waterbody extraction algorithms (except NDWI and mAEWI), the predicted results are less consistent with the actual results, indicating that the models are highly random and unstable.

Finally, compared with other threshold segmentation waterbody extraction studies [13,36–39,55], it can be seen that accuracy of threshold segmentation waterbody extraction models in this paper is relatively low. Comparing the experimental design and process of each study, we consider that there are two main reasons for this phenomenon. The first reason is the scale. Compared with other studies, the range of threshold segmentation waterbody extraction in this study exceeds 3 million km^2, is larger and the background structure is more complex, so the accuracy of waterbody extraction is reduced. The second

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

reason is the data pre-processing in this study. In order to reduce the cloud cover and improve the coverage of the Landsat-9 images to the study region, the method of segmented synthesis is used in this experiment by using the min function, which produces some low-value noise and reduces the data quality, thus affecting the accuracy of threshold segmentation of waterbody extraction. It should be emphasized that the generation of low-value noise is balanced for various models. Although it will affect the accuracy of waterbody extraction, it does not affect the reliability of algorithm comparison results.

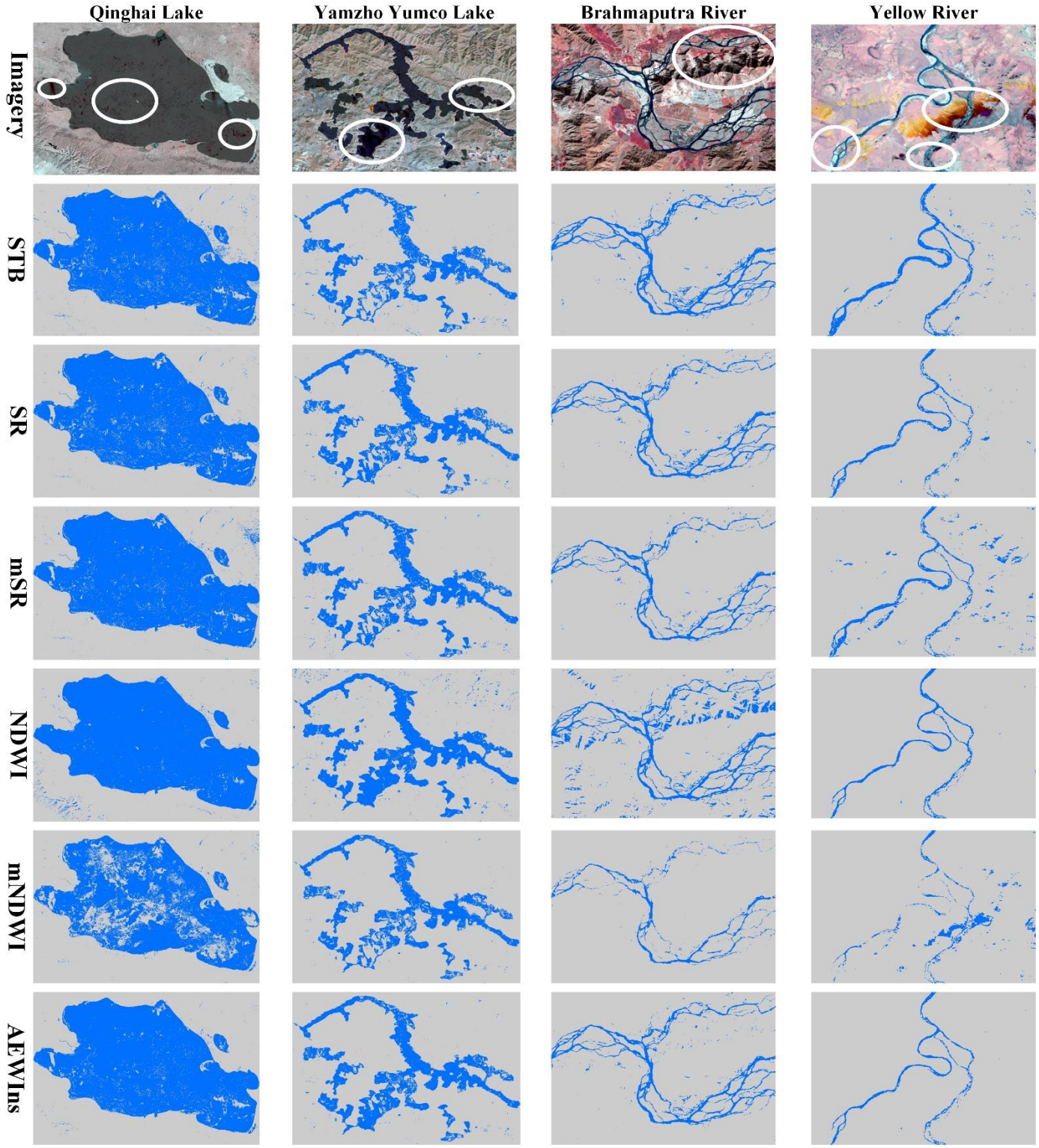
Table 3. Accuracy of waterbody extraction with different models.

