Ensembles of multiple spectral water indices for improving surface water classification

Zhaofei Wen[a, b, *], Ce Zhang[c, d, *], Guofan Shao[b], Shengjun Wu[a], and Peter M. Atkinson[c]

[a] Key Laboratory of Reservoir Aquatic Environment, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China

[b] Department of Forestry and Natural Resources, Purdue University, West Lafayette 47906, USA

[c] Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

[d] UK Centre for Ecology & Hydrology, Library Avenue, Bailrigg, Lancaster LA1 4AP, UK

* Corresponding author

Email address: wenzhaofei@cigit.ac.cn (Zhaofei Wen) and c.zhang9@lancaster.ac.uk (Ce Zhang).

Postal address: No. 266, Fangzheng Avenue, Shuitu Hi-tech Industrial Park, Shuitu Town, Beibei District, Chongqing 400714, China.

Email Addresses: wenzhaofei@cigit.ac.cn (Zhaofei Wen), c.zhang9@lancaster.ac.uk (Ce Zhang), shao@purdue.edu (Guofan Shao), wsj@cigit.ac.cn (Shengjun Wu), and pma@lancaster.ac.uk (Peter M. Atkinson)

Abstract: Mapping surface water distribution and its dynamics over various environments with robust methods is essential for managing water resources and supporting water-related policy design. Thresholding Single Water Index image (TSWI) with a fixed threshold is a common way of using water index (WI) for mapping water for it is easy to use and could obtain acceptable accuracies in many applications. As more and more WIs are available and each has its distinct merits, the real-world application of TSWI, however, often face two practical concerns: (1) selection of an appropriate WI, and (2) determination of an optimal threshold for a given WI. These two issues are problematic for many users who rely either on trial-and-error procedures that are time-consuming or on their personal preferences that are somewhat subjective. To better deal with these two practical concerns, an alternative way of using WIs is suggested here by transforming the current paradigm into a simple but robust ensemble approach called Collaborative Decision-making with Water Indices (CDWI). A total of 145 subsite images (900 × 900 m) from 22 Landsat-8 OLI scenes that covering various water-land environments around the world were used to assess the performance of TSWI and the CDWI. Five benchmark WIs were adopted in five TSWI methods and CDWI method: Normalized Difference Water Index (NDWI), the Modified NDWI (MNDWI), the Automated Water Extraction Indices without considering (AWEI0) and with considering (AWEI1) shadows, and the state-of-the-art 2015 water index (WI2015). Two aspects of performance were analyzed: comparing their accuracies (indicated by both F1-scores and Youden's Index) over various environments and comparing their accuracy sensitivities to threshold. The results demonstrate that CDWI produced higher accuracies than the other five TSWI methods for most application cases. Particularly, more samples (indicated by percentage) produced higher F1-scores by CDWI than the other five TSWI methods, i.e. 67% (CDWI) vs. 15% (TSWINDWI), 54% (CDWI) vs. 22% (TSWIMNDWI), 42% (CDWI) vs. 12% (TSWIAWEI0), 57% (CDWI) vs. 17% (TSWIAWEI1), and 34% (CDWI) vs. 12% (TSWIWI2015). Moreover, the F1-score of the CDWI is much less sensitive to the change of thresholds compared with that of the other five TSWI methods. These important benefits of CDWI make it a robust approach for mapping water. The uncertainty of CDWI method was thoroughly discussed and a general guidance (or look-up-table) for selecting WIs was also suggested. The underlying framework of CDWI could be readily generalizable and applicable to other satellite sensor images, such as Landsat TM/ETM+, MODIS, and Sentinel-2 images.

Keywords: Water index, Threshold, Integrated decision making, Mixed pixels, MNDWI

1. Introduction

Inland water is an important earth resource for providing ecosystem services (Karpatne et al., 2016; Ogashawara et al., 2017), such as being a key habitat for flora and fauna of aquatic ecosystems and support biodiversity conservation (Vörösmarty et al., 2010). It is also a key component of Earth's hydrologic cycle and, as such, can support many aspects of daily life, including drinking water, agricultural irrigation, electricity production, and transportation (Huang et al., 2018). Spatially explicit monitoring of water changes is, therefore, essential for a variety of scientific disciplines and to inform land-use policy and decision-making (Berry et al., 2005; Ma et al., 2010; Pekel et al., 2016).

As remote sensing is well recognized for detecting spatiotemporal patterns of land cover, it has been widely used for monitoring water changes with various purposes, such as water resource inventory, flooding and drought assessment, and urban hydrological evaluation (Allen and Pavelsky 2018; Berry et al., 2005; Shao et al., 2019). Generally, the success of mapping water bodies with remote sensing images relies on the distinct reflectance spectra of water in comparison with other land features: water generally show lower reflectance and a decreasing pattern of reflectance from visible to infrared spectral wavelengths (Bukata et al., 2018). Based on such optical characteristics, various types of water classification methods have been developed which can be broadly grouped into indirect and direct strategies.

The indirect strategy considers water bodies as one of several broad land cover categories, and the water bodies can be extracted from a land use/land cover map derived from image classification methods, such as deep learning, random forest, support vector machine (Cao et al., 2019). The direct classification strategy is to classify an image into water and non-water (land) categories directly. It is easy to use and widely adopted in practice (Allen and Pavelsky 2018; Berry et al. 2005; Cooley et al. 2017; Guo et al. 2017). One of the most common approaches is called Thresholding Single Water Index (TSWI), in which the water index (WI) is derived from two or more spectral bands with a carefully designed algorithm and water pixels would gain high values and the non-water pixels would gain low values (Ji et al., 2009). In the processing of TSWI, selecting a WI and generating corresponding WI image should be done first, and then pixels in such WI image with their values higher than (or lower than in some cases) a predefined appropriate threshold are categorized as water, otherwise non-water (Huang et al., 2018).

As WIs are sensor dependent, only the WIs designed for Landsat images are focused on this research. The Normalized Difference Water Index (NDWI; McFeeters 1996), is considered as the first-generation WI for using TSWI to classify water. It is calculated using the green and near-infrared (NIR) bands of Landsat TM with an equation similar to NDVI which is used for vegetation (Tucker 1979), and the threshold 0 is suggested for thresholding water areas. NDWI was the most widely used index (McFeeters, 2013) before the Modified Normalized Difference Water Index (MNDWI) was introduced by Xu (2006). MNDWI was designed because using NDWI with TSWI cannot efficiently suppress the signal from built-up areas, such that the suggested threshold 0 fails to distinguish water bodies from built-up surfaces accurately. The equation of MDNWI is similar to NDWI, but the NIR band is replaced by the first shortwave infrared (SWIR1) band of Landsat TM imagery. MNDWI is the most widely used WI for a variety of applications, including surface water mapping, land use/cover change analyses, and ecological monitoring research (Allen and Pavelsky 2018; Ji et al., 2009). In certain situations, however, the performance of MNDWI may be relatively poor due to the presence of low reflectance surfaces such as asphalt roads and shadow effects. To overcome such issues, Feyisa et al. (2014) proposed two new WIs, Automated Water Extraction Index with (AWEI1) and without (AWEI0) considering shadows. AWEI0 and AWEI1 are considered highly useful WIs and have been applied with TSWI to extract water bodies from Landsat imagery (Huang et al., 2018; Jiang et al., 2014). Fisher et al. (2016) conducted a comprehensive inter-comparison of the existing WIs and designed the latest water index (WI2015). The WI2015 is derived from linear discriminant analysis and involves all the bands of Landsat TM/ETM+ except for the blue band and it has demonstrated similar accuracy to some of the prevailing WIs.

The driving force behind proposing different WIs indicates the fact that water-land environments in the real-world are very heterogeneous and the stability of applying TSWI with any single WI would vary a lot over different environments (Wu et al., 2018, Yang et al., 2018). Therefore, an average user of TSWI would face two basic concerns: (1) *which WI* should be chosen from existed WIs, and (2) what is the *appropriate threshold* that should be used for a given WI?

In general, the answer to the first concern involves some personal preference because there is no clear guidance of WI selection and a WI performs unsteadily over different water-land environments, such as wetland, mountain, urban, forest, and desert (Fisher et al., 2016; Ji et al., 2009). As a consequence, the same image classified by different TSWI users could produce inconsistent results due to different choices of WIs and the corresponding thresholds (Feyisa et al., 2014; Huang et al., 2018). For the second concern, three types of thresholds have been reported according to the availability of ground reference data, i.e., the real outline of water bodies that were obtained at the same time as the image acquisition time. **Case 1**: If enough reference data is available in an application, the local optimal threshold is suggested because such threshold can be determined (or trained) by the reference data. In most average applications, however, the obtaining of timely reference data could be differcult, especially for highly dynamic water landscapes (e.g., rivers and wetlands during flood events). **Case 2**: If there is no reference data, the locally-adaptive threshold and pre-defined threshold could be the choices. The locally-adaptive threshold is determined by the WI image itself with some segmentation technologies, so that the thresholds can vary self-adaptively for different images (Huang et al., 2018; Li and Sheng 2012; Wen et al., 2020). One obvious shortcoming of locally-adaptive threshold is that it heavily depends on the applied image extent and its land/water ratio, such that threshold can be vastly different for the same location when it is determined from different extents (Zhang et al., 2018). The pre-defined thresholds are often recommended by the original WI inventors or by other experienced authorities. To the best of our knowledge, the pre-defined thresholds are widely used in average water mapping applications for they are super easy to be applied. However, this type of thresholds should be used with caution because they cannot guarantee satisfying results due to the complex water-land environments in the real world (Feyisa et al., 2014; Fisher et al., 2016).

In summary, the application of TSWI faces two common concerns as mentioned above and the ways to deal with them are unsatisfied if there is no sufficient reference data. Thus, alternative solutions have been explored over the past few years (Huang et al., 2018), including the construction of new WIs that are robust and relatively insensitive to threhsold selection or the development new methods using mutliple existing WIs (Sánchez et al., 2018; Wang et al., 2018). The latter is regarded as the most appropriate approach because the combination of multiple WIs could complement their merits and apply to different environments compared with TSWI method (Yang et al., 2015). Such strategy is, to some extent, in line with the collaborative decision-making theroy where multiple variables can produce complementary information to support a more robust result than each individual variable (Kacprzyk and Fedrizzi 2012).

Inspired by these ideas, this research aims to propose a new way of using WIs based on collaborative decision-making theroy to deal with the two concerns mentioned above that exist commonly in TSWI method. Such new approach has the advantages of: (1) less concerned about the WIs selection and (2) less sensitive to WIs thresholds than TSWI method. Specifically, the new approach is transforming the current paradigm of using WIs (i.e., TSWI method) into a simple but highly robust ensemble way of using WIs called Collaborative Decision-making with Water Indices (CDWI). The CDWI (the new way of using WIs) was tested in a variety of water-land environments around the world and assessed by comparing its performances with that of TSWI (the common way of using WIs) using five benchmarked WIs.

