

## A NEW INDEX FOR IDENTIFYING WATER BODY FROM SENTINEL-2 SATELLITE REMOTE SENSING IMAGERY

Wei Jiang<sup>1,2</sup>, Yuan Ni<sup>3</sup>, Zhiguo pang<sup>1,2\*</sup>, Guojin He<sup>4</sup>, June Fu<sup>1,2</sup>, Jingxuan Lu<sup>1,2</sup>, Kun Yang<sup>1,2</sup>, Tengfei Long<sup>4</sup>, Tianjie Lei<sup>1,2</sup>

<sup>1</sup> State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, China-jiangwei@iwhr.com, pangzg@iwhr.com,

<sup>2</sup> Remote Sensing Technology Application Center, Research Center of Flood and Drought Disaster Reduction of the Ministry of Water Resources, Beijing, China-fuje@iwhr.com, lujx@iwhr.com, yangkun@iwhr.com, leitj@iwhr.com;

<sup>3</sup> Sichuan Highway Planning, survey, Design and Research Institute Ltd, Sichuan, China, niyuan\_map@163.com

<sup>4</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China-hegj@radi.ac.cn, longtf@radi.ac.cn;

### Commission III, WG III/1

**KEY WORDS:** Sentinel-2, water body extraction, water index, information extraction, Otsu method, surface water mapping

### ABSTRACT:

Water body plays an irreplaceable role in the global ecosystem and climate system. Sentinel-2 is a new satellite data with higher spatial and spectral resolution. Through analysing spectral characteristics of Sentinel-2 satellite imagery, the brightness of water body in vegetation red edge band and shortwave infrared band shows sharply different than that of the not water body. Therefore, a new type of water index SWI (Sentinel-2 Water Index) was proposed by combining those two bands. Four representative water types, which included Taihu Lake, the Yangtze River Estuary, the ChaKa Salt Lake and the Chain Lake, were selected as experimental areas. Normalized difference water index (NDWI) and Sentinel-2 Water Index (SWI) with Otsu method were employed to extract water body. The results showed that overall accuracy and Kappa coefficient of SWI were higher than that of NDWI and SWI was efficient index to rapidly and accurately extract water for Sentinel-2 data. Therefore, SWI had application potential for larger scale water mapping in the future.

### 1. INTRODUCTION

As an important part of global water cycle, the surface water body has an irreplaceable role in global ecology and climate system (Donchyts et al., 2016; Jiang et al., 2020; Pekel et al., 2016). Surface water body generally contains river, lake and reservoir with obvious seasonal variations (Huang et al., 2018). Therefore, the rapid and accurate monitoring of surface water is of great significance to hydrological process and water resources management (Donchyts et al., 2016). Remote sensing technology has advantages such as large scale, dynamic, real-time and low cost, which has been widely used in regional and global scale surface water mapping (Feng et al., 2016; Ji et al., 2018; Jiang et al., 2018). Considering the requirement of precision and the complexity of the method, the water index method is one of the most convenient and relatively high precision water extraction method (Xu, 2006).

McFeeters made full use of the near infrared and visible light green bands of Landsat MSS data, and proposed a normalized difference water index (NDWI) (McFeeters, 1996). The index can effectively highlight the water body information and eliminate the soil and surface vegetation noise. However, the NDWI will leak out the construction land (Xu, 2006). Xu constructed a modified normalized difference water index (MNDWI) based on Landsat TM/ETM+ data, which not only can overcome the shortcomings of NDWI index, but also enhance water environment characteristic information (Xu, 2006). Feyisa et al proposed a multi band combined water index—AWEI (Automated Water Extraction Index) by testing Landsat TM data (Feyisa et al., 2014). The index can further solve the mountain shadow caused by terrain that the MNDWI index failed to eliminate. Although these indices have good effects on water extraction from Landsat series satellites.

However, as to different sensors, a slight change in spectral range of corresponding wavebands may lead to the change of water extraction results (Xu, 2006).

Sentinel-2 data is a new satellite data launched by ESA, it can obtain global satellite data, had been successfully applied in the field of agricultural statistics (Immitzer et al., 2016), land use, forest monitoring (Drusch et al., 2012). As to the water body extraction, most methods aimed to extract water body with existing water body index (Du et al., 2016; Huang et al., 2016) and the sensitivity of unique band of the Sentinel-2 series satellite to the water body had not been taken considered. Therefore, this study aimed to propose a new water index for Sentinel-2 satellite imagery by analysing spectral characteristics. The experiments between the new index and NDWI index across various water qualities (freshwater lakes, coastal waters, Saline Lake, and ice) were compared and the performance of water body extraction were analysed. This proposed useful water index with Sentinel-2 can be used in large scale surface water mapping, which will be helpful to understand the surface water dynamic change.

