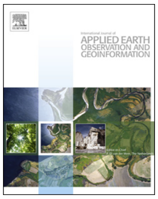
Int J Appl Earth Obs Geoinformation 91 (2020) 102159

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag



Surface water detection in the Caucasus

James Worden, Kirsten M. de Beurs*

Department of Geography and Environmental Sustainability, University of Oklahoma, United States

ABSTRACT

The Caucasus is an important global diversity hotspot and hosts a wide variety of surface water features, including major transboundary wetlands, in addition to large areas with irrigated agriculture and newly developed fishponds. In this study, we aim to establish the best performing methodology to produce surface water maps with a high degree of accuracy in the Caucasus. We evaluate optical data from Landsat 8 in both the dry and wet season for three study areas in the Caucasus. We test the performance of four different optical water indices derived from Landsat data, a method by Zou et al. (2017) also applied to Landsat data, and the European Commission Joint Research Centre (ECJRC) Global Surface Water dataset. We evaluate the performance of each water index using 5744 land cover validation/training points over all three study areas, which we manually classified by evaluating imagery from Google Earth. Using all validation points from all three study areas and both the wet and dry season, we find that the application of a logistic regression model using an optical surface water index (MNDWI) resulted in the most accurate open surface water maps. This approach achieved an overall accuracy of 93.0%, which is better than was found for freely available global surface water products.

1. Introduction

The importance of water, as it relates to living organisms on this planet, cannot be overstated. Water is the essential chemical compound that allows for the creation, existence, and propagation of the species that claim their residence on Earth. This fact especially holds true for the coexistence of human populations in rural and urban developments and the natural environment for the mosquito and transmission of the vector-borne disease of malaria. In 2017, the World Health Organization (WHO) reported 219 million malaria cases and 435,000 malaria-related deaths worldwide(Di Rocco et al., 1997). From the 1800s to around the first 25 years of the twentieth century, malaria was prevalent in Russia and the Soviet Union. It was especially dominant in the Caucasus region of Georgia, Armenia, and Azerbaijan (Bruce-Chwatt, 1959). In 1934, there were almost 600,000 reported incidences of malaria in Azerbaijan (Global Health Group, 2015). After the completion of the Global Malaria Eradication Campaign, there were no reported cases of malaria by 1960 (Mammadov et al., 2016). However, after the collapse of the Soviet Union, the number of reported malaria in cases rose again to 667 cases (in 1994) and jumped to 13,135 cases by 1996 (Di Rocco et al., 1997). After the outbreak of malaria was observed, control and prevention measures were subsequently implemented on a large scale between 1997 and 2011 (Di Rocco et al., 1997).

Malaria is a vector-borne disease that is caused by Plasmodium and distributed by Anopheles mosquito (Okorie et al., 2014). The Anopheles mosquitoes are reliant and limited to breeding habitats that contain standing water. Evaluating precipitation patterns, a relationship can be made between rainfall and mosquito abundance. For example, a numerical simulation has shown that 39% of simulated variances in the abundance of mosquito populations can be attributed to the pattern of intraseasonal rainfall (Bomblies, 2012).

Remotely sensed data are often used to identify surface water in a landscape (Cheng et al., 2006; Lacaux et al., 2007; Mueller et al., 2016; Tulbure et al., 2016; Work & Gilmer, 1976). To observe and measure the variability of the surface water, a water index and threshold can be established for the detection of water bodies and the calculation of the total surface area. Over the last 25 years, numerous remotely sensed indices have been developed for the detection of open surface water (Feyisa et al., 2014; Lacaux et al., 2007; McFeeters, 1996; Xu, 2006). These surface water indices typically use band math to emphasize the spectral reflectance characteristics of water features, while suppressing the characteristics of non-water features. Using a water index to delineate water features from other types of land cover typically requires the use of a cutoff or threshold value to establish the requirements a specific pixel needs to meet to be classified as water. Unfortunately, the performance of these water indices can vary spatially (Ji et al., 2009; Jiang et al., 2014; Kelly & Gontz, 2018). Different water indexes use different band combinations that can provide varying results, dependent on the distribution, type, consistency of surrounding land cover, and mixed pixel composition. Mixed pixels result from the presence of two or more land-cover classes contained in a smaller surface area as compared to the native resolution of the image. Ji (2009) found it necessary to evaluate different water detection indexes to determine the

---

[*] Corresponding author.
E-mail address: kdebeurs@ou.edu (K.M. de Beurs).

https://doi.org/10.1016/j.jag.2020.102159
Received 15 February 2020; Received in revised form 8 May 2020; Accepted 11 May 2020

method that performs best for the study region, as well as, the importance and need to ascertain an appropriate threshold for that region to segregate waterbody features from other land-cover types effectively. Nevertheless, in recent years, several global water datasets have been developed (Pekel et al., 2016).

The Caucasus is an important global diversity hotspot and hosts a wide variety of surface water features, including major transboundary wetlands, in addition to large areas with irrigated agriculture and newly developed fishponds. In this study, we aim to evaluate different spectral indices that delineate bodies of water from other topographical features in the Caucasus, to establish the most effective and accurate method for obtaining surface water area. Thus, different from other studies, we aim to capture predominantly human-made water bodies such as irrigation channels, making it especially important to distinguish these features from impervious surfaces. Besides evaluating different spectral indices, we will also compare our results with water maps from a globally available water dataset.

2. Study region

Georgia (69,700 km^2), Azerbaijan (86,600 km^2), and Armenia (29,743 km^2) make up the southern Caucasus (Transcaucasia), an agrarian, mountainous, region flanked by the Caspian and the Black Sea, with Russia to the north and Iran and Turkey to the south (Fig. 1). The region was part of the Soviet Union until its fall in 1991. The Greater Caucasus Mountain Range and the Lesser Caucasus Mountain Range create substantial elevation differences in all three countries, resulting in highly variable climates, ranging from cold alpine to humid subtropical.

Agriculture is an essential component of the economy in the Caucasus. For example, in Armenia where 62% of the land is allocated to agricultural production (80% of the crops irrigated), agriculture accounted for 19% of Armenia's gross domestic product (GDP) and employed 39% of the labor force (United States Agency International Development, 2017a). In Azerbaijan, 80% of agricultural production comes from irrigated lands, and agriculture employed 39.7% of the country's labor force in 2010 (World Bank Group, 2012c). In Georgia, the agricultural sector employs 52.3% of the country's labor force, which has remained relatively unchanged over time.

After the collapse of the Soviet Union, the agricultural regions in all three countries changed rapidly. For example, in Georgia, he croplands were re-distributed to rural families after 1992, andabout 80 to 90% of the newly created farms were less than 0.01 km^2 in size. After this redistribution about 23% was owned by private farmers, 10% leased to farmers, with the outstanding 67% held by the state. This process fragmented the agricultural sector leaving the irrigation infrastructure to these farms inefficient or insufficient in areas (World Bank Group, 2012a). In Armenia, the Agrarian reform and land privatization also fragmented large agricultural farms into 338,000 smaller farms. Here the area of irrigated lands was halved, also making irrigation and drainage systems unreliable (World Bank Group, 2012a). Poor management in an aging infrastructure led to the majority of the irrigation systems being in poor condition in Azerbaijan (World Bank Group, 2012c). The deterioration of the irrigation and drainage infrastructure in all three countries has led to soil pollution from pesticides and fertilizer, contributing to an increase in soil salinity in the region (World Bank Group, 2012c).

3. Data

3.1. Google Earth Training and Validation

We manually identified 1000 validation and training (30 m by 30 m) grid cells (600 land/400 water) for each of the three study areas using Google Earth images. Since we are explicitly interested in the detection of water, but water is a relatively rare class, we have developed a two-track validation approach. We first select 600 random points on the landscape. In the second step, we randomly select 400 points over areas identified as water according to the European Water Extent product, discussed below. For each of these stratified random samples, we visually evaluated the type of land cover, as well as the percentage of the land cover type that each point contained based on Google Earth imagery (5744 total points). We then randomly divided the classification points into 200 validation and 800 training points. The training points were used to develop the classification of open surface water, and the validation points to assess the accuracy of the water detection product. The majority of the Google Earth imagery was recorded after 2018; however, some imagery ranged back as far as 2008.