2. Test sites and data materials

2.1. Test sites and subsites

Performances of water classification methods are generally affected by two error sources: the applied aquatic environments and their surrounding land features(Wu et al., 2018, Yang et al., 2018). The aquatic environments are often characterized by a variety of watercolors (e.g., dark, yellow, red, and brown, etc.) and water types (e.g., river, reservoir, pond, and ditch, etc.). The surrounding land features are usually recognized as vegetation conditions (high-density vegetation, sparse vegetation, etc.), built-up area (road and buildings), and shadows (cloud shadow, building shadows, and terrain shadows). The combinations of these two error sources make the selection of test sites tricky and time-consuming. Fortunately, many test sites have already been used for validating water classification methods in previous studies and such sites can guide us for selecting test sites in this study. Finally, 22 test sites were carefully selected with some come from Yang et al. (2015) and Feyisa et al. (2014) and some newly selected by considering their spatial representativeness (Fig. 1). These sites scattered around the world and covered a variety of water-land environments (Table 1).

Among each test site, several subsites with 900 × 900 m square size each were selected for preparing test data (as exampled in Figs. 1b, 1c, and 1d). The subsites were mannually selected with expert knowledge in true color composite Landsat-8 OIL images (R: Band 4, G: Band 3, B: Band 2) by following two criteria: (1) the subsites should cover both water and land; (2) the subsites should cover as many different types of watercolors, water types, and land features as possible. Overall, 145 subsites were selected from these 22 test site (Table 1). Although various land features have been covered by these subsites, their sample sizes (or area) varied significantly due to their different frequencies of presences in the real world. For example, vegetated land area could be more likely to be sampled than shadowed land near water bodies. To mitigate such imbalanced sample sizes, 35 additional subsites only covered "uncommon" land features (e.g., built-up land, shadowed land) were selected. Finally, a total of 180 subsites were prepared as the test dataset.

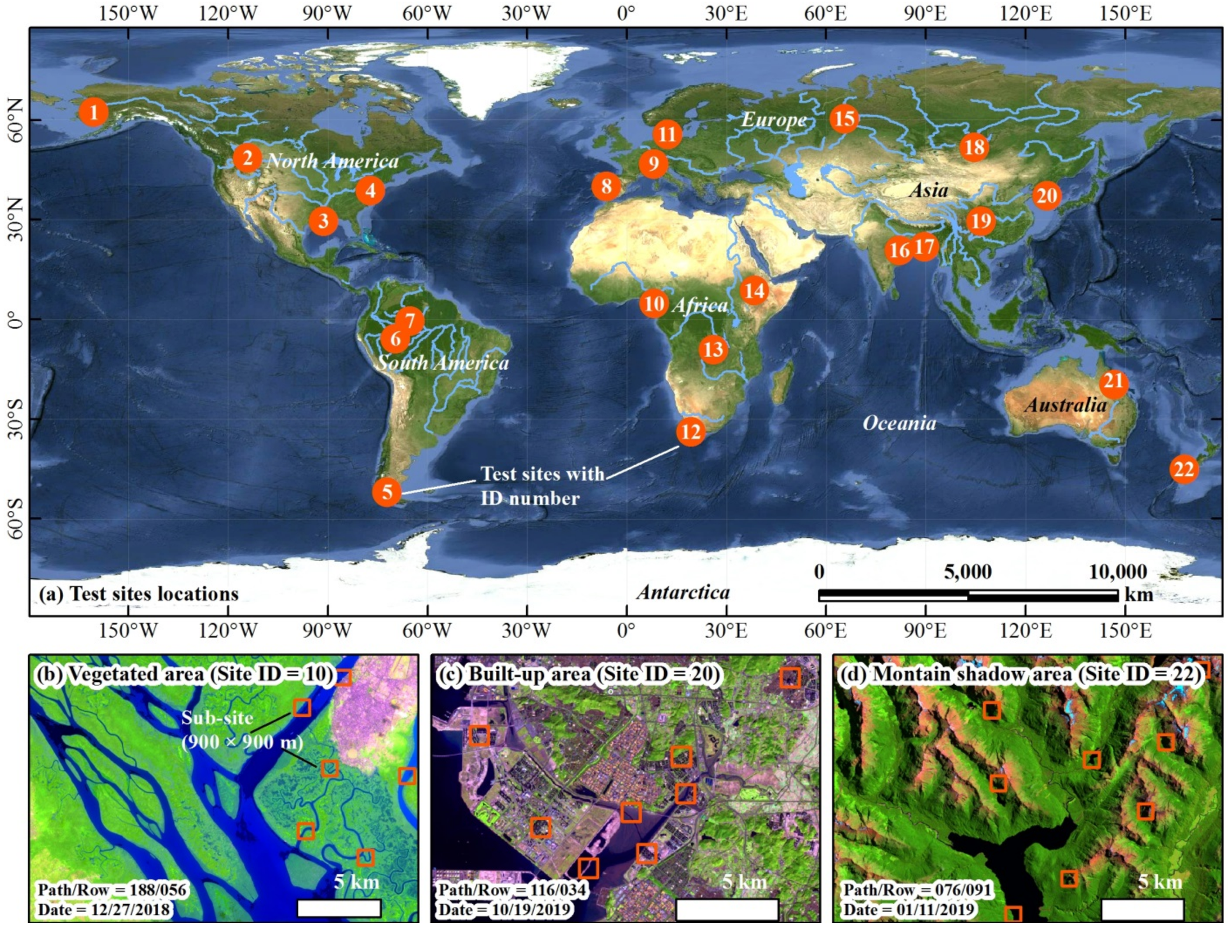

Fig. 1. (a) Locations of the 22 test sites representing three types of water-land environments: water bodies surrounded by vegetated land, built-up land, and shadowed land. The numbers (1 - 22) mark site IDs. (b), (c), and (d) are examples of test site images (R: Band 6, G: Band 5, and B: Band 4 in Landsat-8 OLI image) illustrating water bodies surrounded by vegetated area, built-up area, and mountain shadow area, respectively. The red squares (900 × 900 m) denote subsites that were extracted for preparing test data. All of the 22 test site images are shown in the supplementary Fig. S1.

Table 1 Selected 22 test sites and corresponding Landsat-8 images with different environmental conditions. Watercolors include dark-blue (D), green (G), brown (B), dark-blue-green (DG), dark-blue-brown (DB), and green-brown (GB). Their typical colors are illustrated in the table's header. Water types include river (R), lake/reservoir/pond (LPR), and ditch/creek (DC). Background features include high-density vegetation (HV), moderate-density vegetation (MV), sparse vegetation (SV), built-up area (BA), cloud shadow (CS), building shadow (BS), and terrain shadow (TS).

Site ID	Path/Row	Image Date	Watercolor						Water type			Land features						
			D	G	B	DG	DB	GB	R	LRP	DC	HV	MV	SV	BA	CS	BS	TS
1	075/016	06/13/2019	●		●				●		●	●	●			●		
2	041/026	09/03/2019	●			●							●	●				●
3	023/039	10/23/2019	●	●	●	●			●	●	●	●	●		●			
4	015/033	10/15/2019	●	●				●				●	●	●	●			
5	230/096	02/03/2019	●		●		●			●			●	●				
6	002/061	03/26/2019			●				●		●	●				●		
7	001/060	08/26/2019	●		●							●						
8	203/032	09/18/2019	●	●		●			●	●			●	●				
9	195/027	07/24/2019		●		●				●		●			●	●		●
10	188/056	12/27/2018			●		●		●		●	●	●		●			
11	195/021	04/19/2019		●		●		●		●			●	●	●			
12	175/083	11/14/2018	●					●		●			●	●	●			●
13	174/066	06/19/2019	●			●				●	●	●	●					
14	168/054	02/01/2019	●	●				●		●			●	●				●
15	162/018	05/30/2019	●			●				●	●	●	●	●				
16	142/045	04/16/2019	●				●			●			●	●	●			
17	138/045	03/03/2019			●		●		●		●	●	●					
18	134/024	08/30/2019	●						●	●			●	●	●			
19	127/040	08/26/2018			●	●		●	●	●		●	●		●		●	
20	116/034	10/19/2019			●		●		●	●	●		●	●	●		●	
21	095/073	07/12/2019		●		●				●	●			●	●			
22	076/091	01/11/2019	●							●			●	●				●

2.2. Data Materials

2.2.1 Landsat-8 OLI images

A total of 22 Landsat-8 OLI images with each covered one test site and acquired in different seasons were selected (Table 1). They were standard Landsat-8 surface reflectance level-2 products with 30 m spatial resolution and more information of those products can be found in the Product Guide (2018). The images were firstly downloaded from the USGS Earth Resources Observation and Science Center Science Processing Architecture on Demand Interface (https://espa.cr.usgs.gov/) and then were clipped into sub-images using subsite-defined square polygons (900 × 900 m, see Fig. 1). Only the pixels that entirely contained by the subsite square polygons were selected. In total, 180 clipped subsite images with 153140 pixels of seven-band surface reflectances (range from 0 to 1 in float) were extracted and stored as integer values by scaling 10, 000 (any pixels with values less than 0 or greater than 10,000 were masked).

2.2.2 High spatial resolution images

PlanetScope Analytic Ortho Scene (PSAOS) products were served as reference data for labeling Landsat-8 pixels as water and non-water. PSAOS images have a high spatial resolution (3 m) and very high temporal resolution (1-3 days), which makes them ideal reference data sources. They consist of four bands: blue (455 – 515 nm), green (500 – 590 nm), red (590 – 670 nm), and near-infrared (NIR, 780 – 860 nm). Before distributed to users, they are orthorectified to remove distortion caused by terrain and to eliminate the perspective effect on the ground (not on buildings), as well as to restore the geometry of an image taken at zenith (Planet Labs Inc., 2018).

Each PSAOS image was carefully selected in this study such that their acquisition dates matched exactly the same as that of the corresponding Landsat images (Table 1). In other words, both the PSAOS image and corresponding Landsat-8 image were captured on the same day. All the PSAOS images were obtained from Planet Explorer (https://www.planet.com/explorer/; Planet Team, 2017) and manually georeferenced to the corresponding Landsat-8 image. The geo-referencing errors of PSAOS images were less than one pixel (30 m), which minimized the geolocation error that could potentially propagate to the final classification results.

