Urban lakes change extraction using time series GaoFen-1 satellite imagery

Gan Luo[1,2] , Wei Jiang[1*], Lingjia Liu[2], Qingke Wen[3], Shuo Liu[3], Weichao Sun [3]

[1] Research Center of Flood and Drought Disaster Reduction of the Ministry of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing, China-jiangwei@iwhr.com,luogan_jxsr@163.com

[2] School of Geography and Environment, Jiangxi Normal University, Nanchang, China. liulingjia_office@jxnu.edu.cn

[3] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China-wenqk@radi.ac.cn; liushuo@radi.ac.cn; sunwc@radi.ac.cn

KEY WORDS: GF-1 Satellite, Water Body Extraction, Random Forest Algorithm, Urban Lakes Changes

ABSTRACT:

Urban lakes serve an indispensable role in maintaining the ecological balance of cities, ensuring flood safety, and providing recreational spaces for tourism. With the development of human activities and economic, the extent of urban lakes are inevitably influenced. Currently, the ability to detect detailed temporal changes in urban lake areas using high resolution data still has limitations. This study proposed a novel method by combining time series Gaofen-1 (GF-1) remote sensing data and random forest machine learning algorithm to explore the urban lakes change Zhushan Lake located in Wuhan. The research conducted the extraction of surface water for Zhushan Lake and its surrounding pit-ponds from 2013 to 2020. And then, a quantitative analysis of the characteristics and driving factors of lake changes is conducted. We find that (1) the accuracy of surface water extraction using the random forest classification method consistently exceeded 96%. The Kappa coefficient ranges from a minimum of 0.86 to a maximum of 0.99. (2) A noticeable decline was observed in the water areas of Zhushan Lake and its surrounding pit-ponds, predominantly along the northwestern shoreline and in the eastern pond regions. This decline is primarily attributed to pressures from building construction. The methodology proposed in this study is suitable for the area management of lakes in urban areas.

1. INTRODUCTION

Urban lakes play a pivotal role in the regional natural and human environments, which only supply water for living and production but also exhibit multiple functions, including flood control (Hayashi et al., 2008), urban heat island regulation(Yang et al., 2015), water storage(Song et al., 2013), tourism(Gossling et al., 2012), and climate regulation(Yan et al., 2013). These functions offer significant impetus for the sustainable development of cities. However, urban lakes have faced severe threats due to the urbanization in recent years (Zhang et al., 2018), such as water pollution(Brock et al., 2006) and Lake area shrinks (Kai et al., 2010).Therefore, investigating the changes of urban lake area is of great significance for urban environment sustainable development.

Objectively and accurately observing and understanding the changes in the area of urban lakes is a key indicator of urban land use mapping and ecosystem assessment. Remote sensing, with its high spatial and temporal resolution, wide coverage, long time-series data, diversity, and real-time observation capabilities, has significant advantages in observing changes in urban lakes, making it a crucial tool for exploring the change pattern of urban lakes.

The surface water extent of urban lakes is an important indicator to measure change pattern. Previous study have developed different water extraction methods (He et al., 2018). These methods are mainly categorized into two types: threshold segmentation methods and classifier model methods. The threshold segmentation method can be further divided into single-band method, spectral relationship method, and water body index method. For example, Liu employed the single-band method and multi-source remote sensing data to extract the water area of Lake Chad in Africa (Liu et al., 2013); Jiang combined the spectral features of vegetation red edge and shortwave infrared bands in Sentinel-2 satellite imagery, proposing a new water index termed surface water index (SWI)(Jiang et al., 2021);Zhang employed Landsat imagery and Gaofen data to extract and analyze the spatiotemporal variations in the surface area of Poyang Lake using the surface water index method.(Zhang et al., 2019).