3.2. Landsat

We selected six level 2 Landsat 8 (OLI) satellite images from Earthexplorer (earthexplorer.usgs.gov) of the United States Geological Survey (USGS), at a spatial resolution of 30 m. These level 2 images are already atmospherically corrected, and we selected images that contained less than 30% cloud cover per scene. Three Landsat 8 path/rows, one for each country were used for this project (Table 1). For each country, we selected one image during July and one image during October, which coincides with a dry and wet period. We mask each image based on the provided cloud mask.

Based on extensive literature review, we selected the following water indices for evaluation (Table 2): the Normalized Difference Water Index (NDWI; McFeeters 1996), the modified Normalized Difference Water Index (MNDWI; Xu 2006), the Automated Water Extraction Index; the shadow and non-shadow (AWEIsh and AWEInsh; Feyisa et al. 2014), and a water classification algorithm based on both MNDWI and EVI/NDVI (Zou et al., 2017). The NDWI uses the reflected visible green (0.52-0.60 µm) and the near-infrared (NIR, 0.7-1.4 µm) wavelengths to enhance the spectral characteristics of surface water. Mcfeeters (1996) formulated the NDWI using the Band Ratio Parameter (BRP), defined as subtracting the NIR band from the green band and dividing the difference by the sum of the two bands. This combination of the NIR and green wavelengths gives positive values to water features and zero or negative values for vegetation and soil when applied to a multispectral satellite image (McFeeters, 1996).

Xu (2006) found that NDWI enhanced water features in multispectral satellite images but could not effectively suppress built-up land features present in the scene. To address this issue, Xu (2006) introduced the modified NDWI (MNDWI). An examination of the spectral reflectance patterns for built-up land, vegetation, and lake water exhibited a higher average reflectance in the shortwave infrared (SWIR,1.55-1.75 µm) band range when compared to the green band. The MNDWI uses the same form as the NDWI but instead uses the SWIR band in place of the NIR band to increase the enhancement of open water features. Despite being resilient to built-up land features, the MNDWI has difficulties in distinguishing low albedo urban surfaces and shadows from water (Feyisa et al., 2014). To further suppress the misclassification of water from shadowed and other non-water surfaces, the automated water extraction index (AWEI) was introduced. As a result, two different equations are presented to improve the accuracy of water extraction, while suppressing non-water pixels: AWEIsh and AWEInsh. The AWEInsh index was created to reduce the confusion between water and non-water pixels, as well as dark surfaces included in urban background areas. The subscript "nsh", or non-shadow, identifies the index is best suited for areas where shadows are not present or pose no problems for classification (Table 2).

In some cases, the AWEInsh equation may not effectively eliminate all shadow types and surfaces of low albedo. Surfaces that contain shadows exhibit low reflectance patterns over the spectral range and also vary in the magnitude of this low reflectance as the surface characteristic changes. With these limitations in mind, the second equation,

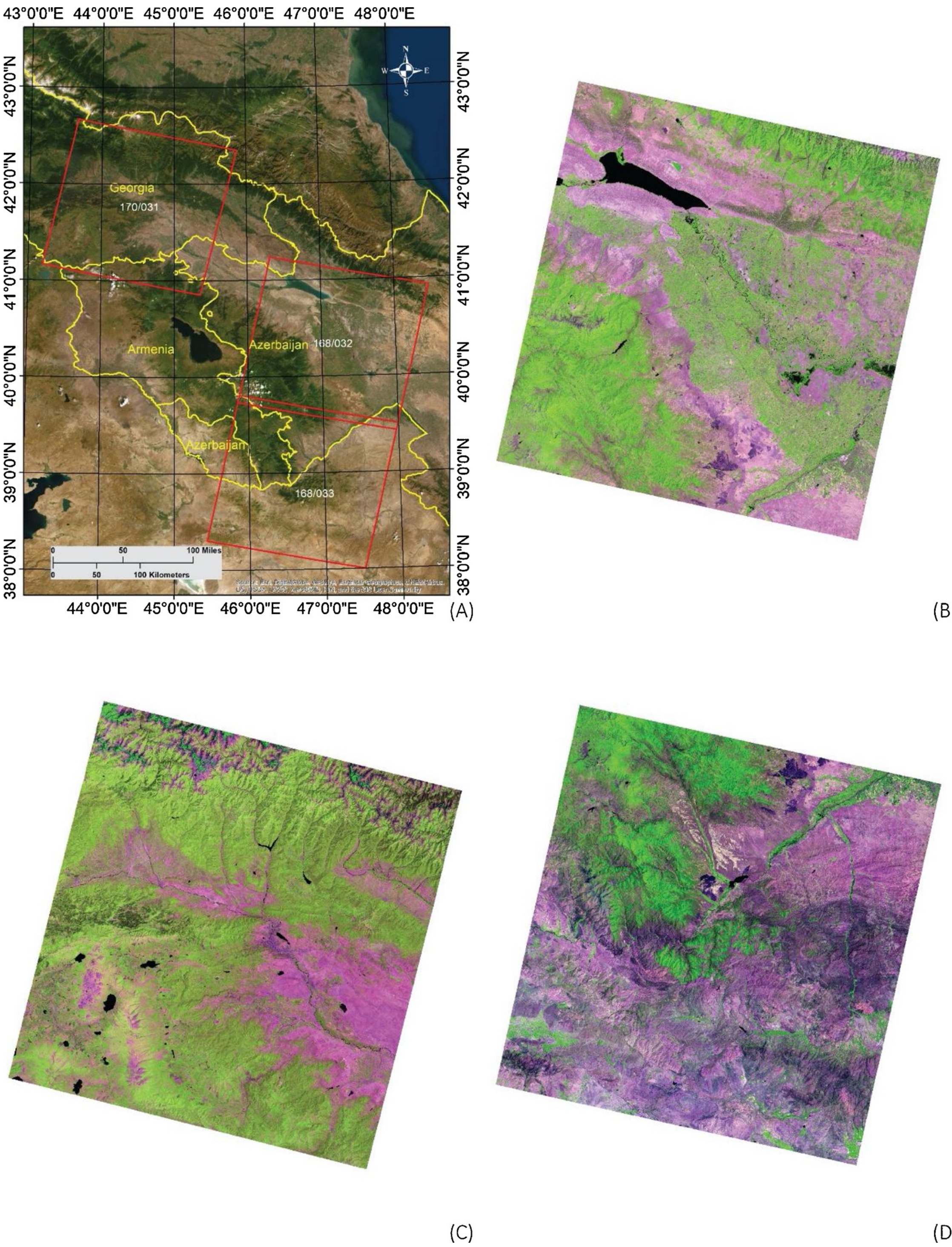


Fig. 1. (A) Overview image of the Caucasus study region, (B) the Azerbaijan Landsat 8 tile 168/032 (6,5,7 RGB), (C) the Georgia Landsat 8 tile 170/031 (6,5,7 RGB), (D) and the Armenia Landsat 8 tile 168/033 (6,5,7 RGB).

Table 1
Summary of path/row and acquisition date of Landsat 8 OLI images from the three study areas located in Armenia, Azerbaijan, and Georgia.