2.2.3 Test dataset preparation

Each test pixel (153,140 in total) holds several attributions: location, source image, band reflectance, WIs values, feature type (water or non-water), percentage of water. The first three attributions were directly obtained from the source Landsat-8 image. WIs values were derived from band reflectance with specific algorithms (detailed in Section 3.1). Feature type and percentage of water were identified with the help of the PSAOS reference images which involved three steps. First, the PSAOS images were displayed in false-color (R: NIR, G: Red, B: Green) and carefully classified into water (including different watercolors) and non-water polygons (including vegetated land, built-up land, or shadowed land) through visual digitization with expert experience. Then, the water area percentage of each corresponding 30 m by 30 m pixel was derived with a series of spatial analysis functions (e.g., create fishnet, clip, etc.) coded in Python script in ArcGIS 10.5 (version 10.5.0.6491; ESRI, 2016). Finally, all the pixels with water percentage higher than 50% were labeled as water, otherwise as non-water (Feyisa et al., 2014; Yang et al., 2015). The non-water type was further identified as vegetated land, built-up land, or shadowed land. In addition, pixels with water percentage equal to 0 (non-water type) or 100% (water) were considered as pure pixels, otherwise as mixed pixels. The numbers of water pixels, non-water pixels, pure pixels, and mixed pixels are listed in Table 2. The dataset is now avaiable at Mendeley Data repository (http://dx.doi.org/10.17632/mfp7jvw7yk.1).

Table 3 Count numbers of water pixels, non-water pixels, pure pixels, and mixed pixels in the test dataset

	Pure pixels	Mixed pixels	Total
Water pixels	47024	5837	52861
Non-water pixels	93973	6306	100279
Total	140997	12143	153140

3. Methods

3.1. The common way of using spectral water indices: TSWI

Although numerous Landsat WIs have been developed over the past three decades, five are prevailing with distinct merits for different water-land environments: NDWI, MNDWI, AWEI0 (also known as AWEI$_{nsh}$), AWEI1 (also known as AWEI$_{sh}$), and WI2015 (Table 3). The application of TSWI for water classification is straightforward: applying a pre-defined threshold to a pre-selected single WI image. Pixels with values larger than the threshold are labeled as water, otherwise, they are labeled as non-water. Please note that the applications of TSWI method using NDWI, MNDWI, AWEI0, AWEI1, and WI2015, are denoted hereafter as TSWI$_{NDWI}$, TSWI$_{MNDWI}$, TSWI$_{AWEI0}$, TSWI$_{AWEI1}$, TSWI$_{WI2015}$, respectively.

Table 3 Five prevailing WIs used in TSWI for mapping water bodies with Landsat-8 OLI images. ρ is surface reflectance and b1, b2, …, b7 are band numbers of Landsat-8 OLI images. The superscript notes "a" and "b" indicate the pre-defined thresholds suggested by the source authors and Fisher et al. (2016), respectively. Note that the pre-defined thresholds suggested by Fisher et al. (2016) were also adopted in the proposed CDWI.

Water index	Equation adjusted for Landsat-8 OLI	Source reference	Pre-defined Threshold^{a}	Pre-defined threshold^{b}
NDWI	$(\rho_{b3}-\rho_{b5}) / (\rho_{b3}+\rho_{b5})$	McFeeters (1996)	0.00	-0.21
MNDWI	$(\rho_{b3}-\rho_{b6}) / (\rho_{b3}+\rho_{b6})$	Xu (2006)	0.00	0.00
AWEI0	$4(\rho_{b3}-\rho_{b6})-0.25\rho_{b5}-2.75\rho_{b7}$	Feyisa et al. (2014)	0.00	-0.07
AWEI1	$\rho_{b2+}2.5\rho_{b3}-1.5(\rho_{b5}+\rho_{b6})-0.25\rho_{b7}$	Feyisa et al. (2014)	0.00	-0.02
WI2015	$1.7204+171\rho_{b3}+3\rho_{b4}-70\rho_{b5}-45\rho_{b6}-71\rho_{b7}$	Fisher et al. (2016)	0.63	0.63

3.2. The ensemble spectral water indices: CDWI

3.2.1 Principle of CDWI

An alternative way of using WIs for water classification is proposed here to handle the common concerns of using TSWI: WI selection and the corresponding threshold determining. The approach is designed as the Collaborative Decision-making with Water Indices (CDWI). It combines a group of weighted and thresholded WI images to generate a new water probability image and a new decision-making probability threshold is applied to extract water. The rationale of the collaborative decision-making principle is that a group of variables can provide potentially complementary information to support a more reliable decision than that based on a single component (Kacprzyk and Fedrizzi 2012). When it comes to handling the concerns of TSWI, CDWI could provide an alternative way of selecting WIs and a potential stable threshold for extracting water. The step-by-step procedure of CDWI is as follows (see also Fig. 2) and the ready-to-use Python script is attached as a supplementary file.

- **Step 1**: Select a group of WIs and calculate corresponding WI images. In this study, the five prevailing WIs were used as listed in Table 3. The reason for selecting these WIs is that they were reported showing complementary merits in classifying water over different water-land environments. For example, MNDWI was designed to separate water from vegetated area and built-up area (Ji, et al., 2009; Xu, 2006), AWEI0 performs better than other WIs in differing built-up land from water, and AWEI1 is good at distinguishing shadow from water (Feyisa et al., 2014).

- **Step 2**: Apply an appropriate pre-defined threshold to each WI image to initially classify water (labeled 1) and non-water (labeled 0). Note that this step is also known as applying TSWI for water classification.
- **Step 3**: Apply an appropriate weight to each initially classified TSWI image. The sum of all weights is 1. TSWI method with better performance needs to be assigned a larger weight to its classified TSWI image.
- **Step 4**: Sum up all weighted images to achieve a new CDWI image. Its pixel values are considered to represent water probability. The larger CDWI pixel value, the greater confidence of the pixel being decided as water.
- **Step 5**: Apply a probability decision-making threshold (T_{CDWI}) to binarize the CDWI image and obtain the final water image.

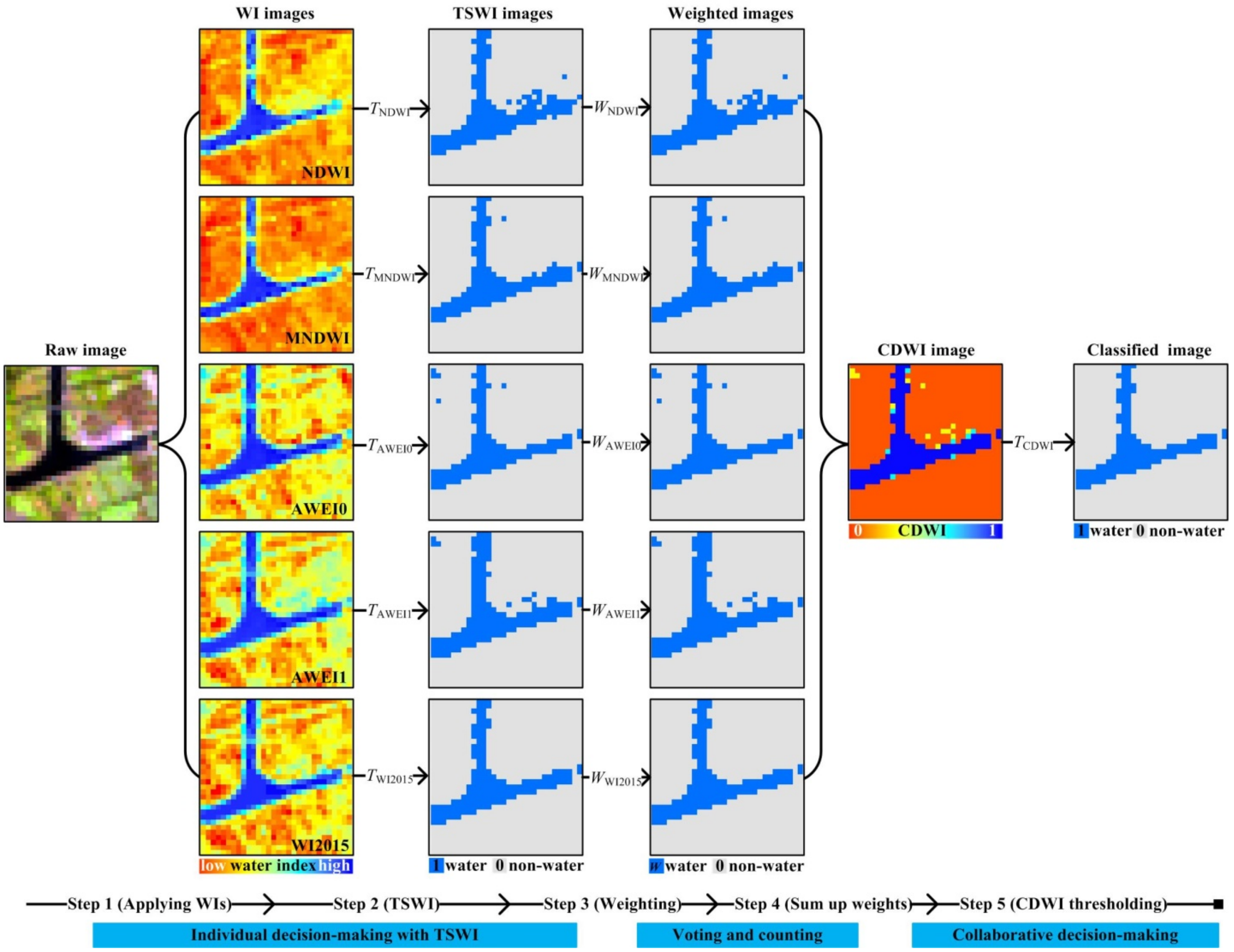

Fig. 2. The workflow of CDWI exemplified with a Landsat-8 OLI subsite image. *T* and *W* stand for threshold and weight, respectively.

From the perspective of the collaborative decision-making process, the workflow of CDWI can be understood as following. Consider there is a decision-making committee named CDWI, and the job of which is to decide whether image pixels are water or not. It has several experienced committee members (i.e., $\mathrm{TSWI}_{\mathrm{NDWI}}$, $\mathrm{TSWI}_{\mathrm{MNDWI}}$, $\mathrm{TSWI}_{\mathrm{AWEI0}}$, $\mathrm{TSWI}_{\mathrm{AWEI1}}$, and $\mathrm{TSWI}_{\mathrm{WI2015}}$ in this study) but with different abilities (weights). In the processing of collaborative decision-making, each committee member would independently make an initial decision (water or non-water) first with TSWI method. Then, each member assigns its weight (*W*) to the corresponding TSWI image. The sum of all weighted TSWI images forms a new CDWI image waiting for the final decision: pixels with values larger than T_{CDWI} are classified as water, otherwise non-water.

3.2.2 CDWI parameters estimation

The application of CDWI requires three types of parameters: (1) the pre-defined WI tresholds (T_{NDWI}, T_{MNDWI}, T_{AWEI0}, T_{AWEI1}, and T_{WI2015}) for applying the five TSWI methods, (2) the weights (W_{NDWI}, W_{MNDWI}, W_{AWEI0}, W_{AWEI1}, and W_{WI2015}) of the five TSWI methods, and (3) the CDWI threshold (T_{CDWI}) for slicing the final CDWI image (Fig. 2). Since the pre-defined thresholds have already been recommended by the previous authors in applying the five TSWI (Table 3), they are directly adopted in this CDWI approach as well. The other two parameters were estimated in the following ways (Fig. 3).