### 2. STUDY AREA AND MATERIALS

Sentinel-2A and 2B satellites were launched at the Kuru space launch centre in June 23, 2015 and March 7, 2017, respectively. It is composed of two satellites constellation, running in the sun synchronous orbit, the orbit height is 786 km and the angle is 98.5 degree and the revisit period is 5 days. The payload has 13 channels, the spectral range is from 0.4  $\mu\text{m}$  to 2.4  $\mu\text{m}$ , and the spatial resolution is 10 m, 20 m and 60 m, as shown in Table 1. The width of imaging is 290 km, and its global land surface imaging data can be updated every 10 day (Drusch et al., 2012). The Sentinel-2 data is the L1C atmospheric top reflectivity

\* Corresponding author

product, which has been processed by geometric correction. The experiment data were download from the website: <https://scihub.copernicus.eu/dhus/#/home>.

Band number	Wavelength range / $\mu\text{m}$	Spatial resolution /m
1	0.433-0.453	60
2	0.4575-0.5225	10
3	0.5425-0.5775	10
4	0.65-0.68	10
5	0.6975-0.7125	20
6	0.7325-0.7475	20
7	0.773-0.793	20
8	0.7845-0.8995	10
8A	0.855-0.875	20
9	0.935-0.955	60
10	1.36-1.39	60
11	1.565-1.655	20
12	2.1-2.28	20

Table 1. Band metadata information of the Sentinel-2 satellite

To analyse the performance of new water index in extracting water body, Taihu lake, the Yangtze river entrance, ChaKa saline lake and the Chain lake were selected as experimental areas. These areas included river, lake, ocean and other conventional water bodies, as well as special water bodies containing sediment, salt and ice. The vegetation, buildings and bare land were included in experiment areas.

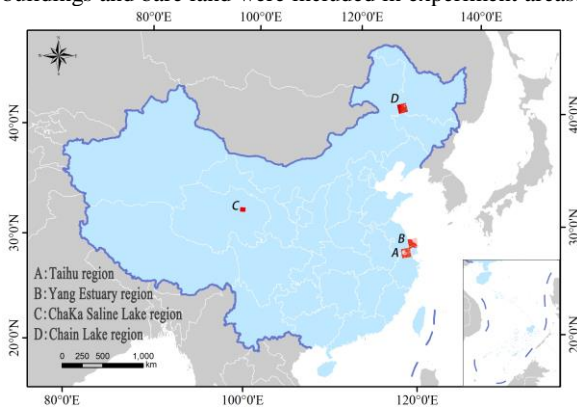


Figure 1. The location of four experimental areas

Figure 1 showed the specific location of the experimental area, where A is the Taihu and the surrounding area. The image of the February 28, 2017 is Sentinel-2A data. The water body in the image covers rivers and lakes, and involves the extraction of urban water bodies and small water bodies. The surrounding background includes vegetation and buildings. B is the Yangtze River entrance and its surrounding area, this image is the Sentinel-2A data acquired on February 28, 2017. The water body in the image covers the oceans and rivers, which involves the extraction of sediment polluted water, and the surrounding background contains vegetation and buildings. C is the ChaKa Saline Lake and its surrounding area, the image of the July 30, 2016 is Sentinel-2A data. A large amount of salt is contained in the water body, and the surrounding background is bare soil. The image is the Sentinel-2A data acquired on April 9, 2017. The water body contains ice that has not been completely melted, and the surrounding background includes bare land and vegetation.

### 3. METHODOLOGY

#### 3.1 Analysis of spectral characteristics

Figure 2 is the spectral characteristic of different land cover type in experimental area A. As to water bodies, visible light is widely transmitted in the water body, and the incident energy of water absorption increased with the increase of wavelength. The reflectance in short wave infrared band low, even close to 0.