	Georgia	Armenia	Azerbaijan
Landsat 8 Path/Row	170/031	168/033	168/032
Landsat 8 - Date 1	07/03/2017	07/28/2017	07/21/2017
Landsat 8 - Date 2	10/07/2017	10/25/2017	10/25/2017

"AWEIsh" was formulated to increase the separability of water from shadows and other dark surfaces. The subscript "sh", or shadow, indicates that the index is designed to improve the accuracy of water feature extraction in scenarios where significant sources of shadow (urban/mountain) are present (Table 2). The AWEIsh may misclassify surfaces as water in areas that contain highly reflective surfaces, including snow, ice, and urban locations with reflective roofs (Feyisa et al., 2014). Zou et al. (2017) use a combination of the MNDWI and vegetation indices, the Enhanced Vegetation Index (EVI) and the Normalized Difference Vegetation Index (NDVI), to reduce the potential

Table 2

Explanation of band combinations based on Landsat 8 OLI data, and potential threshold values from previous studies used to extract water pixels from images

Water Index	Equation	Potential Threshold Values
NDWI McFeeters (1996)	$\frac{(Green - NIR)}{(Green + NIR)}$	0 – 0.337[a] 0.015 – 0.017[b] -0.19 – 0[c] 0.3877[e] -0.21[f]
MNDWI Xu (2006)	$\frac{(Green - SWIR1)}{(Green + SWIR1)}$	0 - 0.09[a] -0.05 - 0.06[c] 0.005 – 0.6[d] 0.35[e] 0.00[f]
AWEIsh Feyisa et al. (2014)	$Blue + 2.5 \times Green - 1.5 \times (NIR + SWIR1) - 0.25 \times SWIR2$	−0.15 - 0.045[d] -0.1 - (-0.03)[c] 0.1112[e] -0.02[f]
AWEInsh Feyisa et al. (2014)	$4 \times (Green - SWIR1) - (0.25 \times NIR + 2.75 \times SWIR1)$	−0.15 - 0.045[d] -0.1 - (-0.02)[c] 0.1897[e] -0.07[f]
Water Classification Algorithm Zou et al. (2017)	[(MNDWI > NDVI or MNDWI > EVI) and (EVI > 0.1)]	n/a

a: (Xu, 2006); **b:** (Ji et al., 2009); **c:** (Jiang et al., 2014); **d:** (Feyisa et al., 2014); **e:** (Acharya et al., 2018); **f:** (Fisher et al., 2016)

commission error of vegetation over wet surfaces. This algorithm requires the MNDWI value of the pixel to be higher than the value of NDVI or EVI and contains an EVI value of less than 0.1 to be classified as water (Table 2).

3.3. European Water Maximum Extent Dataset

We also evaluate the Global Surface Water Explorer (https://global-surface-water.appspot.com/) developed by the European Commission's Joint Research Centre (ECJRC) as it is a freely available global dataset. This water dataset was derived from the entire inventory of the Landsat 5 Thematic Mapper (TM), the Landsat 7 Enhanced Thematic Mapper-plus (ETM+), and Landsat 8 Operational Land Imager (OLI) brightness temperature and top-of-atmosphere reflectance, orthorectified images, that were obtained from March 16, 1984, to October 10, 2015 (Pekel et al., 2016). The Global Surface Water dataset contains several different water metrics. For this study, we selected the maximum water extent product, which renders information on all regions that have ever detected water over the course of the 32-year period and the water monthly history product, which aggregates this data on a monthly scale. The dataset was created by applying an expert system classifier that segregates the pixel to one of three target classes, water, land, and non-valid. This system establishes the parameters to classify bodies of water by using a decision tree based on rules containing a dichotomous conditional and inference framework. The equations describing the conditional parameters are built upon a spectral library, established from visually evaluated samples of 64,254 points over 9,149 Landsat scenes, and enhanced by deriving the NDVI and a standard color-space transformation of the Hue-Saturation-Value (HSV) for the SWIR2 (2.08 – 2.35 μm), NIR, Red and NIR/Green/Blue band combinations. Evaluating the order of reasoning applied by the inference engine gives insight into the factors that lead to pixel classification. This allows for the identification, correction, and improvement of the shortcomings inherent in the developed evidential reasoning through an iterative process. This process is repeated until the system can be improved no further, and subsequently applied to the entire Landsat dataset.

3.4. Shuttle Radar Topography Mission

The Shuttle Radar Topography Mission (SRTM) is a high-resolution digital elevation model (DEM) available for 80% of the Earth's land surface. This data was acquired from a collaborative effort by NASA, the German and Italian Space Agencies, and the National Geospatial-Intelligence Agency (NGA) over a ten-day operational flight period of 149 orbits in February of 2000 (Farr et al., 2007). The SRTM utilized C band (5.6 cm) and X band (3.1 cm) synthetic aperture radar systems, designed to function as a single-pass interferometers. With a 1 arc sec by 1 arc sec sampling resolution, the SRTM Data were to be sampled with a linear vertical absolute/relative height error of fewer than 16 m/10 m and a circular absolute/relative geolocation error of less than 20 m/15 m (Farr et al., 2007). We use the DEM dataset to create a slope raster for the elimination of confusion between water pixels and pixels containing mountain shadows.

3.5. Global Man-made Impervious Surface Dataset

The Global Man-made Impervious Surface (GMIS) data set is created from the freely available satellite imagery from the 2010 Global Land Survey (GLS) (Brown de Colstoun et al., 2017). To mitigate the spectral confusion of urban areas with bare surfaces and fallow lands, a non-urban mask is created from nightlight data and other sources of ancillary data (Homer et al., 2004). To produce training data to train the cubist regression tree algorithm, 1,800 high-resolution scenes were obtained from the NGA WARP (Web-based Access and Retrieval Portal) website to form relationships between the spectral values of Landsat data and subpixel imperviousness (Brown de Colstoun et al., 2017). This dataset is provided in GeoTiff raster files containing two bands: percent imperviousness and standard error in percent imperviousness. This data set is used to create an impervious surface raster mask to eliminate urban pixels that may be misclassified as water.

4. Methods

The goal of this study is to evaluate the European Monthly History Water dataset, as well as a variety of water indices developed in the literature to establish which index performs best for the Caucasus region. Also, we present an alternative method for the detection of surface water in this region. We validate all datasets against validation points selected based on Google Earth images. A mask is applied to omit noise originating from cloud and snow pixels from the logistic modeling process. The snow/cloud mask is then applied along with the urban/shadow mask to remove the misclassification of these phenomena from the final surface area calculation. Fig. 2 provides an overview of the applied methodology.