(1) Weights of the five TSWI methods

According to the principle of CDWI, a TSWI method showing better performance should hold a larger weight. Assessing performances of the five TSWI methods and determining their weights were conducted accordingly as below. First, we prepared 1,000 sample sets with each formed by 1,000 randomly selected pixels from the test dataset: 500 are water and 500 are non-water. Note that the same size of water and non-water pixels can minimize the uncertainty in validation caused by imbalanced sample size (Warmink, et al., 2010). Then, the five TSWI methods with the recommended corresponding pre-defined thresholds (Table 3) were applied to each sample set, and their accuracies were evaluated by F1-score, a harmonic accuracy assessment metric as detailed in Section 3.3.1 (Daskalaki et al., 2006; Zhong et al., 2019). As each sample set produced five F1-scores for the five TSWI methods, and the one holding the maximum F1-score was considered as performed the best and counted one. After this process went for the entire 1, 000 sample sets, each WI would get a final count number (*N*) and the sum of five count numbers equals to 1,000. Finally, the weight of a TSWI method was determined by the proportion of its count value to the sum of all count values. In this study, for example, the weight of $\text{TSWI}_{\text{NDWI}}$ (W_{NDWI} in Fig. 2) was calculated as Eq. (1):

$$W_{\text{NDWI}} = \frac{N_{\text{NDWI}}}{N_{\text{NDWI}} + N_{\text{MNDWI}} + N_{\text{AWEI0}} + N_{\text{AWEI1}} + N_{\text{WI2015}}} = \frac{N_{\text{NDWI}}}{1000} \tag{1}$$

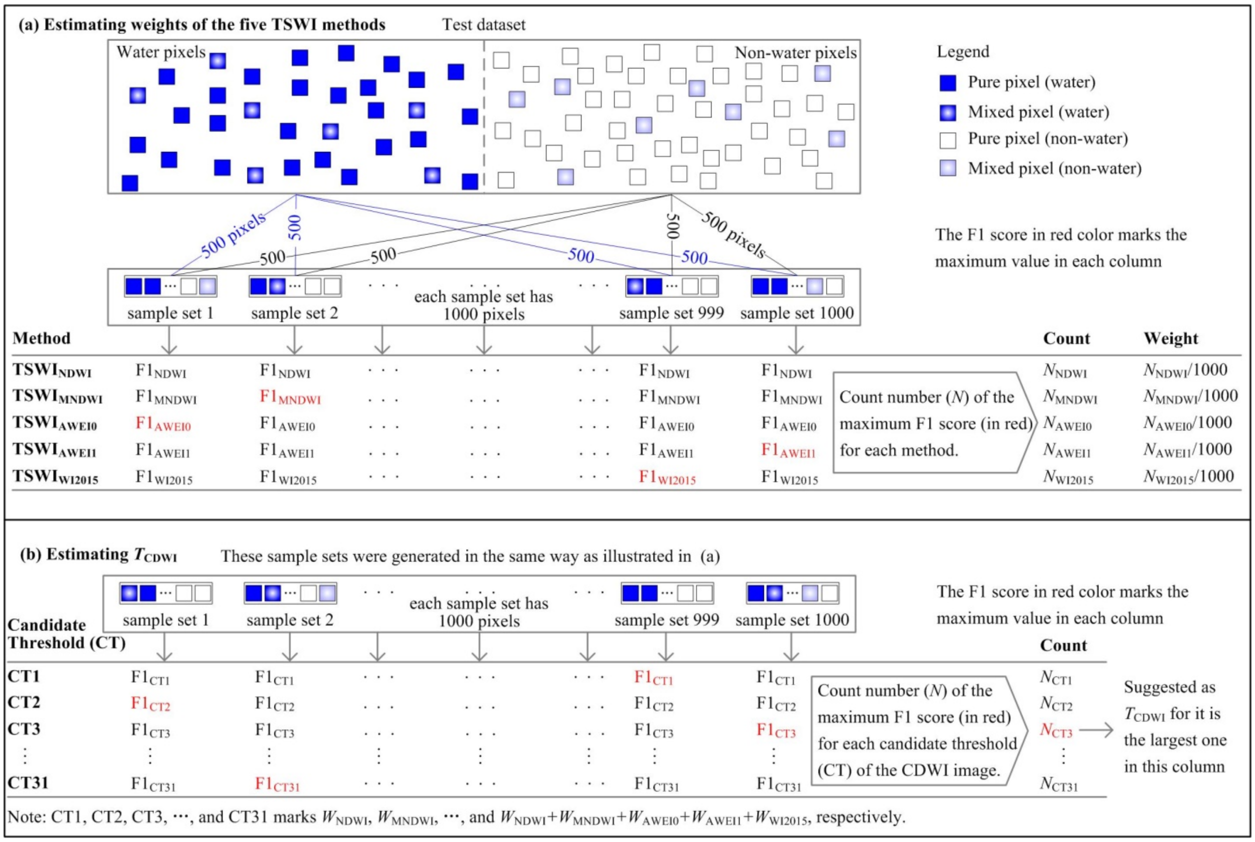

Fig. 3. The workflow of estimating (a) weights of the five TSWI methods and (b) CDWI threshold (T_{CDWI}).

(2) CDWI threshold (T_{CDWI})

Since CDWI image is sum of several weighted TSWI images (Fig. 2), any pixel value of such CDWI image is the sum of one combination weights of TSWI methods. In total, there are 31 different combinations of weights in the case of this study (Fig. 2): W_{NDWI}, W_{MNDWI}, W_{AWEI0}, W_{AWEI1}, W_{WI2015}, $W_{\text{NDWI}}+W_{\text{MNDWI}}$, $W_{\text{NDWI}}+W_{\text{AWEI0}}$, $W_{\text{NDWI}}+W_{\text{AWEI1}}$, ..., and $W_{\text{NDWI}}+W_{\text{MNDWI}}+W_{\text{AWEI0}}+W_{\text{AWEI1}}+W_{\text{WI2015}}$ or 1. Therefore, the final recommended T_{CDWI} should be determined from this list. The determination process is straightforward. First, we generated 1,000 sample sets in the same way as mentioned above. Each sample set would produce 31 F1-scores after applying 31 candidate CDWI thresholds independently. Among these 31 F1-scores, the maximum score and its corresponding threshold was identified and counted. After applying this procedure to all 1,000 sample sets, the threshold which obtained the largest count number was identified as the recommended T_{CDWI}, for it held the most cases of holding the maximum F1-scores than the other candidate thresholds.

3. 3. Performance assessment

3.3.1 Accuracy assessment

As mentioned in the Section 2.1, there are 145 out of 180 subsite images cover both water and land features (the other 35 out 180 subsite images only cover land features). Therefore, the five TSWI methods and the CDWI method were applied to these 145 subsite images to assess their accuracy stabilities over different water-land environments aroud the world. As previous studies suggested, both F1-score and Youden's index (YI) were used to assess accuracies of the six methods (Li et al., 2016; Li et al., 2019; Zhong et al., 2019; Wen et al. 2016). The F1-score is the harmonic average of the producer’s accuracy and user’s accuracy (Daskalaki et al., 2006; Eq. (2)):

$$\text{F1-score} = \frac{2 \times \text{Producer's accuracy} \times \text{User's accuracy}}{\text{Producer's accuracy} + \text{User's accuracy}} \tag{2}$$

The producer’s accuracy is the percentage of correctly classified water pixels from the total number of true water pixels. The user’s accuracy is the percentage of correctly classified water pixels from the total number of classified water pixels. F1-score reaches its best value at 1 and worst at 0. It is considered more objective than overall accuracy (the percentage of correctly classified pixels, both as water and non-water, from the total number of pixels) in our binary classification case because a water body mostly covers a small portion of the image under evaluation. The YI was often used for determining local optimal thresholds (Wen et al. 2016), and it was considered as an indicator of water classification accuracy (Eq. (3)). The larger YI value, the smaller sum of omission error and commission error.

$$\text{YI} = 1 - (\text{Omission error} + \text{Commission error}) \tag{3}$$

3.3.2 Sensitivity to thresholds

Sensitivity to thresholds, defined as how much the accuracy would change by changing the threshold values for a given method (TSWI methods or CDWI method), is indicated by the slope of a threshold-accuracy curve. A robust classification method should, therefore, be less sensitive (low absolute slope value) to threshold changes. For TSWI methods, such thresholds are the pre-defined WI thresholds; for the CDWI method, such thresholds involve both the pre-defined WI thresholds and $T_{CDWI.}$

For a given TSWI method, classification outcome purely affected by the pre-defined thresholds (Fig. 2). Each pre-defined threshold outputs a classification result and one accuracy (F1-score or YI value). The sensitivity analysis, thus, involves selecting different pre-defined thresholds and calculating their corresponding accuracies. To make such selection more objective, the local optimal thresholds of 145 subsite images were served as candidate pre-defined thresholds. For a subsite image, its local optimal threshold was determined as the threshold at which the YI gained the maximum value (Fisher et al., 2016).

For the proposed CDWI method, its accuracy relies on both the five pre-defined WI thresholds (Table 3) and T_{CDWI} (Figs. 2 and 3). To make the sensitivity analysis more clearly, T_{CDWI} was fixed (to the suggested one) in analyzing the sensitivity of CDWI to WI thresholds; while WI thresholds were fixed (to the suggested ones, see Table 3) in analyzing the sensitivity of CDWI method to T_{CDWI}. Each group of the five selected WI thresholds will produce one F1-score of the CDWI. As a WI threshold could be chosen from the 145 candidate local optimal thresholds, 145^5 (=64,097,340,625) different threshold groups could be generated with 145^5 accuracies. To reduce this huge computational burden, the 145 candidate local optimal thresholds were split into 15 equal interval groups and the central value of each group was reselected. Finally, there are 15^5 (=759375) WI threshold groups and 15^5 corresponding CDWI accuracies are obtained. Each selected WI threshold would generate one accuracy for the corresponding TSWI method but 15^4 (=50625) accuracies for the CDWI method. To make them comparable, the mean accuracies of the CDWI method was used for sensitivity analysis.

4. Results

4.1. Suggested CDWI parameters

The parameters of applying the CDWI method were estimated carefully (Fig. 3) and are could be directly used in further applications given that they are evaluated by the dateset collected from various different water-land environment around the world. To estimate the weights of the five TSWI methods, their accuracies were assessed. Overall, $\text{TSWI}_{\text{MNDWI}}$ showed the best performance for classifying water and then followed by $\text{TSWI}_{\text{WI2015}}$, $\text{TSWI}_{\text{AWEI1}}$, $\text{TSWI}_{\text{AWEI0}}$, and $\text{TSWI}_{\text{NDWI}}$. Accordingly, the suggested five weights the TSWI methods were estimated as 0.640, 0.333, 0.019, 0.008, and 0.000, respectively (Fig. 4a). Note that $\text{TSWI}_{\text{NDWI}}$ performed the worst among the five TSWI method and got zero weight, for it held zero cases among 1,000 sample sets that gained the highest F1-scores.