Model	Cover type	Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa coefficient
SBT	Non-water	64.21	85.97	68.93	0.5331
	Water	75.42	52.01		
SR	Non-water	74.75	89.99	76.26	0.4880
	Water	79.91	56.69		
mSR	Non-water	70.36	91.34	74.31	0.4663
	Water	83.87	53.92		
NDWI	Non-water	87.81	97.66	89.89	0.7720
	Water	94.91	76.31		
mNDWI	Non-water	77.41	87.35	76.09	0.4654
	Water	72.89	57.16		
AWEIns	Non-water	80.42	93.57	82.24	0.6099
	Water	86.64	64.67		
AWEIs	Non-water	78.73	97.24	83.37	0.6461
	Water	94.59	64.79		
mAWEI	Non-water	94.52	88.57	87.50	0.6828
	Water	70.50	84.19		
SVM	Non-water	95.60	99.55	96.59	0.9198
	Water	98.96	90.30		
RF	Non-water	95.21	98.87	95.84	0.9021
	Water	97.36	89.38		

In order to more intuitively show the effect and performance of each model for water extraction, this study selects Qinghai LakeYamzho (the largest lake in the QTP), Yumco Lake (complex lake bay morphology), Yellow River (complex river morphology and background) and Brahmaputra River (complex river morphology and background) to compare the results of lake waterbody extraction. The white oval areas on the imageries in Figure 7 are A1, A2, A3, a4, A5, A6, A7, A8 and A9 from left to right. Comparing the deep-water areas from A1 to A5 that are prone to omission error, it can be seen that SVM and RF algorithms can well identify and extract these waterbody areas, while the threshold segmentation waterbody extraction algorithm has more omission errors. However, NDVI and AEWIs have relatively better performance as threshold segmentation waterbody extraction algorithms. There are small waterbodies and obvious shadows in A6 area. Except NDVI and AEWIs methods, all algorithms can better identify the shadows in this area, but MSR and mNDWI models perform poorly in identifying small waterbodies. There are obvious floodplain wetlands in A7, A8 and their upstream areas due to the lateral movement of rivers. Except SVM, RF, NDWI and AEWIns algorithms, other algorithms have commission errors to recognize wetlands as waterbodies. A9 area is an abnormal area caused by image synthesis algorithm. Except mAEWI, mNDWI, MSR and SBT models,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

other algorithms can better resist these abnormal values. In conclusion, we found that SVM, RF and NDWI models can relatively better realize waterbodies recognition in different lake waterbody regions.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