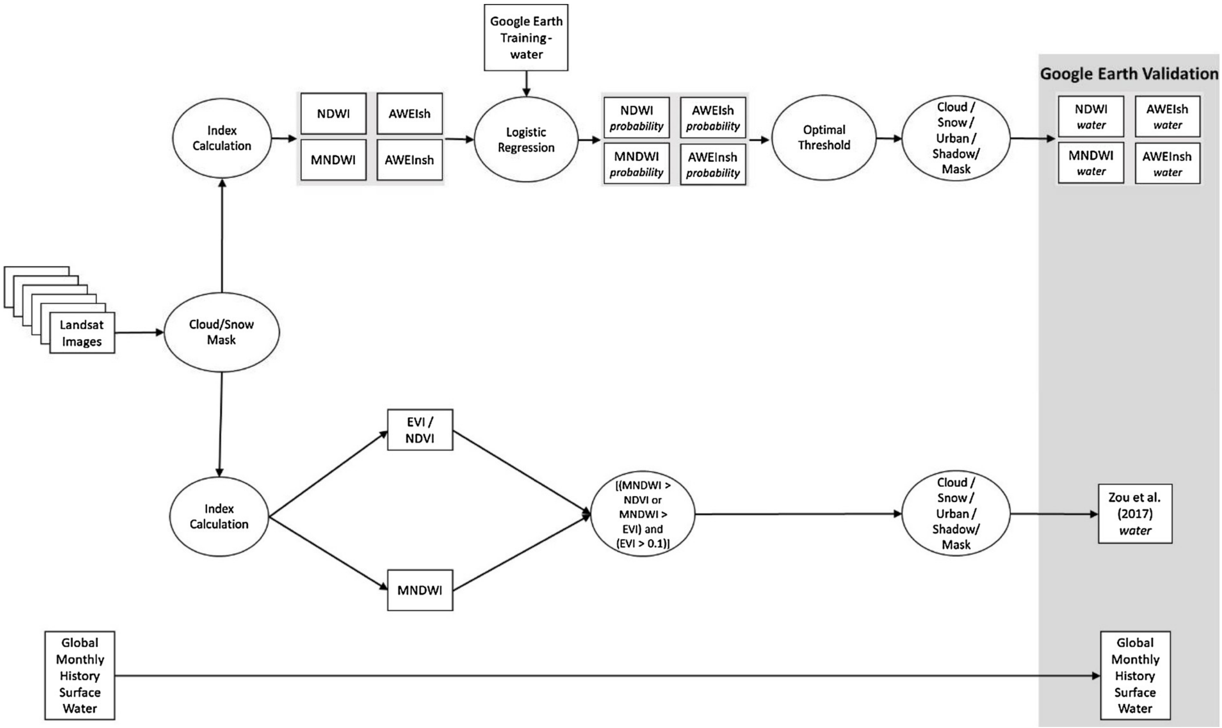


Fig. 2. This figure displays the workflow process for the extraction of surface water body area from water index probability maps. For consistency and continuity, the Cloud/Snow/Urban/Shadow mask was applied when comparing the surface water area of Zou et al. 2017 and the probability maps for the wet/dry seasons.

4.1. *Logistic Regression*

We apply a logistic regression model to link the Google Earth identified water points with the satellite-derived water indices. Logistic regression models have been widely used to provide a probabilistic classification of land cover type based on values of a given set of predictors derived from remotely sensed data (Alsharif & Pradhan, 2014; Cheng et al., 2006; Lee, 2005; Mueller et al., 2016). Others stated that the logistic regression performs well for testing hypotheses and describing the relationship between predictor variables and a categorical outcome variable (Peng et al., 2002). The variables used in logistic regression may be discrete, continuous, or a combination of both classes, and a normal distribution of the data is not required (Lee, 2005). The probability equation for the logistic regression(1) is described as follows:

$$p = \frac{1}{1 + e^{-(bx+a)}} \tag{1}$$

Where b represents the slope value of the index, x represents the index point value, which in our case is the derived water index, a represents the y-intercept value, and p represents the resulting probability value, which provides the probability that an individual pixel contains water.

4.2. *Optimal Probability Cut-Off*

To derive a binary water surface map from a spectral water index, a threshold (or cut off value) is required. The selection of a threshold is critical to the performance of a selected index's ability to identify the land cover feature of interest. A threshold set to a large value has the propensity of creating a product that has an increased omission error and resulting in the identification of a smaller magnitude of the number of water bodies detected. A threshold set to a small value will result in a larger number of water bodies identified in the scene but also is more likely to misidentify land features as water. As Table 2 demonstrates, selecting the threshold value to distinguish between water and other land surfaces is not straightforward, and a constant value cannot be used for all scenes due to the variations of the temporal and subpixel land-cover components in the scenes. This changing of physical and temporal characteristics between scenes makes the threshold value dynamic in nature, making it necessary to evaluate and establish the threshold that is appropriate for the region of study (Ji et al., 2009). There have been several methods used to determine a threshold that provides a cut-off parameter to minimize the commission and omission errors inherent in segregating land cover into classes. Jiang et al. (2014) applied a series of water index thresholds values from -0.1 to 0.1 in increments of 0.01, evaluated via visual inspection and pixel by pixel assessment of the image, along with using high-resolution images from Google Earth$^{\text{TM}}$ as a complementary reference in order to assist in distinguishing confusing water pixels from mountain shadows or urban areas.

In the modification of the normalized difference water index, a series of thresholds tested manually were adjusted and evaluated by confusion matrices produced with a pixel-by-pixel comparison between the predicted and reference images (Xu, 2006). Others also used a "trial and error" method, stepping through a series of thresholds with 0.05 increments between -1 and 1, to find the highest overall accuracy and kappa (Acharya et al., 2018).

The AWEI water extraction index considered multiple thresholds by calculating commission and omission errors and plotting the percentage errors against threshold values (Feyisa et al., 2014). The commission and omission errors were established by comparing the reference data by manual digitizing multiple polygons evenly distributed on true-color composites of Landsat bands to the threshold values selected. The intersection point of the commission and omission errors was regarded as the optimal threshold value for the index in the study area (Feyisa et al., 2014). From the water delineation indices previously described, the method for determining the most accurate threshold for these water detection studies involves a manual pixel-by-pixel or use of manually digitized polygons to evaluate and assess the performance (or overall accuracy) of the selected water index. This process works well for studies that involve evaluating a hand full of images from different areas. Unfortunately, this method proves inefficient when applying such a technique of thresholding to large-scale regions of study across time series on a decadal scale, due to the temporal and physical variations between the areas.

To compensate for the variability of thresholds between scenes, we propose that the threshold selection process should represent the dynamism of variability inherent in each scene. This can be achieved by allowing for the temporal and physical characteristics present in each scene to determine its own threshold value. In general, there are two statistical approaches used to approach the issue of identifying and applying an optimal threshold, the use of the Receiver Operator Characteristic (ROC), and the application of an appropriately selected statistical test (López-Ratón et al., 2014). In this study, we will be focusing on the former. The optimal probability cut-off (OPC) is an iterative process that applies all possible probability cut-off values to the actual and predicted values of the selected index to determine the cutoff value that produces the highest degree of accuracy for classifying water that best fits the regression model established by the validation points. From the resulting OPC value, a comparison of accuracy can be made between the different water indices

4.3. Cloud, Urban areas, and Mountain Shadow Removal

The loss of data due to cloud cover is a natural and unavoidable phenomenon woven into the nature of optical satellite remote sensing. To ameliorate this issue, we remove the clouds present in the scene by applying a cloud mask derived from the pixel QA raster included with the Landsat image. While this method eliminates the data obscured by clouds, the benefit is that the non-selective scattering characteristic of the cloud is not misclassified as water.

Each of the water detection methods applied in this region struggled with the misclassification of water pixels due to the presence of built-up, mountain shadow, and snow cover in the scene. To assist in the removal of mountain shadows in the image, the SRTM digital elevation model was used to create a slope value raster of the region. To effectively remove unwanted mountain shadow from the scene, areas with a slope of more than ten percent were masked out. The distribution and composition of urban land cover vary between locations. For this reason, the Global Man-made Impervious Surface (GMIS) dataset is used. Pixels with more than ten percent impervious surface, according to the GMIS dataset, were masked out.

4.4. Evaluation of Index Performance

To compare the performance of the different water indices, we calculated accuracy measures for each season (wet/dry) and each research area using the Google Earth validation points. The ROC, sensitivity, specificity, concordance, and overall accuracy for all water detection methods and areas are evaluated and compared to determine the best performing index for the region. The ROC curve is a visual representation of classifier performance, derived from plotting the false positive rate (FPR) against the true positive rate (TPR). This curve establishes the boundaries of a test's capability to differentiate between two different states (Zweig & Campbell, 1993). Others disclosed the advantage of using a ROC curve to measure the performance (accuracy) of the different water detection methods (Swets, 1988). To assess the discriminative ability for a model of logistic regression, a concordance statistic is also calculated. The concordance statistic is a unit-less index that indicates the likelihood that a pixel selected at random that has water present will have a higher value of predicted probability, as opposed to a randomly selected pixel that does not contain water (Harrell, 2015). This value is derived from the proportion of the pairs of pixels that contain water and their predicted probability, with the pairs of pixels and probability values that do not contain water (Austin & Steyerberg, 2012). A concordance result with the value of one represents the quality and reliability of a perfect classification model. The TPR (also known as sensitivity) is a ratio that is calculated by dividing the number of true positives (TP the number of correctly classified positive) by the sum of the true positives and false negatives (the number of incorrectly classified positive states). Sensitivity is the measure of the accuracy of an index for water detection. The value range of sensitivity is between 0 and 1. The closer the value is to one, the higher the magnitude of confidence in the actual existence of water in that pixel (Fawcett, 2006). Specificity is a ratio calculated by dividing the number of the true negatives (the number of correctly classified negative states) by the sum of the true negatives and false positives (the number of incorrectly classified negative states). The values range from 0 to 1, and the closer the value is to 1, the higher the confidence in the actual non-existence of water in that pixel (Fawcett, 2006). The overall accuracy of the individual models tested is calculated by summing the number of true positive and true negative classification points and dividing by the total sum of all classification points. The result is then multiplied by 100, and the value obtained is the overall percent accuracy of water classification for the specific detection method. To determine the highest performing water index, each method was evaluated individually for their performance in each category. We assigned a value of one to the highest performing index and a value of zero to all other indices. The index with the highest summed value across all metrics will be determined as the best performing water detection model for use in the Caucasus region.