The threshold method requires the manual setting of segmentation thresholds according to prior knowledge, which can cause errors when extracting water bodies in a complex urban environment area (Zhou et al., 2020). The classifier model method treats the water body as a category of land cover and applies classifier algorithm rules to extract water body from remote sensing imagery. The popular methods include object-oriented, decision tree, support vector machine, neural network and random forest(Jiang et al., 2018). For example, Shen’s study proposed a mountainous water body decision tree extraction model based on GF-5 image data and extracted the water body in the Dongchuan District of Kunming City(Shen et al., 2021). Dong’s research employed Landsat 7 ETM data and a support vector machine to determine the surface water area of Guiyang City, then analyzed its spatiotemporal change characteristics and driving factors(Dong et al., 2022). A multilayer perceptron (MLP) neural network was proposed to identify surface water with Landsat 8 satellite images, which was compared the extraction accuracy with water indices and support vector machines. The results found that the MLP method can achieve better performance compared with water indices and support vector machine(Jiang et al., 2018).

Random forest is a widely used classification algorithm for remote sensing imagery. Previous studies indicate that random forest exhibits significant advantages in classification accuracy, training time, and classifier stability across varying training samples and study areas(Pal et al., 2010). Currently, the extraction of detailed temporal changes in urban lakes is challenging for high spatial resolution remote sensing imagery due to the small size of urban lakes and the complexity of the urban environment. To explore the change pattern of urban lakes using remote sensing data, this study aims to propose a new method by combining time series Gaofen-1 satellite remote

*Corresponding author

sensing data from 2013 to 2020 and random forest machine learning algorithm. And then, the Zhushan Lake, located in Wuhan city is taken as an experimental area to conduct the quantitative analysis of the water areas change pattern.

2. STUDY AREA AND MATERIALS

2.1 Study Area

Wuhan is characterized by its intricate network of rivers, ports, and a myriad of lakes and ponds, earning it the moniker "City of Hundred Lakes" with a record of having 166 lakes. One such lake is Zhushan Lake (Figure 1), which is located in the eastern part of Caidian District, and is among the 43 urban lakes in Wuhan. Zhushan Lake, a crucial water source for Wuhan, not only supplies drinking water to its residents but also plays an indispensable role in agricultural irrigation and in industrial water usage. Moreover, the lake holds significant ecological and tourist value. The surrounding ecosystem offers recreational and natural sightseeing opportunities for local residents, thereby contributing to the city’s ecological balance. However, in recent years, the rapid urbanization process and human activity have impacted the extent of Zhushan Lake.

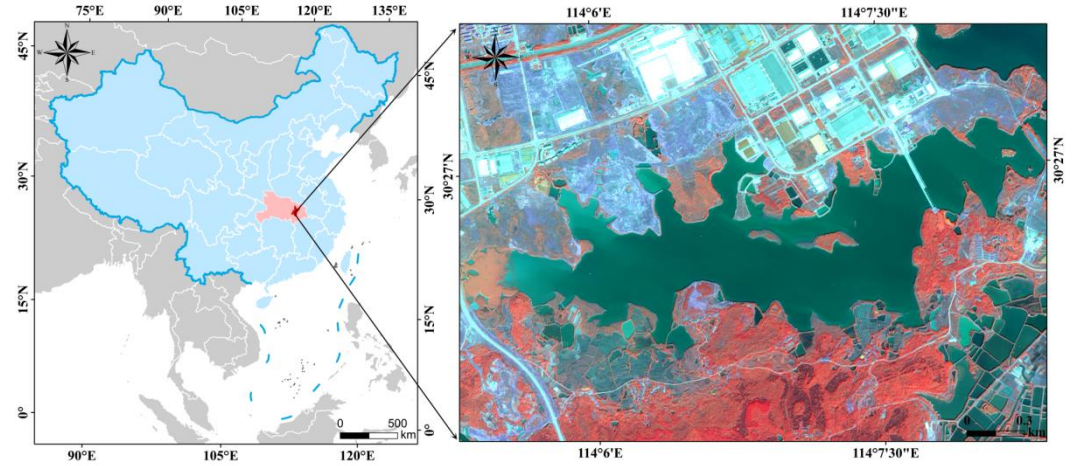

Figure 1. The location of study area

2.2 Data collection and preprocessing

This study collected time series Gaofen-1(GF-1) remote sensing data from 2013 to 2020 (Table 1). To ensure the comparability of lake extraction results across different periods, remote sensing images were preferably selected during the wet season (May to September), with the temporal phase of the remote sensing data kept as consistent as possible. For certain data sets, cloud cover can impact the precision of water body information extraction, which prompted the selection of data from adjacent months. Furthermore, the remote sensing image data underwent preprocessing steps, which included geometric correction, image fusion and image clip, thus achieving an experiment data with 2m spatial resolution.