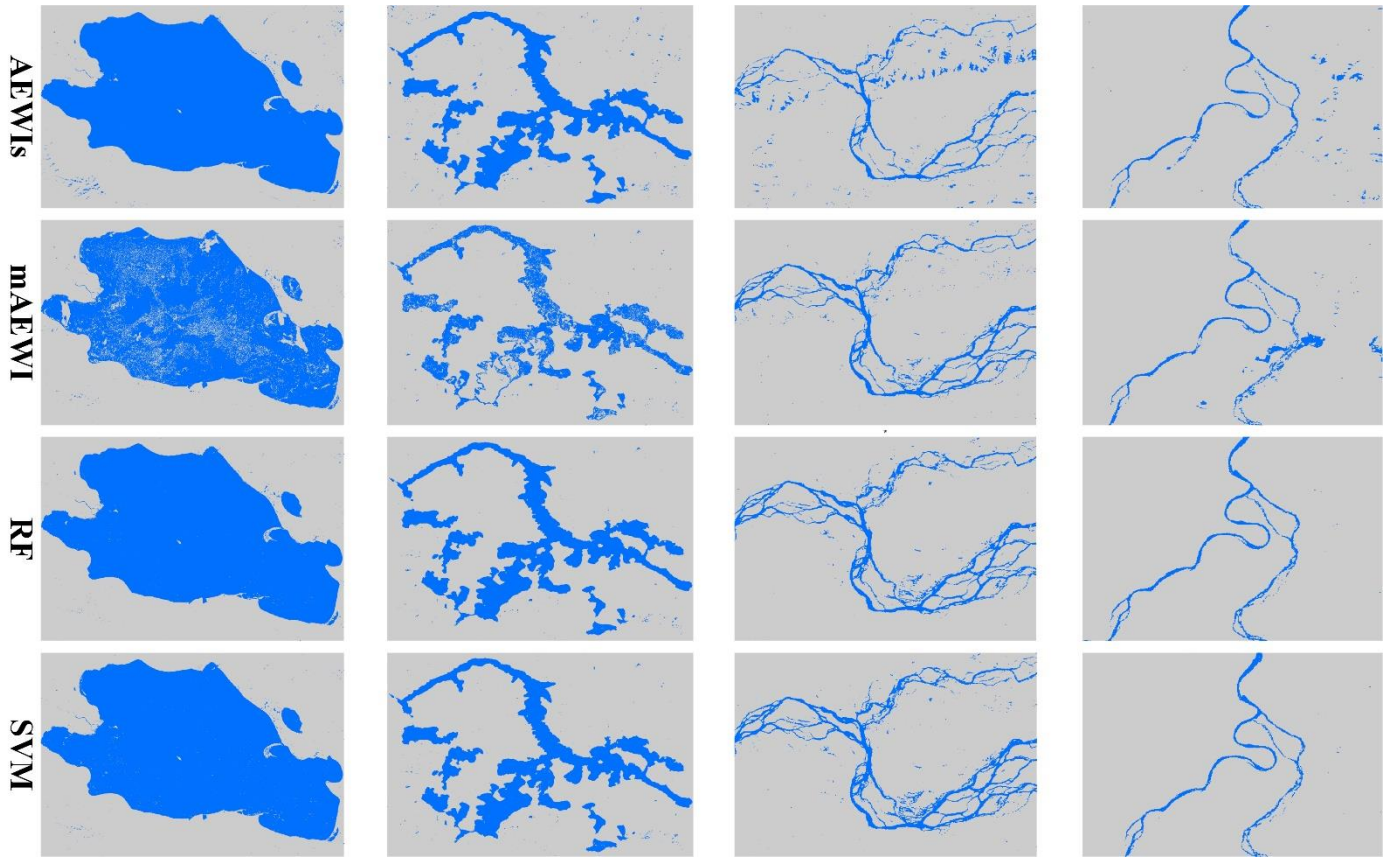


Figure 7. Comparison of the recognition accuracy of different water extraction algorithms for rivers and lakes. L9C2_SR imageries were synthesized in false color in bands 5, 4 and 3, and stretched with 2.5 standard deviations. The white oval circles on the image are the regions where the recognition abnormality may occur, which is A1, A2, A3, A4, A5, A6, A7, A8 and A9 from left to right. The coordinates of the central points of the four images are (100°10′29.0″E, 36°54′6.9″N), (90°44′44.1″E, 28°59′22.7″N), (90°45′34.8″E, 29°20′5.3″N), (102°27′50.5″E, 33°27′20.8″N) from left to right. Regions colored in Blue are waterbody areas and Gray are non-water bodies.

4.3. Analysis of lake waterbody area extraction results

In this research, the lake waterbody dataset was produced by the non-lake waterbody dataset and the SRTM DEM, the former was constructed in Section 3.1 was used to eliminate the non-lake waterbodies in the extraction results [65], and the later was used to correct the shadow influence [42]. Referring to the classification of lake waterbody, the area distribution of lake waterbody in the QTP region the resolution of lake waterbody extraction in previous study [3], the lakes waterbody area in this study are divided into 30-100 km² (type 1), 100-500 km² (type 2) and over 500 km² (type 3). As shown in Table 4, we randomly selected 10 lakes in each category to build a lake waterbody area validation set, and TPLA_V3 data was used as the true value to quantitatively analyze the extraction results of lake waterbody area, so as to measure the accuracy of extracting lake waterbody area of three types by different algorithms.

Table 4. Extraction results of lake waterbody area by different models.

ID	Central location (°)	Area/km²	Type	ID	Central location/°	Area/km²	Type
01	89.4541E, 32.3388N	34.7176	1	16	83.0601E, 35.2735N	246.6113	2
02	90.5172E, 28.9495N	38.6561	1	17	88.7222E, 31.5855N	268.9577	2
03	88.6920E, 32.3169N	44.1437	1	18	90.1922E, 35.7528N	294.7249	2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

04	82.3336E, 31.6280N	54.4298	1	19	89.4425E, 36.3302N	386.9375	2
05	86.2695E, 35.2978N	62.2600	1	20	73.4069E, 39.0290N	413.4563	2
06	88.1369E, 36.1962N	65.5249	1	21	88.9544E, 34.5827N	511.8270	3
07	86.7384E, 31.5679N	67.8710	1	22	97.2666E, 34.9309N	549.8463	3
08	85.2322E, 31.5679N	71.8135	1	23	88.2833E, 31.1572N	555.8818	3
09	85.8104E, 33.6649N	73.8108	1	24	88.3989E, 37.0775N	599.5535	3
10	95.8114E, 36.7416N	80.8892	1	25	97.5895E, 38.2928N	639.1734	3
11	91.1596E, 31.7089N	148.5164	2	26	97.7020E, 34.9062N	656.0598	3
12	87.1744E, 34.5513N	167.8379	2	27	90.4774E, 34.7984N	693.1980	3
13	89.9783E, 32.4493N	188.9746	2	28	85.6116E, 30.9289N	1048.9105	3
14	92.1340E, 35.2207N	203.0216	2	29	90.0623E, 33.4379N	1137.2408	3
15	84.5659 E, 35.4053N	221.7653	2	30	100.1977E, 36.8884N	4538.2366	3

Figure 8 shows the error analysis results of lake waterbody extraction. Among the all algorithms, the accuracy of the lake waterbody area extracted by the RF model is the highest (1.505%), followed by the SVM model (1.624%). In the threshold segmentation algorithm, the NDWI (3.501%) is the most accurate model, followed by AEWIs (6.789%). Comparing the area extraction accuracy of different types of lakes, we found that in almost all models, the error of large lakes is the largest and that of medium lakes is the smallest. In combination with Figure 7, we analyze that there are two main reasons for extraction errors. First, there are many deep-water areas in large lakes, resulting in large omission errors. Second, the spatial resolution of L9C2_SR data is 30m. When identifying small lakes, it is easy to generate commission errors at the mixed pixels at the waterway junction. Therefore, error analysis of small lakes is greatly influenced by error classification, resulting in large error fluctuations.