With regard to T_{CDWI}, it suggests 0.648 as the best for further applications for it obtained the largest number of cases that got the maximum F1-score among all the candidate CDWI thresholds (Fig. 4b). The result means that pixel values larger than 0.648 in the CDWI image (sum of weighted TSWI images) are more likely to be labeled as water than non-water. Furthermore, this T_{CDWI} is the sum weights of MNDWI (W_2 = 0.640) and AWEI0 (W_3 = 0.008), which statistically implies that pixels were classified as water by both $\text{TSWI}_{\text{MNDWI}}$ and $\text{TSWI}_{\text{AWEI0}}$ are more likely to be correctly classified than that only classified eigher by $\text{TSWI}_{\text{MNDWI}}$ or $\text{TSWI}_{\text{AWEI0}}$.

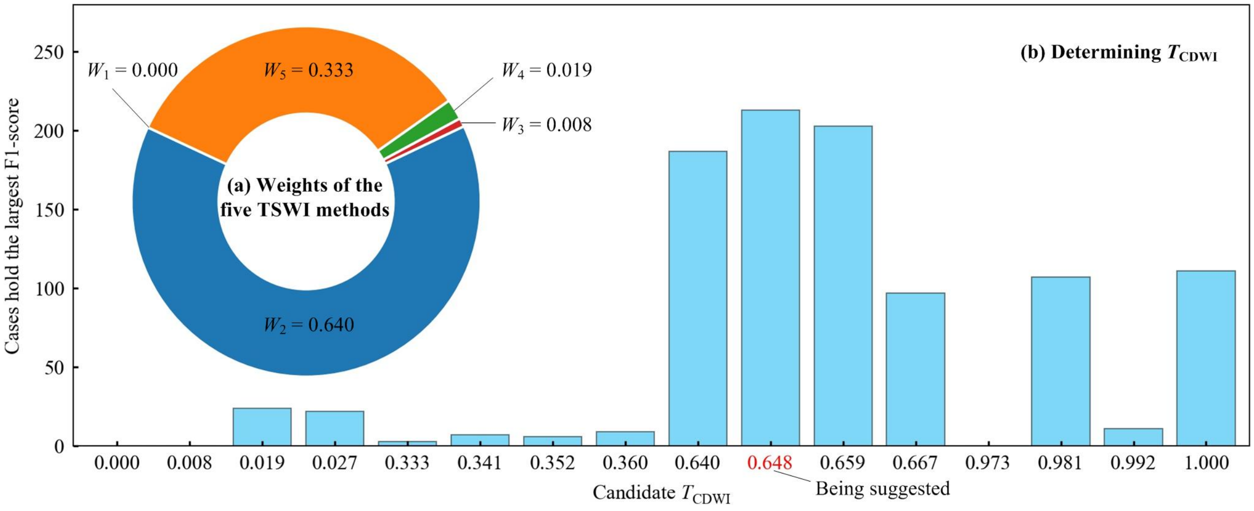

Fig. 4. Suggested parameters of CDWI: (a) wights of the five TSWI methods (W_1, W_2, W_3, W_4, and W_5, see also in Fig. 2), and (b) T_{CDWI}. The red-colored threshold (0.648) in (b) marks the suggested T_{CDWI} for it holds the most cases that obtained the maximum F1-score among all the candidate CDWI thresholds.

4.2. Accuracy assessment over different environments

The accuracies of the six methods were applied to 145 individual subsite images to compare their accuracies over different water-land environments (Fig. 5). All the TSWI methods and CDWI method obtained high accuracies for their F1-scores and YI values greater than 0.9 for most subsites (Fig. 5). Although they all performed relatively well, the differences in their performances can be observed. In general, the number of subsites with their accuracies improved by the CDWI method was much greater than the number of subsites with their accuracies that decreased by the CDWI method. For example, 54% subsite images classified by the CDWI method produced higher F1-scores than that produced by the $TSWI_{MNDWI}$ method, and only 22% subsite images got lower F1-scores by using CDWI than $TSWI_{MNDWI}$ method (Fig. 5b). Moreover, the absolute mean value of decreased accuracies was smaller than that of the increased accuracies. Take YI as an example, such a pattern can be observed as: |-0.029| vs. 0.087 in Fig. 5a, |-0.009| vs. 0.022 in Fig.5b, |-0.011| vs. 0.021 in Fig. 5c, etc. This finding shows that the CDWI method could be more likely to obtain a better water classification result than any TSWI method in general.

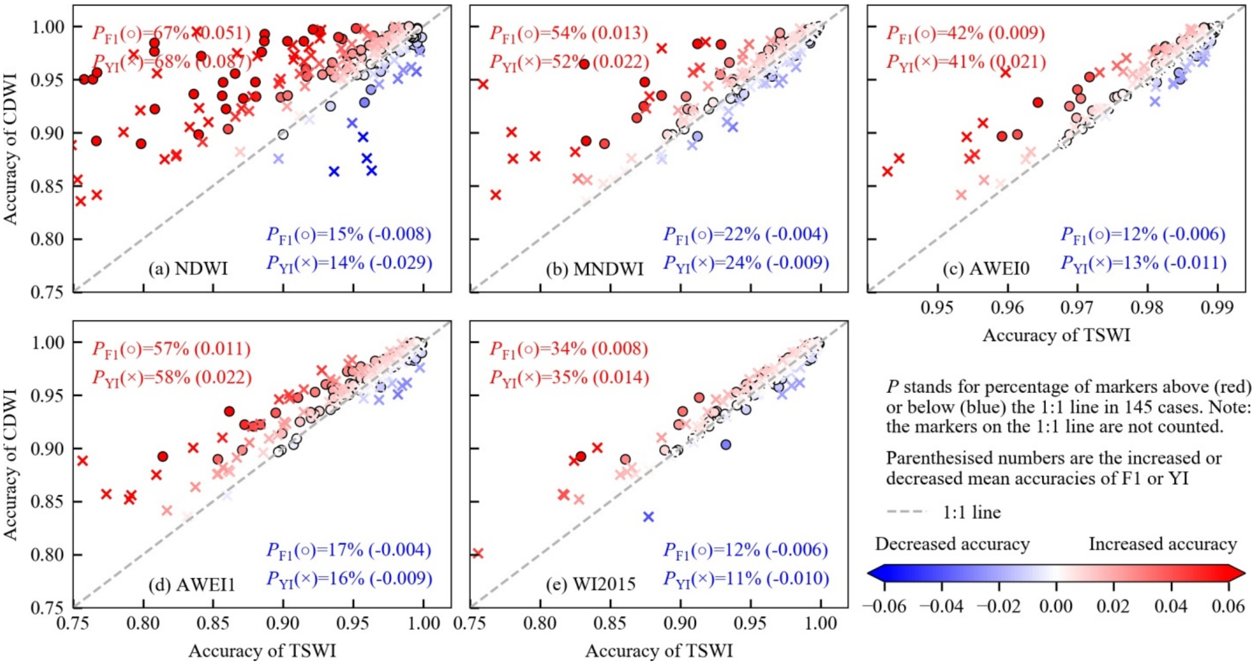

Fig. 5. Accuracy (indicated by F1-score and YI value) comparisons between CDWI and the TSWI method using five WIs: (a) NDWI, (b) MNDWI, (c) AWEI0, (d) AWEI1, and (e) WI2015. Decreased accuracies (blue dots) and increased accuracies (red dots) represent the accuracy difference between the CDWI and TSWI method.

4.3. Sensitivity to threshold

4.3.1 Sensitivity to pre-defined WI thresholds

Each subsite image can obtain a local optimal threshold. For all subsite images, their local optimal threshold varied significantly, as shown in Fig. 6. Generally, the histograms of those local optimal thresholds approximately follow Gaussian distributions. The F1-score of any TSWI method changed dramatically with different pre-defined WI thresholds were used (the blue lines in Fig. 6). Overall, sensitivity curves of all the TSWI methods are in unimodal patterns and peak at their thresholds around the suggested pre-defined thresholds that we used in this study (see Table 3). These sensitivity curves can be broadly categorized into three types: high sensitivity with a steep slope, moderate sensitivity with a moderate slope, and low sensitivity with roughly flat slope. The further distance of a threshold to the suggested pre-defined threshold, the higher sensitivity of a TSWI method to such threshold can be observed (Fig. 6).

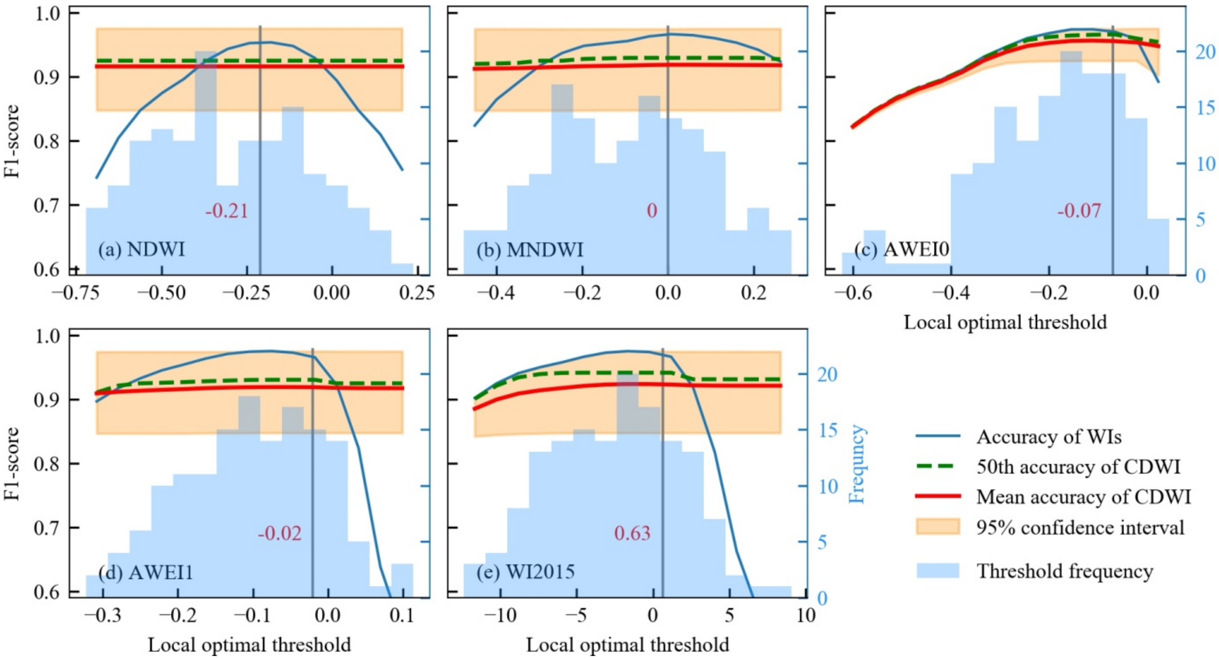

Fig. 6. The sensitivity of F1-score to threhsold for the five TSWI methods and the CDWI method. The sensitivities are indicated by the slope of the sensitivity curve: the threshold-against-F1 curve. The red colored values are the pre-defined WI thresholds suggested by the literatures as listed in Table 3.