Year	Sensor Model	Date of Acquisition	Scene Sequence ID
2013	GF-1_PMS2	2013/5/1	5354
2014	GF-1_PMS1	2014/7/22	399402
2015	GF-1_PMS1	2015/3/25	1035471
2016	GF-1_PMS1	2016/7/25	2641048
2017	GF-1_PMS1	2017/2/19	3350254
2018	GF-1_PMS1	2018/5/16	5031736
2019	GF-1_PMS2	2019/12/3	7154181
2020	GF-1_PMS2	2020/8/1	7995236

Table 1. The metadata of the time series GF-1 remote sensing data

3. METHODOLOGY

3.1 Research method flow

The flowchart of this study is shown in Figure 2, this flowchart contains four steps. The first part involves the collection of time series GF-1 image data and the subsequent preprocessing of these data. In the second part, training samples and feature bands are selected to construct a random forest model for extracting of lake water areas. The third part is dedicated to the verification of the accuracy of the lake water area extraction results. In the final part, indicators for evaluating lake changes are proposed to explore the change pattern and factors of urban lakes.

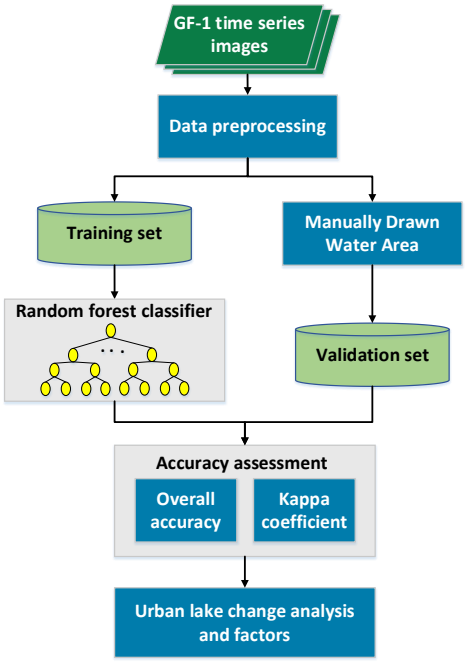

Figure 2. Flow chart of research method

3.2 Random Forest Classification Algorithm

Random forest is an ensemble learning method, composed of multiple decision trees. Each decision tree is trained on a random subset to achieve optimal classification results(Amit et al., 2001). The classification principles are as follows:

For the training sample set X, which contains n samples, each sample has M features. Each sample corresponds to a target set Y ,which is divided into k classes. In the random forest classification algorithm, it is usually divided into two classes.Therefore, we have the training sample set $X = \{x_1, x_2, \dots, x_n\}$, and the target set $Y = \{y_1, y_2, \dots, y_n\}$. From the training sample set X, n' training samples are randomly drawn using the bootstrap sampling method (where $n' < n$) to form a new training subset. For each new training subset, m features are randomly selected ($m < M$), and the optimal feature is selected from these m features for splitting. The mth feature x^m in the training subset D is sorted in ascending order, denoted as $\{A_1, A_2, \dots, A_m\}$, to obtain the set T of $m-1$ split points of the feature x^m, where $\{A_1, A_2, \dots, A_m\}$

$$T_i = \left\{ \frac{A_i + A_{i+1}}{2} \middle| 1 \le i \le m-1 \right\} \quad (1)$$