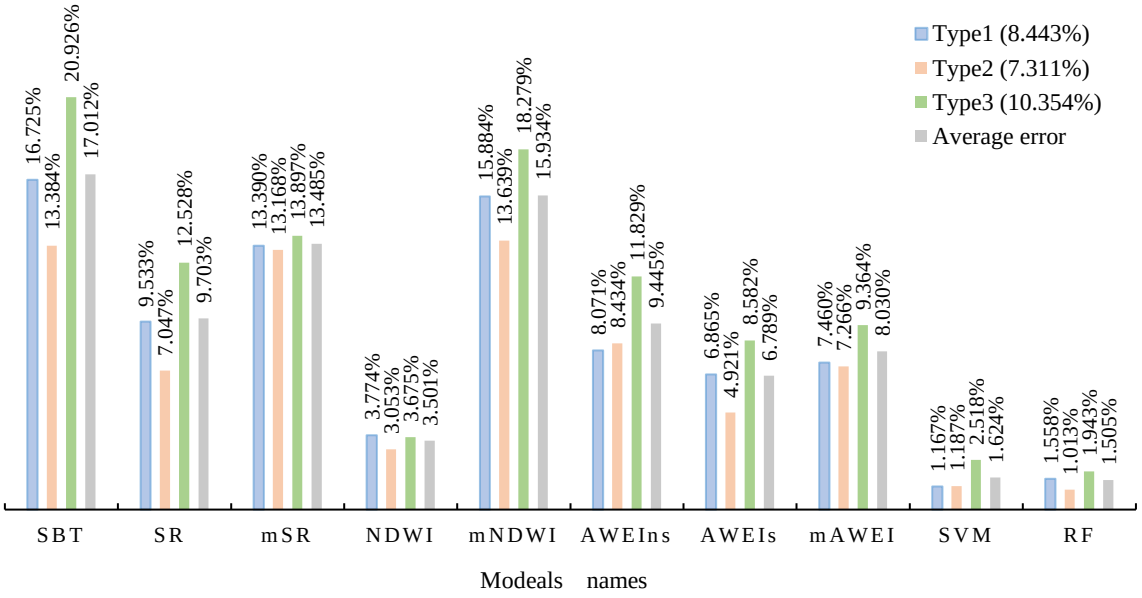


Figure 8. The error analysis (See 3.4 for error analysis formula) of different waterbody extraction algorithms for various types of lake area extraction. The X- axis of the histogram is the name of the model, and the percent labels are the average extraction error of various lake areas under various models. In the percent labels behind the legend are the average error of the lake waterbody area extraction in each algorithm.

5. Conclusions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

It is of great significance for global climate change to quickly and accurately obtain the changes of lakes in the QTP based on remote sensing technology, but there is a certain relationship between the performance of different lake waterbody extraction algorithms and application scenarios with remote sensing data. In order to explore the interaction between Landsat-9 data and the accuracy of different algorithm models for QTP lake waterbody extraction, and determine the algorithm for large-scale QTP lake waterbody area extraction suitable for Landsat-9 imagery, this study selected 10 models widely used waterbody extraction, carried out comparative research leveraging GEE platform, and found the following conclusions.

Affected by clouds and shadows, the Landsat-9 data with limited quality and only 30m resolution, the waterbody extraction model still achieved the best 96.59% overall accuracy and 1.505% average error in the extraction of QTP waterbody and lake waterbody area extraction. It is proved that those algorithms can effectively extract and identify QTP waterbody using Landsat-9 data. With the publication of 15-meter panchromatic data and the acquisition of high-quality images, there is still greater improvement in the accuracy of waterbody and lake waterbody area extraction. Moreover, compared with the threshold segmentation algorithms, the machine learning algorithms have significant advantages in extracting large-scale QTP waterbody and lake waterbody area. Comparing the two machine learning algorithms, under the condition of equal overall accuracy and average error, the operation efficiency of RF on GEE is significantly higher than that of SVM classifier. Therefore, RF algorithm is more recommended in similar studies. With the development of GEE platform, the constraint of computing power on model selection will be greatly reduced. More models with complex calculation but higher accuracy can be considered in future related research. Finally, among the traditional threshold segmentation waterbody extraction algorithms, the best extraction result is the NDWI method. The overall accuracy of waterbody extraction is 89.89%, and the average error of lake waterbody area extraction is 3.501%. The NDWI is recommended method in scenarios with limited samples or high operational efficiency.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

Author Contributions: Conceptualization, L.X. and Z.Y.; methodology, Z.D.; software, J.C.; investigation, S.Y. and L.D.; data curation, Q.K. ; writing—original draft preparation, L.X and J.C.; writing—review and editing, L.Y and C.D.; visualization, L.S.; supervision, Z.D.; project administration, Z.Y.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Major Science and Technology Projects in Anhui Province (NO. 202003a06020002)、Key Research and Development Projects of Anhui Province (NO. 2021003)、Special Support Plan for High-Level Talents of Anhui Province(NO. 2019) and Wuhu key R & D plan(2020ms1-3)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are also thankful to all anonymous reviewers for their constructive comments provided on the study.