5. Results

We have evaluated four different optical water indices (NDWI, MNDWI, AWEIsh, AWEInsh, and the method described in Zou et al. (2017) to identify water in the three different countries that make up the southern Caucasus region. We also have evaluated the performance of these indices on a seasonal scale (July vs. October) and by combining all results within one year. Table 3 displays the results from the logistic regression, consisting of slope, intercept, and optimum probability cut-off for the ECJRC Monthly History Water dataset, water data based on the method by Zou et al. 2017, and each of the water detection indices. Below we will first provide the overall accuracy by country and month for each of the water indices. We then present results on the overall estimated water surface based on the European Monthly History Water dataset from the ECJRC. We will also describe some general errors when detecting water with optical indices.

5.1. Water index accuracy by seasons and country

Based on the Google Earth selected random validation points, we determined the overall accuracy for each water index by season and country (Table 4). Apart from the method described by Zou et al. (2017), we found that the overall accuracy was relatively high for all countries and all indices. Generally, the indices appeared to perform best for Azerbaijan, with overall accuracy above 95% for all indices,

Table 3

Results of Logistic Regression and the Optimal Probabiliy Cut-off values (OPC). An OPC of 0.227 (NDWI), indicates that a water present/absent map is most accurate if any pixel with a probability of more than 0.227 is set to water. Note that the probability cut-off for the European Monthly History Water dataset is close to 0, since any probability that water is detected should be selected as water present.

	Slope	Intercept	Optimal Probability cut-off (OPC)
European Monthly History	4.911	−3.014	0.049
NDWI	6.975	−0.095	0.227
MNDWI	5.864	−0.765	0.182
AWEIsh	9.708e-04	6.642e-01	0.527
AWEInsh	4.938e-04	1.091	0.419
Zou et al. 2017	3.357	−1.568	0.177

Table 4
Overall accuracy for each water index by season and country.

Annual	Georgia	Armenia	Azerbaijan	All Countries
European Monthly History	**89.4%**	**90.7%**	**96.1%**	**92.2%**
July	**88.6%**	**88.9%**	**96.4%**	
October	**90.5%**	**92.6%**	**95.9%**	
MNDWI - Annual	**90.9%**	**93.0%**	**95.7%**	**93.0%**
July	*91.8%*	*92.4%*	*96.4%*	
October	*90.6%*	*94.8%*	*94.9%*	
AWEIsh - Annual	**90.6%**	**93.6%**	**96.3%**	**93.3%**
July	*91.8%*	*92.4%*	*96.4%*	
October	*90.0%*	*94.8%*	*96.1%*	
AWEInsh - Annual	**90.1%**	**93.6%**	**96.5%**	**93.1%**
July	*91.3%*	*92.4%*	*96.4%*	
October	*88.9%*	*95.4%*	*96.6%*	
NDWI - Annual	**90.9%**	**92.4%**	**94.4%**	**92.3%**
July	*91.8%*	*91.3%*	*94.4%*	
October	*90.0%*	*93.6%*	*94.9%*	
Zou et al. 2017	**79.9%**	**78.8%**	**63.0%**	**78.0%**
July	*82.5%*	*77.8%*	*63.0%*	
October	*77.2%*	*80.0%*	*63.5%*	

and worst for Georgia, with all indices revealing an overall annual accuracy of slightly greater than 90%. The method described by Zou et al. (2017) exhibited the weakest performance out of all tested indices (overall accuracy 78%). This method performed the best in Georgia (accuracy 79%) and the worst in Azerbaijan (accuracy 63%) expressing an inverse trend of accuracy compared to the other tested indices in individual countries.

Using the OPC value, we calculated accuracy percentages for each index. To better understand the under and overestimations that make up the overall accuracy, we have calculated these percentages for each index (Table 5). The ECJRC Monthly History Water dataset reveals an underestimation of water pixels, underestimating 15.7% of the water validation points. All the evaluated indices tended to overestimate water pixels, with MNDWI having the most significant percentage of overestimation at 11.6% and AWEIsh the lowest percentage at 9.0%. The water detection method described in Zou et al. 2017 had the highest percentage of underestimated water pixels at 23.2%, while AWEInsh recorded the lowest percentage of underestimated water pixels at 5.1%, based on the collected Google Earth validation data.

The difference in overall accuracy between the four methods and the ECJRC Monthly History product was less than 1% (MNDWI, AWEIsh, AWEInsh, and NDWI). Thus we evaluated other performance metrics of sensitivity/specificity, concordance, and ROC area (Table 6). Evaluating each method by study area on an annual scale and using the other performance metrics (Table 6), we determine that the MNDWI performed the best in Georgia, Armenia, and Azerbaijan. We found that the sensitivity, ROC area, and concordance were highest for the MNDWI index.

5.2. Water surface area

To understand the total surface area estimates, we calculated the total water surface for each index and each country and season (Table 7). The estimation of surface water area between the four water index probability maps was reasonably consistent when compared to each other. With few exceptions, the MNDWI (OPC) map generated the largest surface area values when applied to all tiles for both periods. In contrast, the AWEIsh (OPC) map produced the lowest surface water area over the study region tiles. The NDWI probability water map, however, greatly overestimated the area of surface water in the Armenia tile in both wet/dry periods. The overestimation resulted from the NDWI index having difficulty in differentiating between snow and water pixels in the Armenia tile.

Table 5
Percentages of over and underestimation of validation water pixels.

All Areas Annual	Overall accuracy	Under Estimated	Over Estimated
European 2017 Water Map	92.2%	15.7%	4.27%
MNDWI	93.0%	5.2%	11.6%
AWEIsh	93.3%	5.7%	9.0%
AWEInsh	93.1%	5.1%	10.7%
NDWI	92.3%	6.6%	10.3%
Zou et al. 2017	78.2%	23.2%	10.9%

Table 6
Other performance metrics to evaluate water index performance.

All Areas Annual	Overall Accuracy	Sensitivity	Specificity	Concordance	ROC Area
European Monthly Water History	92.2%	0.843	0.957	0.807	0.900
NDWI	92.3%	0.850	0.956	0.949	0.949
MNDWI	93.0%	0.885	0.950	0.957	0.957
AWEIsh	93.3%	0.870	0.962	0.953	0.952
AWEInsh	93.1%	0.885	0.952	0.938	0.935
Zou et al. 2017	78.2%	0.850	0.956	0.327	0.658

When comparing the OPC surface water estimation to the ECJRC Monthly History water map, we observe an underestimation of surface water pixels detected. Across all study regions and periods, the ECJRC water map identified less water than the probability index maps. This underestimation of the surface water area is consistent with the percentage of underestimated validation points, as described in Table 5.

The method described by Zou et al. 2017 underestimated the area of surface water bodies in all tiles in both wet/dry periods as compared to the other selected water detection methods. This result is also consistent with the method's tendency to underestimate the percentage of total water pixels in the study area (Table 5).