where A_i = i th training sample in ascending order feature value

T_i = i th split point of the feature x^m

Select the m th feature x^m and the i th split point T_i, resulting in two feature space subsets in the training subset D that are divided:

$$D_1(m,t) = \{x | x^m \le T_i\}, D_2(m,t) = \{x | x^m > T_i\} \quad (2)$$

Traverse all the features in the training subset D, and use the calculation of the Gini coefficient to find the optimal split feature x^m and the optimal split point T to construct the decision tree:

$$Gini(D,m) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (3)$$

$$Gini(D) = 1 - \sum_{k=1}^{2} p_k^2 \quad (4)$$

$$P_k = \frac{|C_k|}{|D|} \quad (5)$$

where $|D_1|$, $|D_2|$= number of samples corresponding to the feature space subsets D_1 and D_2.

$|D|$= number of samples in the training sample subset D.

$Gini(D_1)$, $Gini(D_2)$ are the Gini coefficient of the feature space subsets D_1 and D_2 respectively.

p_k = proportion of samples in the dataset that belong to category k.

C_k = subset of samples in the training subset D that belong to the k class.

Repeat the above steps until the predetermined number of decision trees are constructed. When the numerous decision trees of the random forest are constructed, the input to be classified is given, and all the decision trees in the random forest will give a classification result (water body and non-water body) respectively, and vote. The classification result with the most votes will be output as the final classification result

3.3 Accuracy Assessment

In order to validate the precision of the water body extraction performed by the random forest classifier, this study employed a confusion matrix to compute accuracy evaluation indicators. These indicators included the Overall Accuracy (OA) and the Kappa Coefficient are served to assess the accuracy of the extracted water body result.

(1) Overall accuracy (OA)

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

where TP = number of pixels that are actually water body pixels and are also detected as water body pixels.

TN = number of pixels that are actually non-water body pixels and are also detected as non-water body pixels.

FP = number of pixels that are actually non-water body pixels but are detected as water body pixels.

FN = number of pixels that are actually water body pixels but are detected as non-water body pixels.

(2) Kappa coefficient

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (7)$$

$$p_e = \frac{(TP + FN) \times (TP + FP) + (TN + FN) \times (TN + FP)}{n^2} \quad (8)$$

where p_o = overall accuracy OA

n = total number of samples

This study annually selected 500 random sample pixels from the GF-1 imagery from 2013 to 2020. Then, the confusion matrix was established by manual visual interpretation to compute accuracy evaluation indicators.

3.4 Urban lake change index

Four key indicators were selected including lake area, shoreline length, lake shape index and lake fragmentation(Xie et al., 2018). These indicators were utilized to quantitatively analyze the change pattern in this study.

(1) Lake Shape Index (LSI)

$$LSI = \frac{L}{4\sqrt{S}} \quad (9)$$

where L = shoreline length
S = lake area

The Lake shape index serves as a quantitative measure for evaluating the influence of human activities on lake landscapes. A lower LSI value signifies a lake with a relatively simplistic geometric configuration, indicating a heightened susceptibility to the impacts of external activities. Conversely, a higher LSI value represents a lake with a more complex geometric configuration, suggesting a reduced level of human interference.

(2) Lake Fragmentation (LF)

$$LF = \frac{N}{S} \quad (10)$$

where N = Lake number

Lake fragmentation quantifies the degree of fragmentation in a lake, encapsulating the complexity of the lake’s spatial structure.

4. RESULTS

4.1 Extraction result and accuracy assessment

Figure 3 shows the results of lake surface water extraction performance using the random forest classifier and time series GF-1 imagery. Figure 4 employed manual visual interpretation of samples to compute the Overall Accuracy (OA) and Kappa Coefficient. These indictors served to validate the precision of the water body extraction conducted by the random forest algorithm.

Figure 4 reveal that the accuracy of water body extraction using the random forest classification method consistently exceeded 96%. The Kappa coefficient ranged from a minimum of 0.86 to a maximum of 0.99. A comprehensive assessment of these two accuracy evaluation indictors indicated that the random forest method demonstrates good performance in water body extraction of urban lakes.