Conflicts of Interest: The authors declare no conflict of interest.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

References

1. Long, D; Li, X; Li, X; Hang, P; Zhao, F; Hong, Z; Wang, Y; Tian, F. Remote sensing retrieval of water storage changes and underlying climatic mechanisms over the Tibetan Plateau during the past two decades. *Advances in Water Science.* **2022**,1–15.
2. Yao, T.; Thompson, L.; Yang, W.; Yu, W.; Gao, Y.; Guo, X.; Yang, X.; Duan, K.; Zhao, H.; Xu, B. Different Glacier Status with Atmospheric Circulations in Tibetan Plateau and Surroundings. *Nat. Clim. Change.* **2012**, *2*, 663–667.
3. Ma, R; Yang, G; Duan, H; Jiang, J; Wang, S; Feng, X; Li, A; Kong, F; Xu, B; Wu, J; Li, S. China's lakes at present: Number, area and spatial distribution. *Scientia Sinica(Terrae).* **2011**, *41*, 394–401.
4. Zheng, G.; Yang, Y.; Yang, D.; Dafflon, B.; Yi, Y.; Zhang, S.; Chen, D.; Gao, B.; Wang, T.; Shi, R.; et al. Remote Sensing Spatiotemporal Patterns of Frozen Soil and the Environmental Controls over the Tibetan Plateau during 2002–2016. *Remote Sens. Environ.* **2020**, *247*, 111927, doi:10.1016/j.rse.2020.111927.
5. Zhang, G.; Yao, T.; Xie, H.; Yang, K.; Zhu, L.; Shum, C.K.; Bolch, T.; Yi, S.; Allen, S.; Jiang, L. Response of Tibetan Plateau Lakes to Climate Change: Trends, Patterns, and Mechanisms. *Earth-Sci. Rev.* **2020**, *208*, 103269.
6. Immerzeel, W.W.; Van Beek, L.P.; Bierkens, M.F. Climate Change Will Affect the Asian Water Towers. *science.* **2010**, *328*, 1382–1385.
7. Immerzeel, W.W.; Lutz, A.F.; Andrade, M.; Bahl, A.; Biemans, H.; Bolch, T.; Hyde, S.; Brumby, S.; Davies, B.J.; Elmore, A.C. Importance and Vulnerability of the World's Water Towers. *Nature.* **2020**, *577*, 364–369.
8. Song, C; Zhan, P; Ma, R Progress in remote sensing study on lake hydrologic regime. *Journal of Lake Sciences.* **2020**, *32*, 1406–1420.
9. Lu, S; Wu, B; Yan, N; Li, F; Wen, M; Wang, J Progress in River Runoff Monitoring by Remote Sensing. *Advances in Earth Science.* **2010**, *25*, 820–826.
10. Ma, B.; Wu, L.; Zhang, X.; Li, X.; Liu, Y.; Wang, S. Locally Adaptive Unmixing Method for Lake-Water Area Extraction Based on MODIS 250 m Bands. *Int. J. Appl. Earth Obs. Geoinformation.* **2014**, *33*, 109–118.
11. Ling, F.; Li, X.; Foody, G.M.; Boyd, D.; Ge, Y.; Li, X.; Du, Y. Monitoring Surface Water Area Variations of Reservoirs Using Daily MODIS Images by Exploring Sub-Pixel Information. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 141–152.
12. Hong, S.; Jang, H.; Kim, N.; Sohn, H.-G. Water Area Extraction Using RADARSAT SAR Imagery Combined with Landsat Imagery and Terrain Information. *Sensors.* **2015**, *15*, 6652–6667.
13. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A New Technique for Surface Water Mapping Using Landsat Imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35.
14. Wang, J.; Ding, J.; Li, G.; Liang, J.; Yu, D.; Aishan, T.; Zhang, F.; Yang, J.; Abulimiti, A.; Liu, J. Dynamic Detection of Water Surface Area of Ebinur Lake Using Multi-Source Satellite Data (Landsat and Sentinel-1A) and Its Responses to Changing Environment. *Catena.* **2019**, *177*, 189–201.
15. Lou, L.; Chen, C.; Gao, X.; Liu, K.; Li, M.; Fu, Y. Comparative Research on Water Body Extraction Methods Based on SPOT Data. In Proceedings of the Proceedings of the 7th China High Resolution Earth Observation Conference (CHREOC 2020); Springer, 2022; pp. 247–255.
16. Ogilvie, A.; Belaud, G.; Massuel, S.; Mulligan, M.; Le Goulven, P.; Calvez, R. Surface Water Monitoring in Small Water Bodies: Potential and Limits of Multi-Sensor Landsat Time Series. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 4349–4380.
17. Ji, H; Fan, X; Wu, G; Liu, Y Accuracy comparison and analysis of methods for water area extraction of discrete lakes. *Journal of Lake Sciences.* **2015**, *27*, 327–334.
18. Sivanpillai, R.; Miller, S.N. Improvements in Mapping Water Bodies Using ASTER Data. *Ecol. Inform.* **2010**, *5*, 73–78.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