Evaluating the OPC method, we find that there are strengths and weaknesses between the performance of the different water probability maps. Fig. 3 displays the different water detection capabilities between the indices and the ECJRC Monthly History water map. The ECJRC Monthly History water map detects the least amount of surface water in this scene, while all water probability maps using an OPC prove more successful in identifying the water present in the canal. The MNDWI water probability map generates a more complete representation of the surface water area as compared to the other water index probability maps.

Urban areas and mountain shadows pose a threat to the accuracy of water index detection methods, as these areas can have positive water index values causing a misclassification of the water features presents in the scene. Figs. 4 and 5 highlight the misclassification in areas that contain urban and mountain shadow land features. In Fig. 4, we see that the best performing probability index, as it relates to urban misclassification, is the AWEInsh water map. The other index water probability maps had varying levels of success with the misclassification of urban areas, with the NDWI water probability performing the poorest and the AWEInsh water probability performing the best out of the set. However, the success of the AWEInsh probability water map does not carry over to areas containing mountain shadows. Fig. 5 displays the performance of the water index probability maps when applied to a region Northwest of Tbilisi, Georgia. In this location, the NDWI probability index performed the best, while the AWEInsh performed the worst as it relates to mountain shadow classification.

Evaluating Figs. 3–5, we observe variation in the performance of the water index probability maps depending on the landcover type present in the scene. To reduce the misclassification of built-up urban areas, an urban area mask is utilized. Fig. 6 shows the water index probability

Table 7
Surface water area using OPC value and the ECJRC Monthly History water map, from each of the two periods (wet/dry) over three study areas. A Pixel QA raster snow/cloud mask was applied to all Landsat tiles to remove pixels containing clouds from the final area calculation.

		Georgia (km^2)	Armenia (km^2)	Azerbaijan (km^2)
European Water dataset	Monthly Water History	210.1 / 212.5	65.0 / 51.5	644.0 / 610.0
Landsat 8	NDWI	233.2 / 238.7	143.4 / 186.3	742.0 / 644.4
Landsat 8	MNDWI	251.9 / 233.5	82.9 / 62.4	744.7 / 623.5
Landsat 8	AWEIsh	222.4 / 217.2	66.0 / 50.9	663.9 / 570.2
Landsat 8	AWEInsh	237.0 / 256.7	72.3 / 58.0	735.0 / 619.0
Landsat 8	Zou et al. 2017	44.3 / 99.3	37.9 / 28.5	552.7 / 456.4

maps before and after the urban areas mask was applied.

Many studies have used water indices with other methods to reduce the misclassification of shallow water bodies (Jiang et al., 2014; Tulbure et al., 2016; Zou et al., 2017). This study focuses on the unique characteristics of water that are lost by the derivation of a generic water index. To establish a more meaningful relationship between surface water in the region, and the water index values that are generated, the specific spectral and environmental characteristics of the water must be considered. Differences in atmospheric conditions, the variation of the incident angle of the sun, and changes in the chemical and biophysical properties of the water can influence the pattern of reflected wavelengths exhibited by surface water bodies (Zhang et al., 2010). Applying a logistic regression to the water index incorporates the specific spectral properties of water in that region into a range of probabilistic values that are as unique as the environmental conditions in which they exist; resulting in more separation between the values of vegetated and water land cover types (Fig. 7).

Looking at the spectral reflectance tendencies of water, MNDWI values of water are generally greater than zero (Xu, 2006). The equations for the AWEI products use coefficients in order to force water pixel values to be positive and negative values pixels for non-water pixels, allowing for an initial threshold value of zero for varying environmental conditions (Feyisa et al., 2014). To compare the effectiveness of the water probability maps with OPC value to the selected water indexes, a default threshold of zero was chosen in Fig. 8 for the extraction of surface water. Despite having a value of zero for the threshold, the MNDWI, NDWI, AWEIsh, and AWEInsh had difficulties extracting the water pixels in the scene. Comparing the results of the water probability map and the water index maps, the OPC value water map exhibits a more complete representation of the water canal's location and surface water area, than the traditional thresholding of the raw water index values. The inability of these indices to effectively detect the water in this canal by using a default threshold of zero is due to the water in this canal having negative index values.

To accurately extract the water present in the canal, an adjustment of the water threshold value is required. After applying the lowest potential threshold implemented in the previously discussed studies for the selected water index (Table 2), the water indices present difficulties in detecting the water present in the canal (Figure 9). Comparing the results of the water probability map and the water index maps, the OPC value water map exhibits a more complete representation of the water canal's location and surface water area, than the traditional thresholding of the raw water index values.

6. Discussion

Establishing a threshold for any type of land cover detection method is complicated. Selecting a threshold can be a time consumptive process and may lead to a somewhat subjective decision of the derived threshold value that is to be used (Ji et al., 2009). This issue stems from the lack of a stable threshold value for water across different types of water bodies in different locations (although see Fisher et al., 2016). Others found that stable spectral profiles can be found in deep clean bodies of water while becoming unstable in shallow/narrow water bodies (Sivanpillai & Miller, 2010). Once an appropriate threshold is determined, the threshold value is only representative of the conditions temporally present in that scene but does not necessarily hold true when applied to different scenes in different conditions at different times. Ji et al. (2009) suggested that an adjustment of the threshold value, to suit the conditions present better, could improve the extraction results of the water index. However, these adjustments would be difficult in the automation of a time series that requires a manual adjustment of the threshold between images (Huang et al., 2018). Zou et al. 2017 addressed this issue by leveraging vegetation indices values (NDVI and EVI) against a water index (MNDWI) to alleviate the need to establish a static threshold for water delineation. However, an EVI threshold 0.1 was applied to exclude noise from vegetated wetland pixels. Zou et al. (2017) found an overall water accuracy of 94% derived from a confusion matrix containing 3216 water and 6726 non-water ground reference pixels. Our analysis reveals that this high accuracy is location-specific as a result of the conditions inherently found in the study area. The study area of Oklahoma has more human-made lakes than any other state and has 55,646 miles of shoreline among the lakes and ponds (Oklahoma Water Resources Board, 2018). While the method works well in Oklahoma, the validity of this method deteriorates when applied to water bodies in the Caucasus that contains vegetation, resulting in an overall accuracy of just 78%. Our method of selecting an OPC value from a logistic regression can allow for an automated adjustment of the cut-off threshold values based on the probability which minimizes the misclassification of the confusion matrix.

Tulbure et al. (2016) employed the use of multiple explanatory variables in a random forest classification tree algorithm. These explanatory variables consisted of all top of atmosphere (TOA) bands, selected water indexes (NDWI, MNDWI, AWEI), vegetation indexes (NDVI, EVI), brightness temperature, as well as slope and hillshade datasets derived from the SRTM digital elevation model. The resulting water surface product produces an overall classification accuracy of 99.9%. The validation of the proposed surface water dynamic product was evaluated based on a stratified random sampling of 500 water/non-water points(Tulbure et al., 2016). The water stratum employed was derived from a water body map from Geoscience Australia (2006) and a maximum extent water map where water was observed at some point during a 26-year period. The high degree of accuracy found by Turbule et al. (2016) might partly be the result of having a low amount of reference water pixels (0.5%) as compared to the non-water pixels (99.4%) used in the confusion matrix. To avoid a small number of training/validation water pixels, we used a stratified random sampling of 400 water and 600 non-water pixels for each study area based on a global maximum water extent product (Pekel et al., 2016). Once completed, each sampled pixel was then classified manually using Google Earth data. After the removal of pixels obscured by cloud and snow, the validation of the logistic regression model was based on 1095 total (756 non-water, 339 water) validation points.