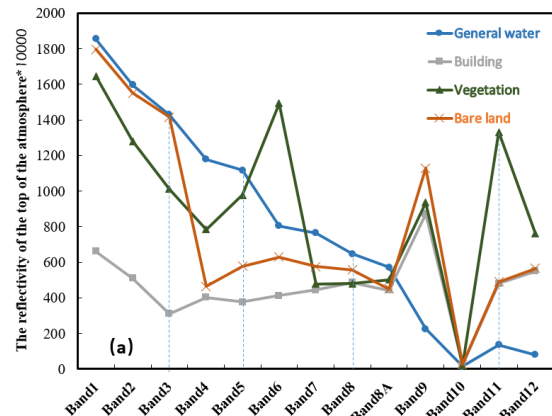


Figure 2. The spectral characteristics of different land cover type from Sentinel-2 satellite imagery

By using the reflectance difference of water body in different bands, the brightness value of water body can be enhanced and the background noise can be suppressed. According to Sentinel-2 data, the spatial resolution of shortwave infrared band is 20m and the spatial resolution in the infrared and green bands are 10m. Therefore, NDWI index and MNDWI index can be directly calculated due to the different band resolution.

According to the band spatial resolution and spectral characteristics, a water index for the Sentinel-2 data in 20m resolution was constructed and the formula was as follows:

$$SWI = \frac{\rho_{vre} - \rho_{SWIR}}{\rho_{vre} + \rho_{SWIR}} \quad (1)$$

In the formula,  $\rho_{vre}$  represents the red edge of the vegetation (corresponding to fifth band of the Sentinel-2), and  $\rho_{SWIR}$  represents the short wave infrared band (corresponding to eleventh band of the Sentinel-2).

#### 3.2 Otsu threshold method

Otsu method is an efficient image binarization algorithm, it uses the maximum variance between classes to distinguish objects and background(Otsu, 1979). When the value of the target water is quite different from that of the background, the water body and the not water body are easily distinguished(Otsu, 1979).

The principle of the algorithm is: it is assumed that the classification of the image I of the gray value of  $\{1,2 \dots M\}$  and the image is divided into the threshold

$C_0 = \{1, 2 \dots T\}$  and  $C_1 = \{T + 1, T + 2 \dots M\}$ .

Then the probability of each gray level  $I$  is as follows:

$$\rho_i = \frac{N_i}{N} (\rho_i \geq 0) \quad (2)$$

Among them,  $N = n_1 + n_2 + \dots + n_M$  is the total number of image pixels, and  $n_i$  is the number of pixels corresponding to the gray level. The probability of the pixels number of  $C_0$  and  $C_1$  in the entire image is  $W_0$ ,  $W_1$ , and the average gray value is  $\mu_0$ ,  $\mu_1$ , respectively:

$$W_0 = \sum_{i=1}^T \rho_i = \omega(t) \quad (3)$$

$$W_1 = \sum_{i=T+1}^M \rho_i = 1 - \omega(t) \quad (4)$$

$$\mu_0 = \sum_{i=1}^M \frac{i\rho_i}{W_0} \quad (5)$$

$$\mu_1 = \sum_{i=T+1}^M \frac{i\rho_i}{W_1} \quad (6)$$

The total average gray value of the whole image is  $U$ :

$$U = W_0 * \mu_0 + W_1 * \mu_1 \quad (7)$$

Inter class variance of  $C_0$  and  $C_1$ :

$$g = W_0(\mu_0 - U)^2 + W_1(\mu_1 - U)^2 \quad (8)$$

The maximum threshold  $T$  of the inter class variance  $g$  is the best threshold  $T_{good}$  in the experimental area:

$$T_{good} = \arg \max (g) \quad (9)$$

The water area was detected by the best threshold  $T_{good}$ , and the water area  $W$  was extracted by two values.

$$W = \begin{cases} 1, & I \geq T_{good} \\ 0, & I < T_{good} \end{cases} \quad (10)$$

## 4. RESULTS

### 4.1 Water index difference

To compare the performance of extracting water, the index accuracy of NDWI and SWI is analysed by the difference index between the water body and the background. The difference index of water and background is employed to analyse the accuracy of NDWI and SWI index, and the experimental area water and not water samples are choose to calculate the

corresponding mean. According to the difference between the water and not water, the separation of water and not water can be used to evaluate the validity of index.

Table 2 is the mean and contrast value of water and not water samples of NDWI and SWI in each experimental area. It can be seen that the difference between the mean value of NDWI and the mean value of SWI is not significant, but the mean value of SWI is much larger than that of NDWI. Therefore, the difference of water and not water of the SWI index is greater than the NDWI index, which indicate that the SWI index is better robust and accurate in the water extraction.