19. Fujisada, H.; Urai, M.; Iwasaki, A. Manual-Based Improvement Method for the ASTER Global Water Body Data Base. *Remote Sens.* **2020**, *12*, 3373.
20. Liu, H.; Zheng, L.; Jiang, L.; Liao, M. Forty-Year Water Body Changes in Poyang Lake and the Ecological Impacts Based on Landsat and HJ-1 A/B Observations. *J. Hydrol.* **2020**, *589*, 125161.
21. Lei, Z.; Bing, Z.; Junsheng, L.; Qian, S.; Fangfang, Z.; Ganlin, W. A Study on Retrieval Algorithm of Black Water Aggregation in Taihu Lake Based on HJ-1 Satellite Images. In Proceedings of the IOP Conference Series: Earth and Environmental Science; IOP Publishing, 2014; Vol. 17, p. 012100.
22. Martins, V.S.; Kaleita, A.L.; Gelder, B.K.; Nagel, G.W.; Maciel, D.A. Deep Neural Network for Complex Open-Water Wetland Mapping Using High-Resolution WorldView-3 and Airborne LiDAR Data. *Int. J. Appl. Earth Obs. Geoinformation.* **2020**, *93*, 102215.
23. Cooley, S.W.; Smith, L.C.; Stepan, L.; Mascaro, J. Tracking Dynamic Northern Surface Water Changes with High-Frequency Planet CubeSat Imagery. *Remote Sens.* **2017**, *9*, 1306.
24. Lu, D.; Hetrick, S.; Moran, E. Impervious Surface Mapping with Quickbird Imagery. *Int. J. Remote Sens.* **2011**, *32*, 2519–2533.
25. Ekercin, S. Water Quality Retrievals from High Resolution Ikonos Multispectral Imagery: A Case Study in Istanbul, Turkey. *Water. Air. Soil Pollut.* **2007**, *183*, 239–251, doi:10.1007/s11270-007-9373-5.
26. Guo, H.; He, G.; Jiang, W.; Yin, R.; Yan, L.; Leng, W. A Multi-Scale Water Extraction Convolutional Neural Network (MWEN) Method for GaoFen-1 Remote Sensing Images. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 189.
27. Zhang, Z.; He, H.; Yu, C.; Zhang, W.; Li, L.; Meng, L. Using the Modified Two-Mode Method to Identify Surface Water in Gaofen-1 Images. *J. Appl. Remote Sens.* **2018**, *13*, 022003.
28. Yang, F.; Guo, J.; Tan, H.; Wang, J. Automated Extraction of Urban Water Bodies from ZY-3 Multi-Spectral Imagery. *Water.* **2017**, *9*, 144.
29. Yao, F.; Wang, C.; Dong, D.; Luo, J.; Shen, Z.; Yang, K. High-Resolution Mapping of Urban Surface Water Using ZY-3 Multi-Spectral Imagery. *Remote Sens.* **2015**, *7*, 12336–12355.
30. Ferrentino, E.; Nunziata, F.; Buono, A.; Urciuoli, A.; Migliaccio, M. Multipolarization Time Series of Sentinel-1 SAR Imagery to Analyze Variations of Reservoirs' Water Body. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 840–846.
31. Yang, J.; Wang, X.; Wang, J.; Ye, C.; Xiong, J. Water Extraction of Hyperspectral Imagery Based on a Fast and Effective Decision Tree Water Index. *J. Appl. Remote Sens.* **2021**, *15*, 042605.
32. Luo, X.; Xie, H.; Tong, X.; Pan, H. A Water Extraction Method Based on Airborne Hyperspectral Images in Highly Complex Urban Area. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP); IEEE, 2017; pp. 1–4.
33. Zhu, Z.; Wulder, M.A.; Roy, D.P.; Woodcock, C.E.; Hansen, M.C.; Radeloff, V.C.; Healey, S.P.; Schaaf, C.; Hostert, P.; Strobl, P.; et al. Benefits of the Free and Open Landsat Data Policy. *Remote Sens. Environ.* **2019**, *224*, 382–385, doi:10.1016/j.rse.2019.02.016.
34. Bannari, A.; Ali, T.S.; Abahussain, A. The Capabilities of Sentinel-MSI (2A/2B) and Landsat-OLI (8/9) in Seagrass and Algae Species Differentiation Using Spectral Reflectance. *Ocean Sci.* **2022**, *18*, 361–388, doi:10.5194/os-18-361-2022.
35. Song, C.; Huang, B.; Ke, L.; Richards, K.S. Remote Sensing of Alpine Lake Water Environment Changes on the Tibetan Plateau and Surroundings: A Review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 26–37, doi:10.1016/j.isprsjprs.2014.03.001.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