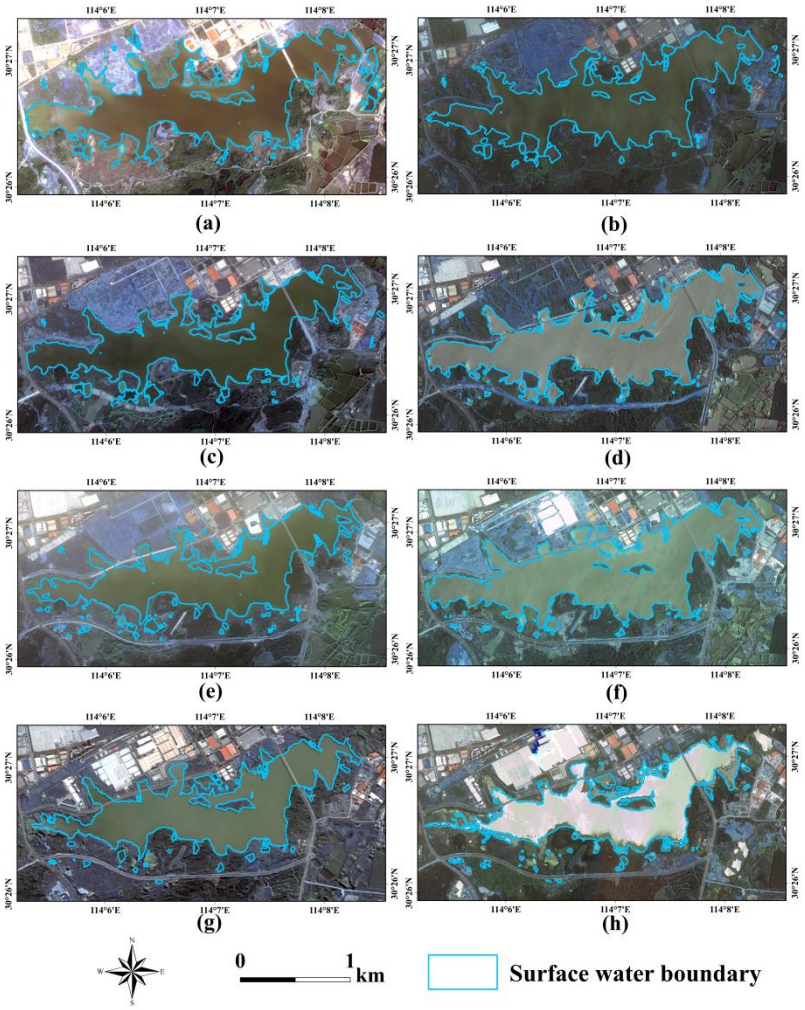

Figure 3. Zhushan Lake surface water extraction ((a),(b),(c),(d),(e),(f),(j) and (h) represent the annual extraction result from 2013 to 2020,respectively).