	NDWI			SWI		
	Water	Not water	Contrast value	Water	Not water	Contrast value
Taihu Lake	0.38	-0.20	0.58	0.78	-0.12	0.90
Yangtze River Estuary	0.22	-0.23	0.45	0.89	-0.09	0.98
ChaKa Saline Lake	0.13	-0.21	0.34	0.81	-0.17	0.98
Chain Lake	0.34	-0.26	0.60	0.84	-0.31	1.15

Table 2. Mean value and its contrast value of water and not water

Figure 3 is the statistical map of the maximum, minimum, mean and standard deviation of water and not water samples regarding to NDWI and SWI. A, B, C and D are the experimental areas of Taihu, Yangtze River Estuary, ChaKa Saline Lake, and Chain Lake. From the figure 3, we can see that NDWI and SWI can distinguish most of the water and not water area, however, the maximum or minimum value of the not water body of the NDWI index is almost intersecting with the standard deviation of water body, and the maximum or minimum value of the not water body of the SWI index is not intersecting with the standard deviation of the water body. Therefore, the extraction of water body using NDWI index is easy to produce a large number of misclassification. The SWI index can be extracted to an accurate water body.

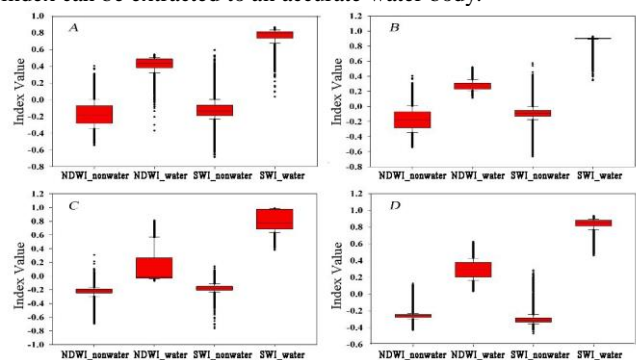


Figure 3. Comparison of the maximum, minimum, and standard deviations of the NDWI and SWI index (A is Taihu Lake, B is Yangtze River Estuary, C is ChaKa Saline Lake, D is Chain Lake.)