36. Frazier, P.S.; Page, K.J.; others Water Body Detection and Delineation with Landsat TM Data. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 1461–1468.
37. Du, Y，Zhou, C. Automatically Extracting Remote Sensing Information Method for Water Bodies. *National Remote Sensing Bulletin.* **1998**, 264–269.
38. Yang, S; Xue, C; Liu, T; Li, T A Method of Small Water Information Automatic Extraction from TM Remote Sensing Images. *Acta Geodaetica et Cartographica Sinica.* **2010**, *39*, 611–617.
39. McFEETERS, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432, doi:10.1080/01431169608948714.
40. Xu, H. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033, doi:10.1080/01431160600589179.
41. Yan, P; Zhang, Y; Zhang, Y A Study on Information Extraction of Water System in Semi-arid Regions with the Enhanced Water Index (EWI)and GIS Based Noise Remove Techniques. *Remote Sensing Information.* **2007**, 62–67.
42. Chen, W; Ding, J; Li, Y; Niu, Z Extraction of Water Information based on China-made GF-1 Remote Sense Image. *Resources Science.* **2015**, *37*, 1166–1172.
43. Chandrasekar, K.; Sesha Sai, M.V.R.; Roy, P.S.; Dwevedi, R.S. Land Surface Water Index (LSWI) Response to Rainfall and NDVI Using the MODIS Vegetation Index Product. *Int. J. Remote Sens.* **2010**, *31*, 3987–4005, doi:10.1080/01431160802575653.
44. Ma, R.; Duan, H.; Hu, C.; Feng, X.; Li, A.; Ju, W.; Jiang, J.; Yang, G. A Half-century of Changes in China's Lakes: Global Warming or Human Influence? *Geophys. Res. Lett.* **2010**, *37*.
45. Tao, S.; Fang, J.; Ma, S.; Cai, Q.; Xiong, X.; Tian, D.; Zhao, X.; Fang, L.; Zhang, H.; Zhu, J. Changes in China's Lakes: Climate and Human Impacts. *Natl. Sci. Rev.* **2020**, *7*, 132–140.
46. Zou, Z.; Dong, J.; Menarguez, M.A.; Xiao, X.; Qin, Y.; Doughty, R.B.; Hooker, K.V.; Hambright, K.D. Continued Decrease of Open Surface Water Body Area in Oklahoma during 1984–2015. *Sci. Total Environ.* **2017**, *595*, 451–460.
47. Jin, S.; Liu, Y.; Fagherazzi, S.; Mi, H.; Qiao, G.; Xu, W.; Sun, C.; Liu, Y.; Zhao, B.; Fichot, C.G. River Body Extraction from Sentinel-2A/B MSI Images Based on an Adaptive Multi-Scale Region Growth Method. *Remote Sens. Environ.* **2021**, *255*, 112297, doi:10.1016/j.rse.2021.112297.
48. Chen, C; Fu, J; Sui, X; Lu, X; Tan, A. Construction and Application of Knowledge Decision Tree After a Disaster for Water Body Information Extraction from Remote Sensing Images. *National Remote Sensing Bulletin*. **2018**, *05*, 792-801.
49. Razaque, A.; Ben Haj Frej, M.; Almi'ani, M.; Alotaibi, M.; Alotaibi, B. Improved Support Vector Machine Enabled Radial Basis Function and Linear Variants for Remote Sensing Image Classification. *Sensors* **2021**, *21*, 4431.
50. Zhang, S.L.; Chang, T.C. A Study of Image Classification of Remote Sensing Based on Back-Propagation Neural Network with Extended Delta Bar Delta. *Math. Probl. Eng.* **2015**, *2015*.
51. Ko, B.C.; Kim, H.H.; Nam, J.Y. Classification of Potential Water Bodies Using Landsat 8 OLI and a Combination of Two Boosted Random Forest Classifiers. *Sensors* **2015**, *15*, 13763–13777.
52. Yu, L.; Wang, Z.; Tian, S.; Ye, F.; Ding, J.; Kong, J. Convolutional Neural Networks for Water Body Extraction from Landsat Imagery. *Int. J. Comput. Intell. Appl.* **2017**, *16*, 1750001.
53. Sui, H.; Chen, G.; Hua, L. An Automatic Integrated Image Segmentation, Registration and Change Detection Method for Water-Body Extraction Using HSR Images and GIS Data. *Int Arch Photogramm Remote Sens Spat Inf Sci* **2013**, *7*, W2.
54. Qiao, C.; Luo, J.; Sheng, Y.; Shen, Z.; Zhu, Z.; Ming, D. An Adaptive Water Extraction Method from Remote Sensing Image Based on NDWI. *J. Indian Soc. Remote Sens.* **2012**, *40*, 421–433.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