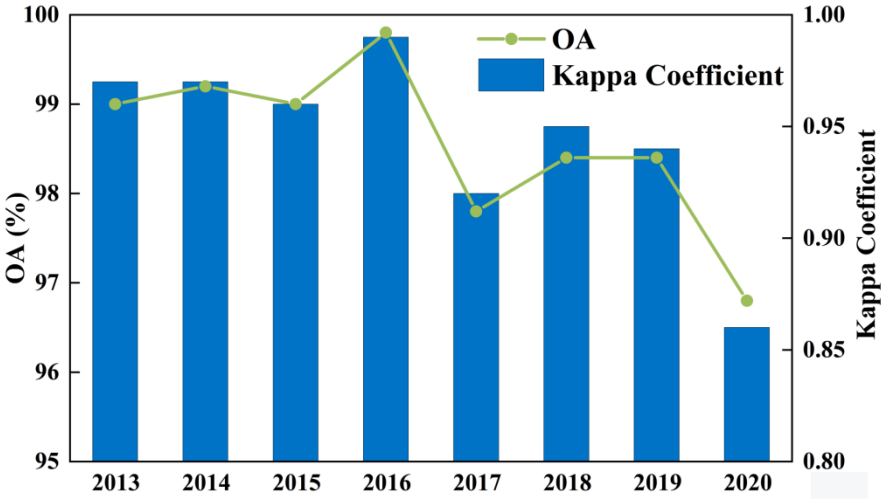

Figure 4. Accuracy verification of surface water extraction

4.2 Urban lake change analysis

Table 2 presents the values of four indicators of change-shoreline length, lake area, lake shape index, and lake fragmentation-for Zhushan Lake from 2013 to 2020. These results reveal a noticeable decline in both shoreline length and lake area since 2013. By 2020, the water area of Zhushan Lake and its surrounding pit-ponds has decreased by 1.11 km^2 compared to 2013, and the shoreline length of the ponds had shortened by 7.63 km. This corresponds to a shrinkage rate of 28.73% for the water area and a shortening rate of 16.4% for the shoreline. The lake shape index is higher in 2020 than that of in 2013, suggesting that the lake was more severely impacted by external activities in 2020. Conversely, the lake fragmentation was greater in 2013 than in 2020, indicating a relatively better biodiversity in 2013.

Observations of water extraction results from 2013 to 2020 reveal a significant reduction in the area of lakes along the northwestern shores and ponds in the eastern part of the study area. In the northwestern region, lake infilling began from 2014. Starting in 2018, construction of a factory commenced, and by 2019, the construction of this factory was completed. In the eastern region, infilling of pit-ponds started from 2013. From 2015 to 2018, construction began on the infilled the lake area, and this project was essentially completed until 2018. Consequently, it can be concluded that the water area changes of Zhushan Lake and its surrounding pit-ponds are primarily due to anthropogenic infilling of lakes and ponds. The infilled areas are mainly used for building construction.

Year	Shoreline Length (km)	Lake Area (km^2)	Lake Shape Index	Lake Fragmentation
2013	46.53	3.87	5.91	23.75
2014	36.34	3.66	4.75	11.21
2015	38.87	3.42	5.25	15.78
2016	38.47	3.66	5.02	16.10
2017	38.84	3.13	5.48	27.11
2018	34.79	3.76	4.49	12.77
2019	33.99	3.12	4.81	12.82
2020	38.90	2.76	5.85	22.82

Table 2. Urban lake change index

5. CONCLUSION

This study, focusing on Zhushan Lake and the surrounding pit-ponds in Wuhan City, utilized time series GF-1 remote sensing imagery and random forest algorithm to extract water bodies of Zhushan Lake. The following conclusions are obtained:

(1) The application of the random forest algorithm for water body extraction achieved high extraction precision. The OA exceeded 96% and the overall Kappa coefficient was 0.95.

(2) From 2013 to 2020, the water body area of Zhushan Lake and its surrounding ponds decrease by 1.11 km^2, and the shoreline length reduced by 7.63 km. The shrinkage rate of the water body area is 28.73%, and the shortening rate of the pond shoreline is 16.4%.

(3) Through the study of key changing areas, it was found that Zhushan Lake and its surrounding pit-ponds are mainly impact by building construction.

These results of this study demonstrated the feasibility and accuracy of using the random forest algorithm for water body extraction, providing a methodological reference for exploring the factors influencing changes in river and lake water environments.

ACKNOWLEDGEMENTS

This research was financially supported by the National Natural Science Foundation of China (42301450), the National Key Research and Development Program of China (2022YFC3301600 and 2023YFE0110300) and Scientific and Technological Innovation Talent Project of the Disaster Reduction Center of IWHR (WH0166B012023).

REFERENCES

Amit, Y. & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation*, 9, 1545 – 1588.

Brock, C., Ebel, C., & Patoine, A. (2006). Landscape-scale effects of urban nitrogen on a chain of freshwater lakes in central North America. *Limnology and Oceanography*, 51, 2262-2277.