As with the methods described (Zou et al., 2017l; Tulbure et al., 2016), the monthly water map from the ECJRC boasted a high degree of accuracy when evaluated by statistical performance metrics. However, despite having a high degree of accuracy, the performance of the ECJRC Monthly History water map's performance can vary significantly in the

Fig. 3. This figure displays the difference in water detection ability based on the OPC maps of a canal just East of the city of Agjabedi, Azerbaijan. (A) displays the water canal in Agjabedi, Azerbaijan. (B) the ECJRC Monthly History water map. (C) MNDWI water probability map. (D) NDWI water probability map. (E) AWEIsh water probability map. (F) AWEInsh water probability map.

detection in the type of water body being evaluated. In Figs. 3 and 9, we observe that the ECJRC Monthly History water product retains poor performance when detecting the water present in the water canals in Azerbaijan. The misclassification of water in Figs. 3 and 9 are consistent with the differences in surface water area detection in Tables 5 and 7. We must apply caution with the use of statistical performance metrics when evaluating the effectiveness of a water detection method. A high degree of water detection accuracy is excellent, but it means very little if the method does not detect the landcover type the technique was designed to identify.

In addition to the misclassification errors of built-up areas and mountain shadows experienced by our proposed method, floating vegetation in water bodies has also provided a source of misclassification error in the extraction of surface water area. Variation in floating aquatic vegetation can shift the surface reflectance of wetland areas, causing surface water containing a large percentage of floating

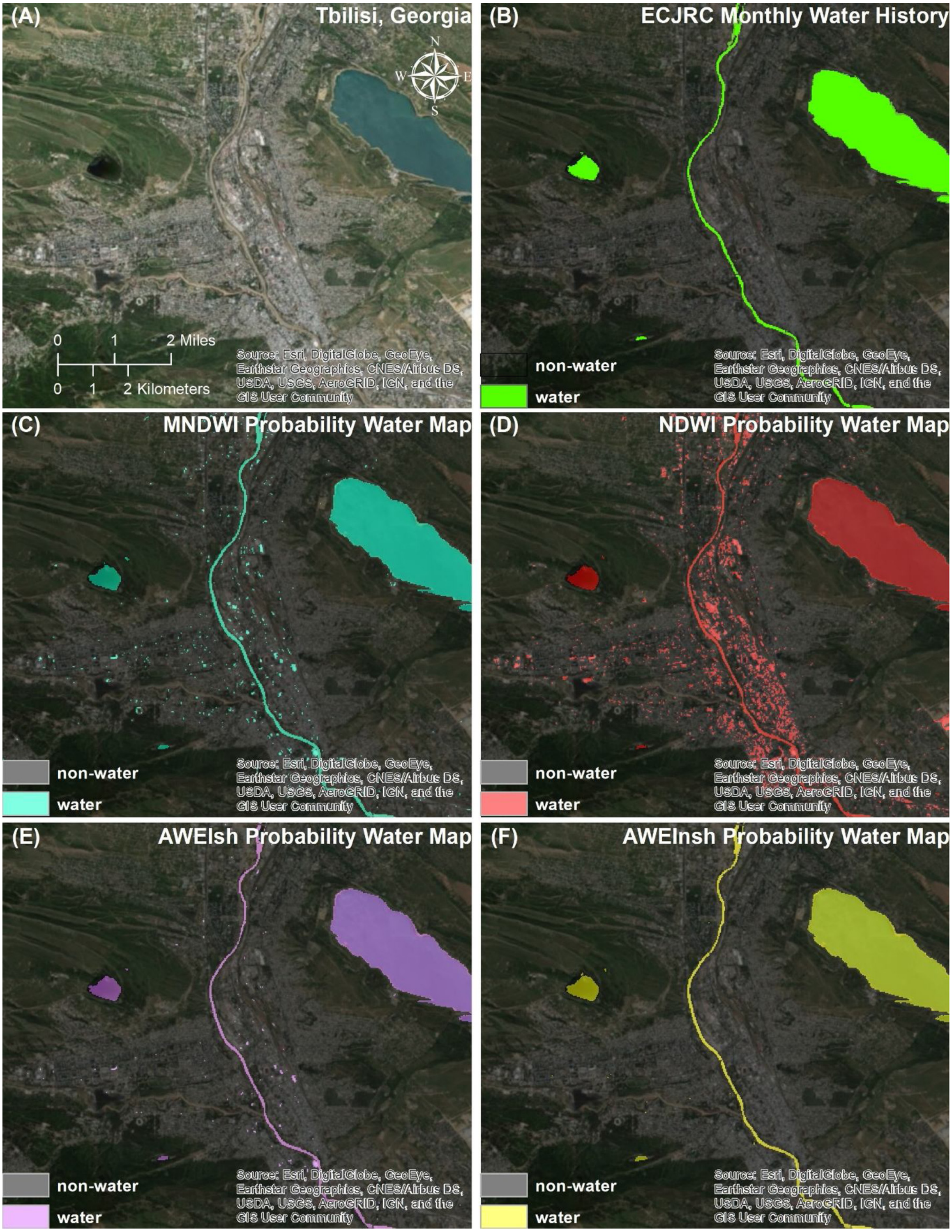


Fig. 4. This figure shows the difficulties of urban pixel confusion of the OPC method between the selected water index probability maps and the ECJRC dataset for the month of July. (A) shows a true-color image of Tbilisi, Georgia. (B) the ECJRC Monthly History water map. (C) the MNDWI water probability map. (D) the NDWI water probability map. (E) the AWEIsh water probability map. (F) AWEInsh water probability.

vegetation to be evaluated as dry land (Jones, 2015). The wetland area of Lake Arh Gol, Azerbaijan, is one such source of confusion between vegetation and the surface water area residing underneath. In this wetland area, the amount of floating vegetation is excluded by the logistic regression model, due to the high spectral reflectance of the NIR and red bands on the surface of the water.

Despite the overall accuracy obtained in this water detection method, the limiting factor in all remotely sensed data is the spatial resolution capabilities of the sensing platform. The spectral values obtained in a pixel are not only influenced by the percentage of the dominant land cover class present in the pixel but is also influenced by the proportions of the subordinate landcover classes present (Ji et al., 2009). As technology progresses and the spatial resolution gets finer, the number of mixed pixels present in the scene will decrease, resulting