In contrast, the proposed CDWI method showed the least sensitive to threshold. That is, no matter what threshold were used, the accuracies of the CDWI method changed slighter than those of any TSWI method. For example, when the threshold changed from -0.45 to 0.26, the F1-score of $\text{TSWI}_{\text{MNDWI}}$ changed from 0.82

to 0.97, whereas the mean F1-score of CDWI method changed from 0.912 to 0.918 (Fig. 6b). This low sensitivity-to-threshold of the CDWI method indicate that the uncertainties related to threshold determination can be significantly reduced compared to the TSWI methods. Such characteristics of CDWI method could make users less worrying about whether the selected thresholds are the optimal ones or not in applications without reference data.

4.3.2 Sensitivity to T_{CDWI}

Overall, the sensitivity of CDWI accuracy to T_{CDWI} is relatively low (Fig. 7). The mean F1-score of the CDWI method changes from 0.940 to 0.956 as the T_{CDWI} changing from 0.008 to 1. 000. It generally shows a "∩" pattern with short increasing, long-flatten, and a slightly decreasing trend in order. In terms of YI value, it also shows a similar sensitivity-to-T_{CDWI} pattern as of F1-score. It is noteworthy that the accuracy produced by combined T_{CDWI} (i.e., summed by two or more TSWI weights) is overall larger than that produced by single T_{CDWI} (i.e., single TSWI weight), which is explained here. All the T_{CDWI} values are denoted by the x-axis ticklabels in Fig. 7, the single T_{CDWI} values are 0.000 (W_1), 0.008 (W_2), 0.019 (W_3), 0.333 (W_4), and 0.640 (W_5); the rest are combined T_{CDWI} values. It is observed that the mean F1-score produced by the 0.027 (W_2+W_3) is larger (0.952) than the mean F1-score produced either by 0.940 (W_2) or 0.951 (W_3). This observation goes for our suggested T_{CDWI} (0.648, W_2+W_5) in the study (Fig. 4): its mean F1-score and YI value are both larger than that produced by corresponding single T_{CDWI}: 0.008 (W_2) and 0.640 (W_5).

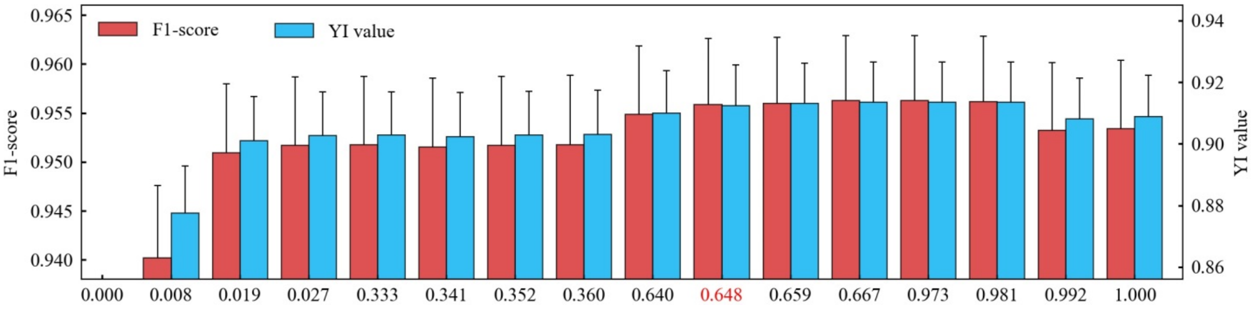

Fig. 7. The sensitivity of CDWI accuracy (F1-score and YI value) to the T_{CDWI}. The red-colored threshold (0.648) marks the suggested T_{CDWI} in this study (see Fig. 4).

5. Discussion

5.1 Uncertainty analysis

5.1.1 Pure pixels vs. mixed pixels

One commonly recognized uncertainty of a water classification method may come from water-land mixed pixels or water-land boundary pixels (Comber et al., 2012; Yang et al., 2015). To better understand how the CDWI works, we compared the performances of the six methods in classifying both pure pixels and mixed pixels of the 145 subsite images (Fig. 8). It is observed that all the TSWI methods and CDWI method performed worse for mixed pixels than for pure pixels. Because the TSWI methods were developed based on the principle that water and land features have distinct reflectance properties: water shows a decrease in reflectance from the visible to infrared wavelengths, while land features (e.g., vegetation) often do not show such reflectance pattern(Xiong et al., 2018). Moreover, those WI methods are “hard” classification methods using a Boolean set (i.e., 0 or 1) to restrict each pixel to either water or non-water types (Yang, et al., 2015). Therefore, classifying mixed pixels often introduce more errors to the result than classifying pure pixels with TSWI methods, due to the averaging of the reflectance properties of the water and non-water components (Fisher et al., 2016). How to reduce the class uncertainty of mixed pixels in classifying water is accordingly a research topic for many researchers.

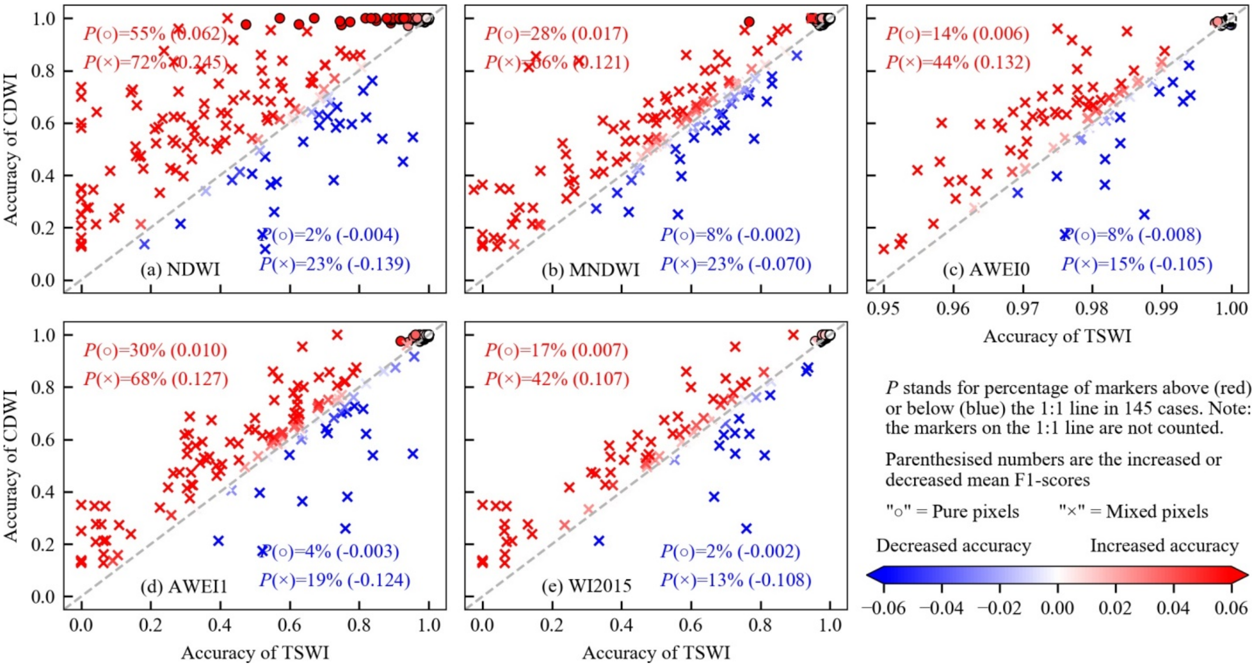

Fig. 8. Accuracy (indicated by F1-score) comparison between CDWI and the five TSWI methods for both pure pixels (o) and mixed pixels (×) of subsite images: (a) NDWI, (b) MNDWI, (c) AWEI0, (d) AWEI1, and (e) WI2015. Decreased accuracies (blue dots) and increased accuracies (red dots) represent the accuracy difference between the CDWI and an individual TSWI method.