Dong, F., Zhang, Y., Huang, D., et al. (2022). Analysis of temporal and spatial changes of surface water in karst mountain area - Taking Guiyang City as an example. *Henan Science*, 41(13).

Gossling, S., Peeters, P., Hall, C.M., Ceron, J.P., Dubois, G., Lehmann, L.V., & Scott, D. (2012). Tourism and water use: supply, demand, and security. An international review. *Tourism Management*, 33, 1-15.

Hayashi, S., Murakami, S., Xu, K., & Watanabe, M. (2008). Effect of the three Gorges dam project on flood control in the Dongting Lake area, China, in a 1998-type flood. *Journal of Hydro Environment Research*, 2, 148-163.

He, G.,Wang, G.,Long, T.,Peng, Y.,Jing, W.,Yin, R.,Jiao, W.,Zhang, Z. (2018)Opening and Sharing of Big Earth Observation Data: Challenges and Countermeasures[J].*Bulletin of Chinese Academy of Sciences*,2018,33(8):783-790.

Jiang, W., He, G., Long, T., Ni, Y., Liu, H., Peng, Y., Lv, K., & Wang, G. (2018). Multilayer Perceptron Neural Network for Surface Water Extraction in Landsat 8 OLI Satellite Images. *Remote Sensing*, 10(5), 755.

Jiang, W., He, G., Long, T., Ni, Y. (2018)."Detecting water bodies in landsat8 oli image using deep learning.Proceedings of the ISPRS Technical Commission III Midterm Symposium on "*Developments,Technologies and Applications in Remote Sensing*".Ed.. , 2018, 896-899.

Jiang, W., Ni, Y., Pang, Z., Li, X., Ju, H., He, G., Lv, J., Yang, K., Fu, J., & Qin, X. (2021). An Effective Water Body Extraction Method with New Water Index for Sentinel-2 Imagery. *Water*, 13(12), 1647.

Liu, T., Liu, R., & Ge, Q. (2013). Monitoring of Lake Chad surface water change based on multi-source remote sensing data. *Progress in Geography*, 32(06).

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217-22

Shen, C., Gan, S., Li, X., et al. (2021). Research on water body extraction method based on GF-5 hyperspectral feature analysis. *Global Positioning System*, 46(05).

Song, C., Huang, B., & Ke, L. (2013). Modeling and analysis of lake water storage changes on the Tibetan Plateau using multi-mission satellite data. *Remote Sensing of Environment*, 135, 25-35.

Xie, C., Huang, X., Wang, L., Fang, X., & Liao, W. (2018). Spatiotemporal change patterns of urban lakes in China’s major cities between 1990 and 2015. *International Journal of Digital Earth*, 11(11), 1085-1102.

Xu, K., Kong, C., Liu, G., Wu, C., Deng, H., Zhang, Y., & Zhuang, Q. (2010). Changes of urban wetlands in Wuhan, China, from 1987 to 2005. *Progress in Physical Geography: Earth and Environment*, 34(2), 207-220.

Yan, D., Schneider, U.A., Schmid, E., Huang, H., Pan, L., Dilly, O. (2013). Interactions between land use change, regional development, and climate change in the Poyang Lake district from 1985 to 2035. *Agricultural Systems*, 2013 (119), 10-21.

Yang, B.H., Meng, F., Ke, X.L., & Ma, C.X. (2015). The impact analysis of water body landscape pattern on urban heat island: a case study of Wuhan City. *Advances in Meteorology*, 2015.

Zhang, W., Cui, C., Li, L., et al. (2019). Monitoring and analysis of Poyang Lake surface area based on long time series remote sensing data. *Hydrology*, 39(03).

Zhang, Z., Liu, F., Zhao, X., et al. (2018). Urban Expansion in China Based on Remote Sensing Technology: A Review. *Chinese Geographical Science*, 28, 727 – 743.

Zhou, P., Xie, Y., Jiang, G., et al. (2020). Research progress on water body information extraction from remote sensing images. *Remote Sensing Information*, 35(05).