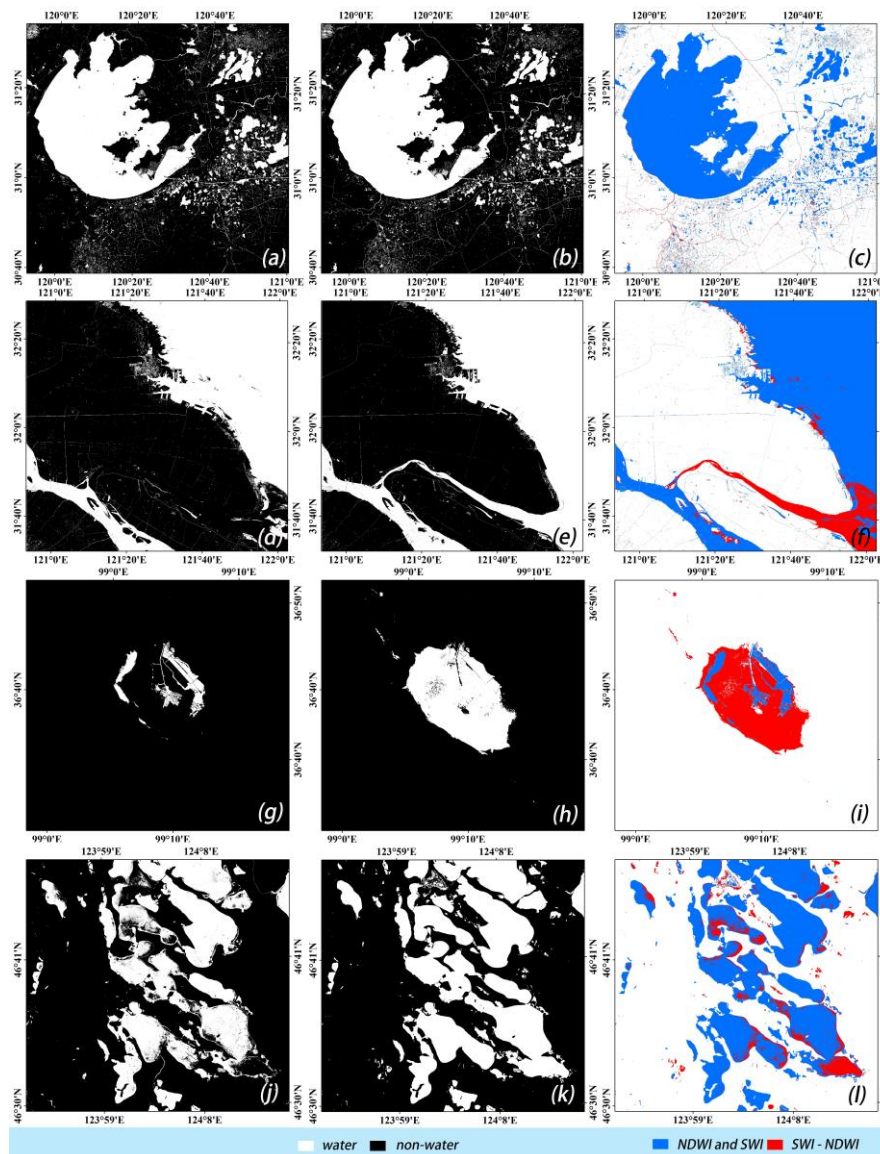


Figure 4. Water extraction of NDWI and SWI index and its difference ((a)(d)(g)(j))respectively represent water bodies extracted by NDWI in Taihu Lake、the Yangtze River Estuary、ChaKa Saline Lake、Chain Lake; (b)(e)(h)(k) respectively represent water bodies extracted by SWI in Taihu Lake、the Yangtze River Estuary、ChaKa Saline Lake、Chain Lake; (c)(f)(i)(l) respectively represent similarities and differences of water body extraction based on NDWI and SWI in Taihu Lake、the Yangtze River Estuary、ChaKa Saline Lake、Chain Lake).

		Mapping precision		User precision		Overall accuracy	Kappa coefficient
		Water	Not water	Water	Not water	Water	Not water
Taihu Lake	NDWI	95.80%	89.68%	79.72%	98.05%	91.50%	0.808
	SWI	99.11%	90.28%	79.86%	99.62%	92.75%	0.833
Yangtze River Estuary	NDWI	96.43%	81.63%	72.65%	97.83%	86.6%	0.722
	SWI	98.11%	89.63%	86.19%	98.63%	93%	0.868

Table 3. Overall accuracy for each classification result

## 4.2 Water body extraction performance

Otsu method can automatically extract water body by adaptive threshold. The performance of water extraction with Otsu method are shown in Figure 4. Figure4(a), Figure4 (d), Figure4 (g) and Figure4 (j) are corresponding to the water extraction results of the NDWI index in Taihu Lake, the Yangtze River entrance, the ChaKa Saline Lake and the Chain Lake, respectively. Figure4(b), Figure4 (e), Figure4 (h) and Figure4 (k) are corresponding to the water extraction results from the SWI index of four experiment areas, respectively. The blue region of four experiment areas in Figure4 (c), Figure4 (f), Figure4 (i) and Figure4 (l) correspond to the part of the water body extracted jointly by NDWI and SWI index, the red region corresponds to the part of the water body that the SWI index is more than the NDWI index.

Figure 4 (a), Figure 4 (b), Figure 4 (c) show that SWI and NDWI index can extract most of the water bodies, but for the extraction of large river water, the NDWI index obviously miss some water. This is due to the fact that the NDWI index has little difference between the water value and the not water value at these river area. Figure 4 (d), Figure 4 (e) and Figure 4 (f) show that the NDWI index extraction performance of water body with high sediment concentration and it is easy to mix the water body and building or bare soil, etc. While the SWI index still can accurately extract water in clean water or high sediment concentration. Figure 4 (g), Figure 4 (h) and Figure 4 (i) show that the NDWI index has a poor extraction performance for a water body with high salinity. While the performance of SWI extraction is better. Because the SWI index is sensitive to water body in the short wave infrared band, it can accurately extract a small amount of water with high humidity in the bare land. Figure 4 (j), Figure 4 (k) and Figure 4 (l) show that both the NDWI index and the SWI index can extract most of the water containing ice. Because NDWI has high reflectivity on the ice surface, it is easily confused with the high reflection of the bare ground, causing the leakage of ice water. The SWI index can accurately extract water with floating ice.