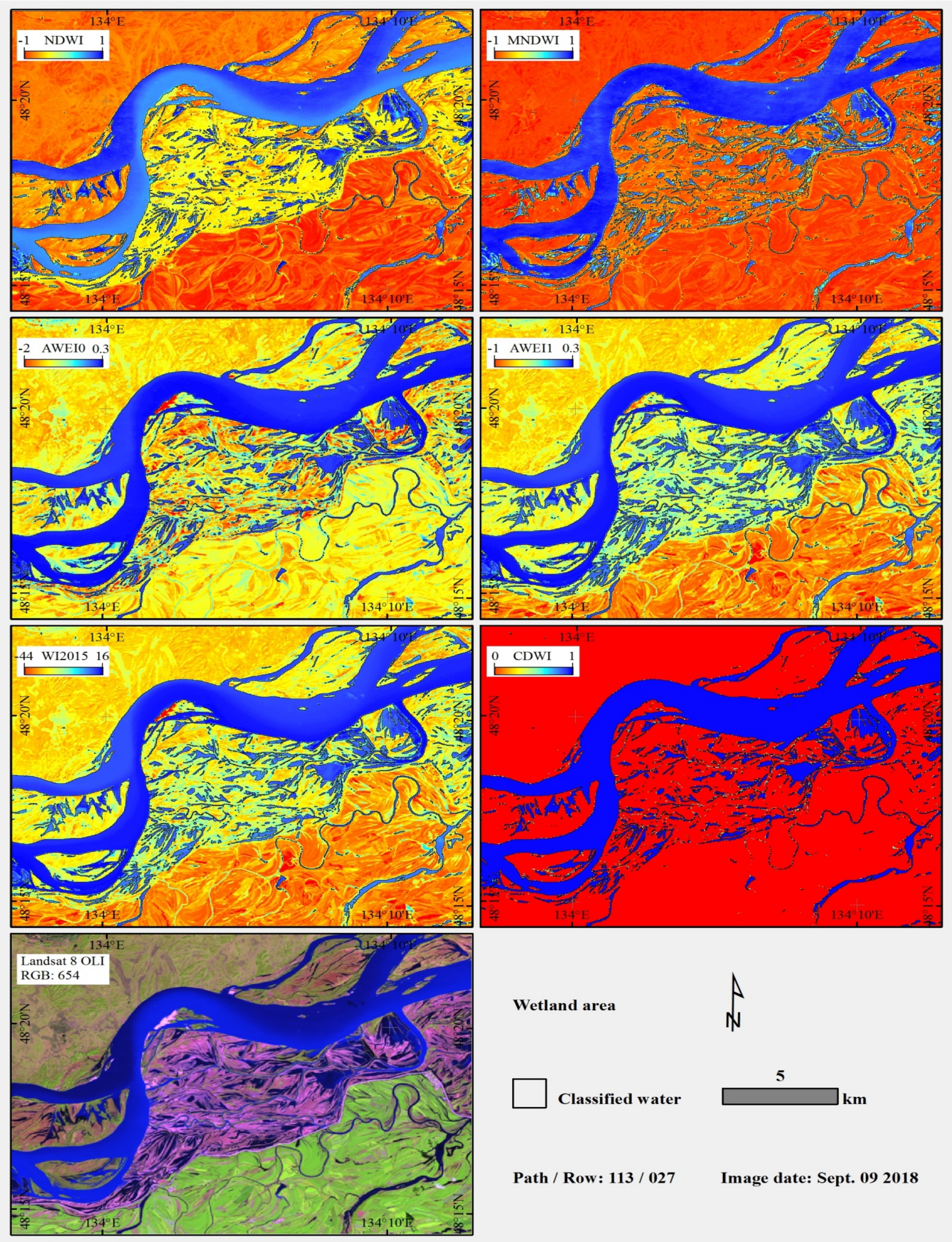

Fig. 9. An example application of using the five TSWI methods and CDWI method in a heterogeneous wetland environment (More examples are illustrated in Supplementary Figs. S2-S6).

Various techniques have been developed in attempts to reduce the uncertainty of mixed pixels in water classification. Some are based on the idea of "soft" classification such as sub-pixel classification and fuzzy classification (Dewi et al., 2016; Xiong et al., 2018). Some use machine learning techniques by taking mixed pixels into the training process (Foody and Mathur 2006). In this study, however, the CDWI achieved higher performance than the other TSWI methods for classifying water from mixed pixels (Fig. 8). It looks like this study provides an alternative way of reducing the uncertainty of mixed pixels. For a mixed pixel labeled as water (i.e., water percentage larger than 50%), the processing of CDWI could be considered as accumulating the probability of a water pixel that being correctly classified. That is, the decision of a mixed pixel be water or non-water is not only based on a single result of an individual TSWI method (except it has large weight) but collectively decided by the results of several TSWI methods. Based on these understandings, it is highly recommended to apply CDWI to the cases where mixed pixels are very common, such as small water bodies (e.g., pond), or water bodies with large perimeters-area ratios (e.g., dike, creek, tide channel, and mountainous reservoir) as shown in Fig. 9.

5.1.2 Different compositions of land features

We observed in some subsites that CDWI performed worse than TSWI methods as illustrated in Fig. 5 (the blue dots below the 1:1 line). One reason could be the parameters of CDWI were estimated from a simulated general scenario, not from the specific scenario of each subsite. Such a general scenario was simulated by 1, 000 sample sets, with each of them formed by 1,000 randomly selected water and non-water pixels from the test dataset. Since the dataset collected from various water-land environments around the world (Fig. 1 and Table 1), a general scenario could consist of water with different colors, and land features with most covered by vegetation and some parts covered by built-up land and shadows. However, for some specific scenarios, the proportion of land components may differ a lot from the general scenario. For example, an urban is mostly occupied by built-up land and building shadows and a small portion of vegetated land. In such a case, the suggested parameters of CDWI in Fig. 4 could not perform well than the ones carefully designed for an urban area, like AWEI0 and MNDWI (Feyisa et al., 2014).

To explore more application scenarios, we first simulated a variety of land environments that consisted of different fractions of three typical land features, namely vegetated land, built-up land, and shadowed land. For each simulated land environment, the corresponding five WI weights are estimated in the same way as that for the general scenario (Fig. 10). Overall, the performances (indicated by TSWI weights) of both $\mathrm{TSWI_{NDWI}}$ and $\mathrm{TSWI_{AWEI1}}$ are not sensitive to any kind of land environment and gained the lowest weights (0 or near to 0). When the fraction of shadowed land larger than 10%, $\mathrm{TSWI_{MNDWI}}$ gained the largest weights than the other

TSWI methods. It implies that in the scene with a large portion of shadows, image classified by $\text{TSWI}_{\text{MNDWI}}$ method should be assigned dominate weight than that by the other TSWI methods in applying CDWI method; or if one just wants to use TSWI method, the $\text{TSWI}_{\text{MNDWI}}$ should also be suggested for guiding WI selection in applying TSWI method. It also shows that the AWEI0 is sensitive to the fraction of built-up land: the more built-up land in an application, the higher weight of the $\text{TSWI}_{\text{AWEI0}}$ gains (Feyisa et al., 2014; Fisher et al., 2016). In an extreme scenario such as in urban areas, it is suggested to assign the largest weight to $\text{TSWI}_{\text{AWEI0}}$ than other TSWI methods in applying the CDWI method. We recommend that the above findings and Fig. 10 could be served as a general guidance or a look-up-table for selecting WI in water classification applications using either TSWI or CDWI methods.

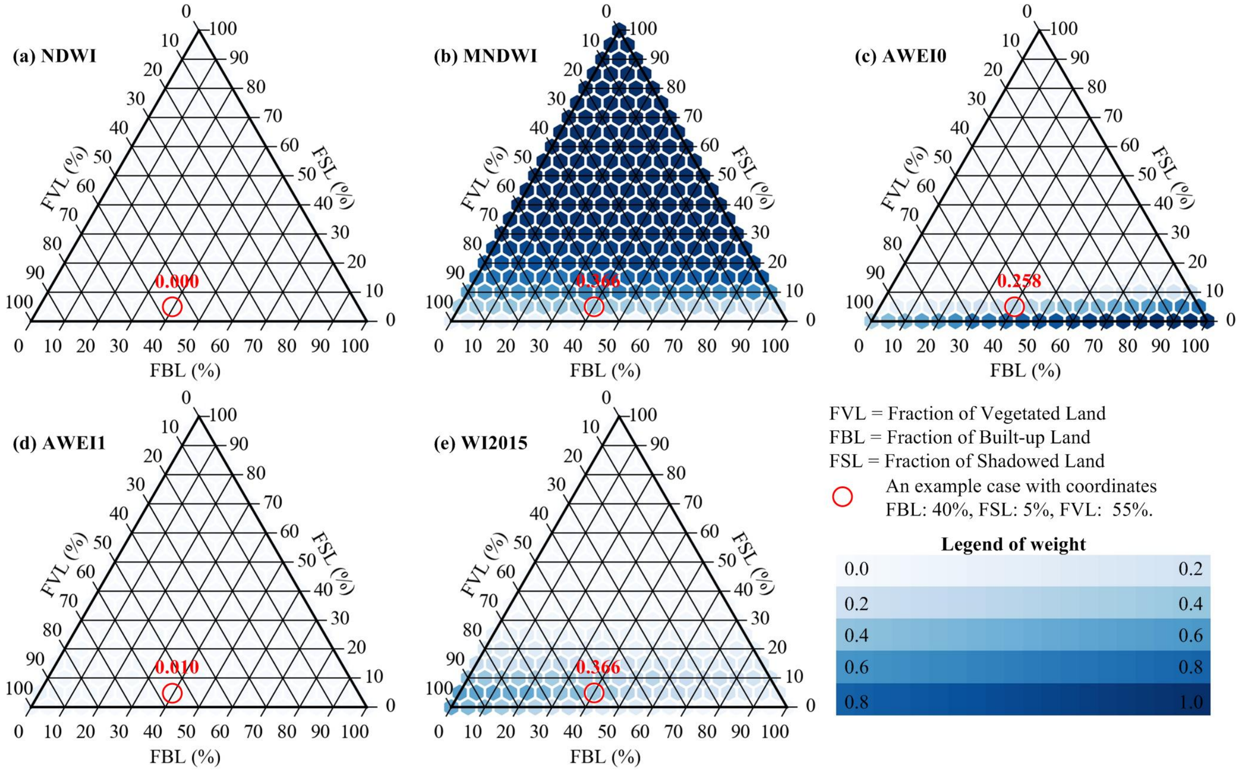

Fig. 10. Ternary plot of TSWI weights (*W*) for different fraction combinations of vegetated land, built-up land, and shadowed land. (a-e) denotes the five TSWI methods with using the five WIs: (a) NDWI, (b) MNDWI, (c) AWEI0, (d) AWEI1, and (e) WI2015.

5.2. Transferability of the CDWI

Different from the common WI methods which were designed for a specific sensor with fixed equations

(Feyisa et al., 2014; McFeeters 1996; Xu 2006), the CDWI could also be considered as a new framework that could readily be used in many applications involving different sensors. First, both the number and the form of TSWI methods involved in the CDWI are not fixed and can be adjusted according to practical conditions. For example, the existing water indices that are not used in this study, such as TCW (Crist 1985), WRI (Rokni et al., 2014), TSUWI (Wu et al., 2018) and MBWI (Wang et al., 2018), could be integrated readily into the CDWI method in further applications. Likewise, as newly designed water indices become available, they can be brought into the framework of CDWI. Moreover, any water classification maps either obtained by TSWI methods or by more sophisticated methods (e.g., Random Forest and Support Vector Machine; see Acharya et al., 2016; Ireland et al., 2015) can be included in the CDWI method to determine the final water classification results. Second, although the proposed CDWI method is tested and demonstrated on Landsat-8 OLI images, it is also suggested for application to Landsat TM/ETM+ images because the TSWIs used here were all originally designed for Landsat TM/ETM+ images (Huang et al., 2018). Since these TSWI methods work well on the Landsat-8 OLI images in this study, they should be suitable for Landsat TM/ETM+ images as well. Third, the framework of the CDWI method can be applied to other types of images with different bands than the Landsat images, such as MODIS (Sharma et al., 2015), Sentinel-2A/B (Du et al., 2016), and HJ-1A/B images (Lu et al., 2011). Because the image bands of these images are very different from those of the Landsat images, their sensor-dependent water indices should be carefully selected before using the CDWI method.

In summary, the proposed CDWI method has four critical potential advantages:

(1) The operation procedure of CDWI is straightforward, applied with basic raster algebra. Users can expand any TSWI methods into the CDWI framework.

(2) The robustness of the CDWI is higher than that of the TSWI methods making it suitable for a wide range of applications over different water-land environments.

(3) The accuracy of the CDWI is less sensitive to the threshold (both pre-defined WI thresholds and T_{CDWI}) selection compared to the TSWI methods, such that the need for tedious parameter tuning of the threshold is reduced or avoided.