55. Zhang, G.; Li, J.; Zheng, G. Lake-Area Mapping in the Tibetan Plateau: An Evaluation of Data and Methods. *Int. J. Remote Sens.* **2017**, *38*, 742–772.
56. Deng, Y.; Jiang, W.; Tang, Z.; Ling, Z.; Wu, Z. Long-Term Changes of Open-Surface Water Bodies in the Yangtze River Basin Based on the Google Earth Engine Cloud Platform. *Remote Sens.* **2019**, *11*, 2213.
57. Li, Y.; Niu, Z.; Xu, Z.; Yan, X. Construction of High Spatial-Temporal Water Body Dataset in China Based on Sentinel-1 Archives and GEE. *Remote Sens.* **2020**, *12*, 2413.
58. Tang, H.; Lu, S.; Ali Baig, M.H.; Li, M.; Fang, C.; Wang, Y. Large-Scale Surface Water Mapping Based on Landsat and Sentinel-1 Images. *Water* **2022**, *14*, 1454.
59. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, doi:10.1029/2005RG000183.
60. Zhang, G.; Luo, W.; Chen, W.; Zheng, G. A Robust but Variable Lake Expansion on the Tibetan Plateau. *Sci. Bull.* **2019**, *64*, 1306–1309.
61. Zhang, G.; Yao, T.; Xie, H.; Zhang, K.; Zhu, F. Lakes' State and Abundance across the Tibetan Plateau. *Chin. Sci. Bull.* **2014**, *59*, 3010–3021.
62. Zhang, Y; Li, B; Liu, L; Zheng, D. Redetermine the Region and Boundaries of Tibetan Plateau. *Geographical Research*. **2021**, *06*,1543-1553.
63. Lehner, B.; Grill, G. Global River Hydrography and Network Routing: Baseline Data and New Approaches to Study the World's Large River Systems. *Hydrol. Process.* **2013**, *27*, 2171–2186.
64. Liu, L.; Xiao, X.; Qin, Y.; Wang, J.; Xu, X.; Hu, Y.; Qiao, Z. Mapping Cropping Intensity in China Using Time Series Landsat and Sentinel-2 Images and Google Earth Engine. *Remote Sens. Environ.* **2020**, *239*, 111624.
65. Zhang, G; Wang, M; Zhou, T; Chen, W. Progress in Remote Sensing Monitoring of Lake Area，Water Level，and Volume Changes on the Tibetan Plateau. *National Remote Sensing Bulletin*. **2022**, *26*, 115–125.
66. Nie, Y; Yu, M; Lan, T. Research on Water Information Extraction based on MAWEI Index. *Journal of Earth Environment*. **2019**, *10*, 281–290.
67. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259.
68. Yu, L.; Porwal, A.; Holden, E.-J.; Dentith, M.C. Towards Automatic Lithological Classification from Remote Sensing Data Using Support Vector Machines. *Comput. Geosci.* **2012**, *45*, 229–239.
69. Ji, L.; Zhang, L.; Wylie, B. Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *Photogramm. Eng. Remote Sens.* **2009**, *75*, 1307–1317.
70. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66.
71. Huang, M.; Yu, W.; Zhu, D. An Improved Image Segmentation Algorithm Based on the Otsu Method. In Proceedings of the 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing; IEEE, 2012; pp. 135–139.
72. Liu, C.; Frazier, P.; Kumar, L. Comparative Assessment of the Measures of Thematic Classification Accuracy. *Remote Sens. Environ.* **2007**, *107*, 606–616, doi:10.1016/j.rse.2006.10.010.
73. Li, J.; Sheng, Y. An Automated Scheme for Glacial Lake Dynamics Mapping Using Landsat Imagery and Digital Elevation Models: A Case Study in the Himalayas. *Int. J. Remote Sens.* **2012**, *33*, 5194–5213, doi:10.1080/01431161.2012.657370.
74. Raschka, S.; Mirjalili, V. Python Machine Learning: Machine Learning and Deep Learning with Python. *Scikit-Learn TensorFlow Second Ed. Ed* **2017**, *10*, 3175783.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2022 doi:10.20944/preprints202206.0020.v1

75. Albatal, R.; Little, S. Empirical Exploration of Extreme SVM-RBF Parameter Values for Visual Object Classification. In Proceedings of the International Conference on Multimedia Modeling; Springer, 2014; pp. 299–306.
76. Hsu, C.-W.; Lin, C.-J. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425.
77. Zheng, X; Qian, F. Support Vector Machine with Gauss Kernel Classified Method and Model Selection. *Computer Engineering and Applications*. **2006**, *42*, 77–79.
78. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31.
79. Lawrence, R.L.; Wood, S.D.; Sheley, R.L. Mapping Invasive Plants Using Hyperspectral Imagery and Breiman Cutler Classifications (RandomForest). *Remote Sens. Environ.* **2006**, *100*, 356–362.
80. Liu, Y; Gao, Y. Surface Water Extraction in Yangtze River Basin based on Sentinel Time Series Image. *National Remote Sensing Bulletin*. **2022**, *26*, 358–372.