## 4.3 Accuracy assessment

The experimental area of ChaKa Saline Lake and Chain Lake is a comparison of extraction performance for special water bodies, and its surrounding surface types are single. Visual discrimination can better evaluate the extraction performance of water and not water. For ChaKa Saline Lake, SWI index not only can extract more complete salty water, but also can extract effectively a small amount of water from bare soil slope due to its sensitivity to water humidity. Therefore, the SWI index in experimentation area has better extraction performance. The NDWI index has a significant miss of salt water in the region. For Chain lakes, the smoothness and continuity degree of SWI index for extracting Lake Boundary is better than that of NDWI index.

For the experimental area of Taihu and the Yangtze River entrance, the type of ground objects is rich and the composition of the water is varied. Qualitative evaluation of water extraction performance obviously cannot meet the accuracy evaluation of water bodies, and quantitative accuracy evaluation of water extraction requires the selection of similar or more high-resolution optical imagery as reference images. Using the 10m resolution data of the same image of Sentinel-

2A data as the reference image, 400 test pixels in the experimental area are randomly selected to verify the man-machine interaction, and the results are shown in Table 3.

For the two experimental areas, the mapping accuracy index (user accuracy, overall accuracy and Kappa coefficient) of SWI index are higher than that of NDWI index. For Taihu experimentation area, the accuracy of NDWI and SWI index is not very different, which can accurately extract most of the water. For the experimentation area of Yangtze River Estuary, the accuracy of NDWI index is obviously lower than that of SWI index. This is because NDWI index discriminating the water body containing sediment into bare soil or building, which resulted that its accuracy is lower than the SWI index. Therefore, from the accuracy verification, the extraction performance of SWI index is better than the NDWI index.

## 5. CONCLUSION

By analysing the spectral response mechanism of Sentinel-2A data, a new water index -SWI index is proposed by using the vegetation red edge band (corresponding to fifth band of the Sentinel-2) and shortwave infrared band (corresponding to eleventh band of the Sentinel-2). Four typical experimental areas are selected to compare the extraction performance of SWI and NDWI index. The conclusion is as follows:

(1) SWI index has better performance to separate urban area, sediment, salt and ice from water body.

(2) The extraction accuracy of SWI index is higher than that of NDWI index. It shows that the SWI index can accurately extract water and not water.

Compared with the results of two water index, SWI is an efficient Sentinel-2 water extraction index, and it has the potential of automatic extraction in large scale water body.

## ACKNOWLEDGEMENTS

This research was financially supported by the National Natural Science Foundation of China (51779269), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA190090300), the National Key Research and Development Program of China—rapid production method for large-scale global change products (2016YFA0600302) and the Innovation Talent Project of Research Center of Flood and Drought Disaster Reduction of the Ministry of Water Resources.

## REFERENCES

- Donchyts, G., Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., Van De Giesen, N., 2016. Earth's surface water change over the past 30 years. *Nature Climate Change*. 6(9), 810-813.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*. 120,25-36.
- Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*. 8(4), 354.

Feng, M., Sexton, J.O., Channan, S., Townshend, J.R., 2016. A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic-spectral classification algorithm. *International Journal of Digital Earth*. 9(2), 113-133.

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*. 140,23-35.

Huang, C., Chen, Y., Zhang, S., Li, L., Shi, K., Liu, R., 2016. Surface water mapping from Suomi NPP-VIIRS imagery at 30 m resolution via blending with Landsat data. *Remote Sensing*. 8(8), 631.

Huang, C., Chen, Y., Zhang, S., Wu, J., 2018. Detecting, extracting, and monitoring surface water from space using optical sensors: A review. *Reviews of Geophysics*. 56(2), 333-360.

Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sensing*. 8(3), 166.

Ji, L., Gong, P., Wang, J., Shi, J., Zhu, Z., 2018. Construction of the 500 - m resolution daily global surface water change database (2001–2016). *Water Resources Research*. 54(12), 10,270-210,292.

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., Pang, Z., Guo, H., Long, T., Ni, Y., 2020. Surface water map of China for 2015 (SWMC-2015) derived from Landsat 8 satellite imagery. *Remote Sensing Letters*. 11(3), 265-273.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*. 17(7), 1425-1432.

Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*. 9(1), 62-66.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*. 540(7633), 418-422.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*. 27(14), 3025-3033.