(4) The framework underlying the CDWI is not WI dependent and sensor dependent. It has the potential to be applied to other indices (e.g., impervious surface index) and other sensors (e.g., Landsat TM/ETM+,

MODIS, and Sentinel-2).

6. Conclusions

The TSWI methods are widely adopted in water mapping applications due to their potential ease-of-use and generally acceptable performances. However, two concerns need to be carefully considered before applying them in practice: the selection of WI and the determination of an appropriate threshold for the given WI. In practice, answers to these two concerns could be affected by several subjective factors, such as experiments and personal preference. To overcome these two concerns, a new ensemble way of using WIs for water mapping approach that integrates five widely used WIs is proposed, namely the CDWI, based on the collaborative decision-making principle.

A total of 145 subsite images were selected representing different geographical areas with distinct water-land environments and different seasonal patterns. The performances of the CDWI method and the five TSWI methods were assessed in terms of accuracy and robustness. It was found that (1) the CDWI produced higher or comparable accuracies than the five benchmark TSWI methods for most cases, making it less sensitive to application scenarios and, thus, suitable for more different water-land environments. (2) The accuracy of the CDWI is much less sensitive to the pre-defined WI thresholds chosen for the TSWI methods; (3) The underlying framework of CDWI has great potential for transferability and further application. For example, it can be modified readily by adding new WIs in the future. Moreover, the principle underlying the CDWI method is not sensor-dependent and, thus, the proposed CDWI can be applied to different types of images, such as Landsat TM/ETM+, MODIS, Sentinel-2A/B and HJ-1A/B images in future applications.

Acknowledgments

This work is supported by the Open Fund of the State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (No: 18R07), the Open Fund of the State Key Laboratory of Lake Science and Environment, Chinese Academy of Sciences (No: 2018SKL006), and by the National Natural Science Foundation of China (No: 41501096, 51779241, and 41401051). The authors thank the anonymous reviewers for providing constructive comments and feedbacks on this paper.

References

Acharya, T., Lee, D., Yang, I., Lee, J., 2016. Identification of water bodies in a Landsat 8 OLI image using a J48 decision tree. Sensors 16, 1075.

Acharya, T.D., Subedi, A., Lee, D.H., 2018. Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. Sensors 18, 2580.

Allen, G.H., Pavelsky, T.M., 2018. Global extent of rivers and streams. Science 361, 585-588.

Berry, P., Garlick, J., Freeman, J., Mathers, E., 2005. Global inland water monitoring from multi-mission altimetry. Geophys. Res. Lett. 32, L16401.

Bukata, R.P., Jerome, J.H., Kondratyev, A.S., Pozdnyakov, D.V., 2018. Optical properties and remote sensing of inland and coastal waters. CRC Press.

Cao, Z., Ma, R., Duan, H., Xue, K., 2019. Effects of broad bandwidth on the remote sensing of inland waters: Implications for high spatial resolution satellite data applications. ISPRS J. Photogramm., 153, 110-122.

Comber, A., Fisher, P., Brunsdon, C., Khmag, A., 2012. Spatial analysis of remote sensing image classification accuracy. Remote Sens. of Environ., 127, 237–246.

Cooley, S.W., Smith, L.C., Stepan, L., Mascaro, J., 2017. Tracking Dynamic Northern Surface Water Changes with High-Frequency Planet CubeSat Imagery. Remote Sens. 9, 1306.

Crist, E.P., 1985. A TM Tasseled Cap equivalent transformation for reflectance factor data. Remote Sens. Environ. 17, 301-306.

Daskalaki, S., Kopanas, I., Avouris, N.M., 2006. Evaluation of Classifiers for an Uneven Class Distribution Problem. Appl. Artif. Intell., 20, 381-417.

Dewi, R.S., Bijker, W., Stein, A., Marfai, M.A., 2016. Fuzzy Classification for Shoreline Change Monitoring in a Part of the Northern Coastal Area of Java, Indonesia. Remote Sens. 8, 190.

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 Sens. 8, 354.

ESRI., 2016. ArcGIS desktop: release 10.5, Environmental Systems Research Institute: CA, http://www.esri.com (accessed April 2nd 2020).

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861-874.

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 Sens. Environ. 140, 23-35.

Fisher, A., Flood, N., Danaher, T., 2016. Comparing Landsat water index methods for automated water classification in eastern Australia. Remote Sens. Environ. 175, 167-182.

Foody, G.M., Mathur, A., 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. Remote Sens. Environ. 103, 179-189.

Guo, Q., Pu, R., Li, J., Cheng, J., 2017. A weighted normalized difference water index for water extraction using Landsat imagery. Int. J. Remote Sens. 38, 5430-5445.

Huang, C., Chen, Y., Zhang, S., Wu, J., 2018. Detecting, extracting, and monitoring surface water from space using optical sensors: a review. Rev. Geophys. 56, 333-360.

Ireland, G., Volpi, M., Petropoulos, G., 2015. Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: A case study from a Mediterranean flood. Remote Sens. 7, 3372-3399.

Ji, L., Zhang, L., Wylie, B., 2009. Analysis of dynamic thresholds for the normalized difference water index. Photogramm. Eng. Remote Sens. 75, 1307-1317.

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J., Xiao, T., 2014. An automated method for extracting rivers and lakes from Landsat imagery. Remote Sens. 6, 5067-5089.

Kacprzyk, J., Fedrizzi, M., 2012. Multiperson decision making models using fuzzy sets and possibility theory. Springer Science & Business Media.

Karpatne, A., Khandelwal, A., Chen, X., Mithal, V., Faghmous, J., Kumar, V., 2016. Global Monitoring of Inland Water Dynamics: State-of-the-Art, Challenges, and Opportunities. In J. Lässig, K. Kersting, & K. Morik (Eds.), Computational Sustainability (pp. 121-147). Cham: Springer International Publishing.

Li, J., Sheng, Y., 2012. 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. 33, 5194-5213.

Li, L., Yan, Z., Shen, Q., Cheng, G., Gao, L., Zhang, B., 2019. Water Body Extraction from Very High Spatial Resolution Remote Sensing Data Based on Fully Convolutional Networks. Remote Sens. 2019, 11, 1162.

Li, X., Chen, W., Cheng, X., Wang, L., 2016. A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery. Remote Sens. 8, 514.

Lu, S., Wu, B., Yan, N., Wang, H., 2011. Water body mapping method with HJ-1A/B satellite imagery. Int. J. Appl. Earth Obs. 13, 428-434.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425-1432.

Ma, R., Duan, H., Hu, C., Feng, X., Li, A., Ju, W., Jiang, J., Yang, G., 2010. A half-century of changes in China's lakes: Global warming or human influence? Geophys. Res. Lett., 37, L24106.

McFeeters, S.K., 2013. Using the Normalized Difference Water Index (NDWI) within a Geographic Information System to Detect Swimming Pools for Mosquito Abatement: A Practical Approach. Remote Sens. 5, 3544-3561.

Ogashawara, I., Mishra, D.R., Gitelson, A.A., 2017. Chapter 1 - Remote Sensing of Inland Waters: Background and Current State-of-the-Art. In D.R. Mishra, I. Ogashawara, & A.A. Gitelson (Eds.), Bio-optical Modeling and Remote Sensing of Inland Waters (pp. 1-24). Elsevier

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, 418.

Planet Labs Inc., 2018. Planet Imagery Product Specifictions. San Francisco, CA.

Planet Team, 2017. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://www.planet.com/explorer/ (accessed on March 27, 2020).

Product Guide, 2018. Landsat 8 Surface Reflectance Code (LASRC) Product. Department of the Interior U.S. Geological Survey.

Rokni, K., Ahmad, A., Selamat, A., Hazini, S., 2014. Water Feature Extraction and Change Detection Using Multitemporal Landsat Imagery. Remote Sens. 6, 4173-4189.

Sánchez, G. C., Dalmau O., Alarcón T. E., Sierra B., Hernández C., 2018. Selection and Fusion of Spectral Indices to Improve Water Body Discrimination. IEEE Access. 6, 72952-72961.

Shao, Z., Fu, H., Li, D., Altan, O., Cheng, T., 2019. Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation. Remote Sens. Environ. 232, 1113-1138.

Sharma, R.C., Tateishi, R., Hara, K., Nguyen, L.V., 2015. Developing Superfine Water Index (SWI) for Global Water Cover Mapping Using MODIS Data. Remote Sens. 7, 13807-13841.

Smith, R.C., Baker, K.S., 1981. Optical-Properties of the Clearest Natural-Waters (200-800 Nm). Appl. Opt. 20, 177-184.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127-150.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., 2010. Global threats to human water security and river biodiversity. Nature 467, 555.

Wang, X., Xie, S., Zhang, X., Chen, C., Guo, H., Du, J., Duan, Z., 2018. A robust Multi-Band Water Index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. Int. J. Appl. Earth Obs. 68, 73-91.

Warmink, J. J., Janssen, J. A. E. B., Booij, M. J., Krol, M. S., 2010. Identification and classification of uncertainties in the application of environmental models. Environ. Model. Softw. 25, 1518-1527.

Wen, D., Huang, X., Zhang, L., Benediktsson, J.A., 2016. A Novel Automatic Change Detection Method for Urban High-Resolution Remotely Sensed Imagery Based on Multiindex Scene Representation. IEEE Trans. Geosci. Remote Sens. 54, 609-625.

Wen, Z., Yang, H., Zhang, C., Shao, G.F., Wu, S.J., 2020. Remotely Sensed Mid-Channel Bar Dynamics in Downstream of the Three Gorges Dam, China. Remote Sens. 12, 409.

Wu, W., Li, Q.Z., Zhang, Y., Du, X., Wang, H.Y., 2018. Two-Step Urban Water Index (TSUWI): A New Technique for High-Resolution Mapping of Urban Surface Water. Remote Sens. 10, 1704.

Xiong, L.H., Deng, R.R., Li, J., Liu, X.L., Qin, Y., Liang, Y.H., Liu, Y.F., 2018. Subpixel Surface Water Extraction (SSWE) Using Landsat 8 OLI Data. Water 10, 653.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025-3033.

Yang, Y., Liu, Y., Zhou, M., Zhang, S., Zhan, W., Sun, C., Duan, Y., 2015. Landsat 8 OLI image based terrestrial water extraction from heterogeneous backgrounds using a reflectance homogenization approach. Remote Sens. Environ. 171, 14-32.

Yang, X.C., Qin, Q.M., Grussenmeyer, P., Koehl, M., 2018. Urban surface water body detection with suppressed built-up noise based on water indices from Sentinel-2 MSI imagery. Remote Sens. Environ. 219, 259-270.

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3, 32-35.

Zhang, F.F., Li, J.S., Zhang, B., Shen, Q., Ye, H.P., Wang, S.L., Lu, Z.Y., 2018. A simple automated dynamic threshold extraction method for the classification of large water bodies from landsat-8 OLI water index images. Int. J. Remote Sens. 39, 3429-3451.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430-443.