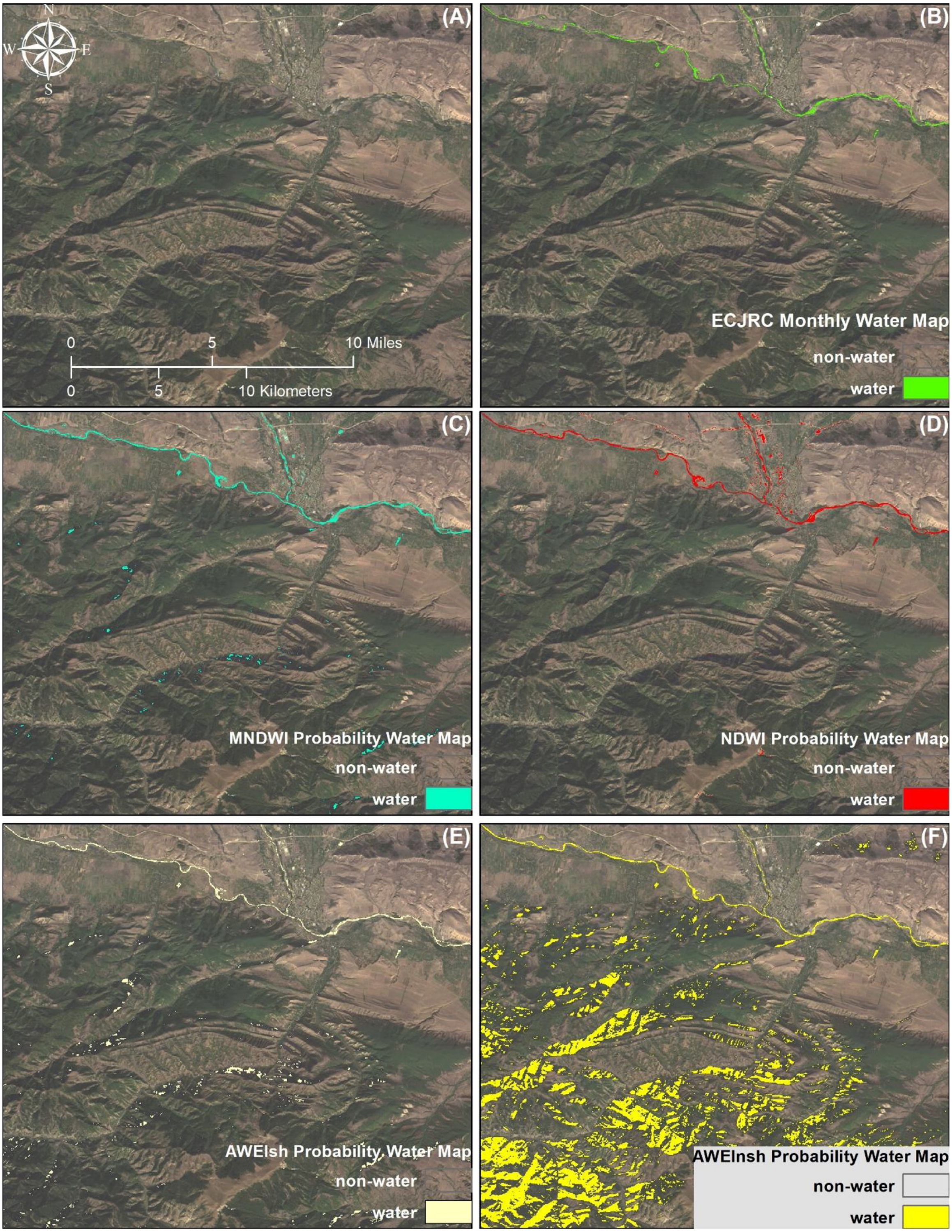


Fig. 5. This figure shows the difficulties of mountain shadow pixel confusion of the OPC value method between the selected water index probability maps. (A) shows a true-color image of mountain region Northwest of Tbilisi, Georgia. (B) Displays water detected by the ECJRC Monthly History water map. (C) Displays water detected by the MNDWI water probability map. (D) Displays water detected by the NDWI water probability map. (E) Displays water detected by the AWEIsh water probability map. (F) Displays water detected by the AWEInsh water probability.

in a more accurate mapping of surface water bodies. Huang et al.'s (2018) review of surface water detection from space using optical sensors, concluded that the future of water monitoring techniques involves the integration of multisource data. Passive remote sensing platforms rely on the magnitude of reflected wavelengths received at sensor from the Earth's surface to create the multispectral datasets that are used in the application of spectral indices to delineate between landcover types. Water tends to absorb almost all incident radiation, resulting in reflectance values of a lesser magnitude when compared to reflectance values of vegetation and urban landcover. Mountain and urban shadows present in multispectral images, tend to reduce the amount of incident radiation available to be reflected and received by the sensor. The wavelength absorbance of water and the reduced magnitude of reflected wavelengths from shadows, result in similar

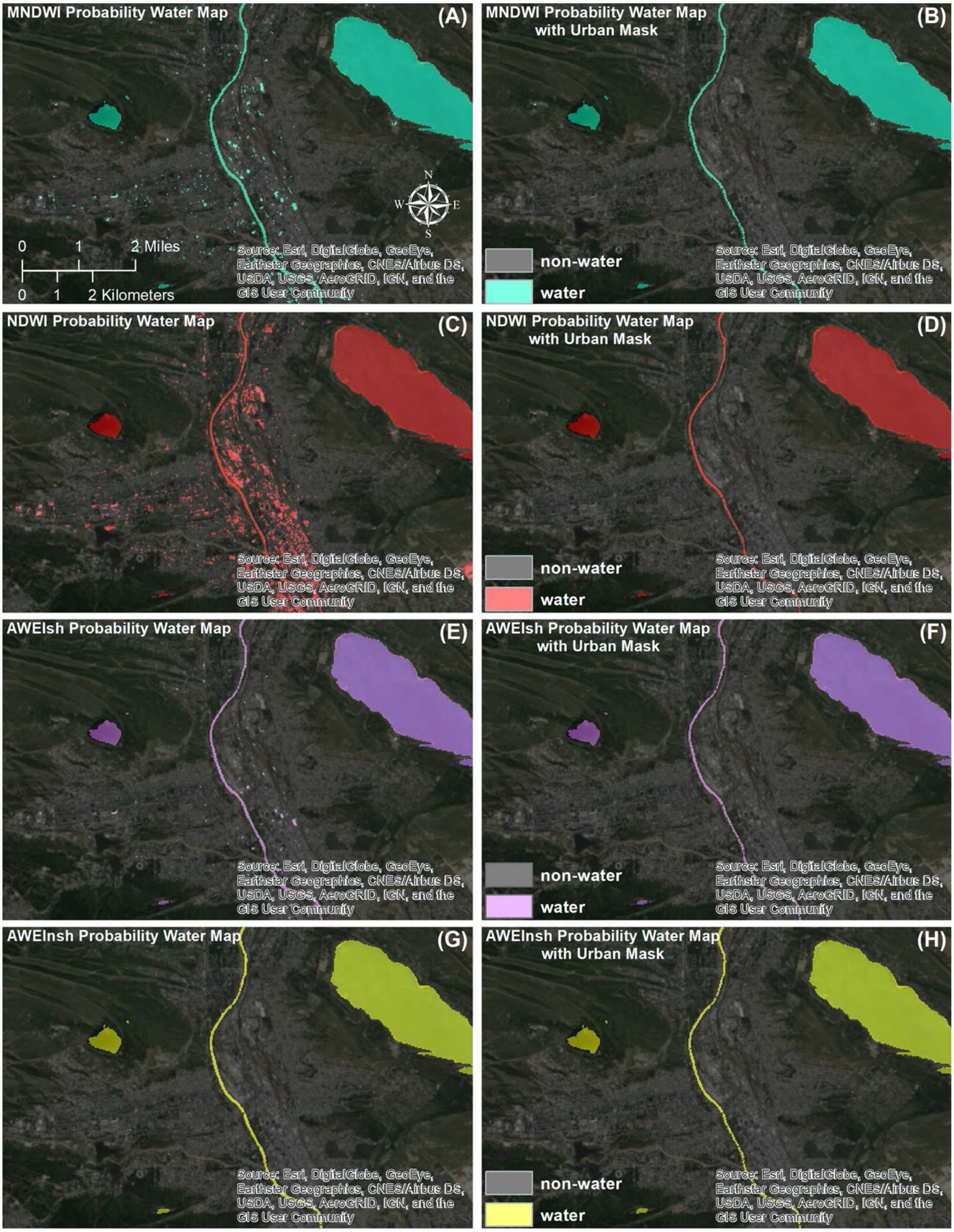


Fig. 6. Displays the misclassification of urban pixels when OPC water map (left) and the application of the urban area mask for the resulting water probability index (right). MNDWI water probability map (A) and the MNDWI water probability map with the urban mask applied (B); the NDWI water probability map (C) and NDWI water probability map with the urban mask applied (D); the AWEIsh water probability map (E) and AWEIsh water probability map with the urban mask applied (F); the AWEInsh water probability map (G) and AWEInsh water probability map with the urban mask applied (H).

water index values derived from the selected bands received by the sensor, leading to misclassification of pixels in water detection indices that rely on the normalized differencing of specific bands. To ameliorate the potential for misclassification of urban and mountain shadow, contemporary water detection methods will require the use of multiple datasets ranging from passive and active sensing platforms to produce water detection products with greater levels of accuracy.

7. Conclusion

The goal of this study was to establish the best performing water detection index in the Caucasus region. Using all validation points from all three study areas and both the wet and dry season, we find that the application of a logistic regression model using an optical surface water index (MNDWI) resulted in the most accurate open surface water maps

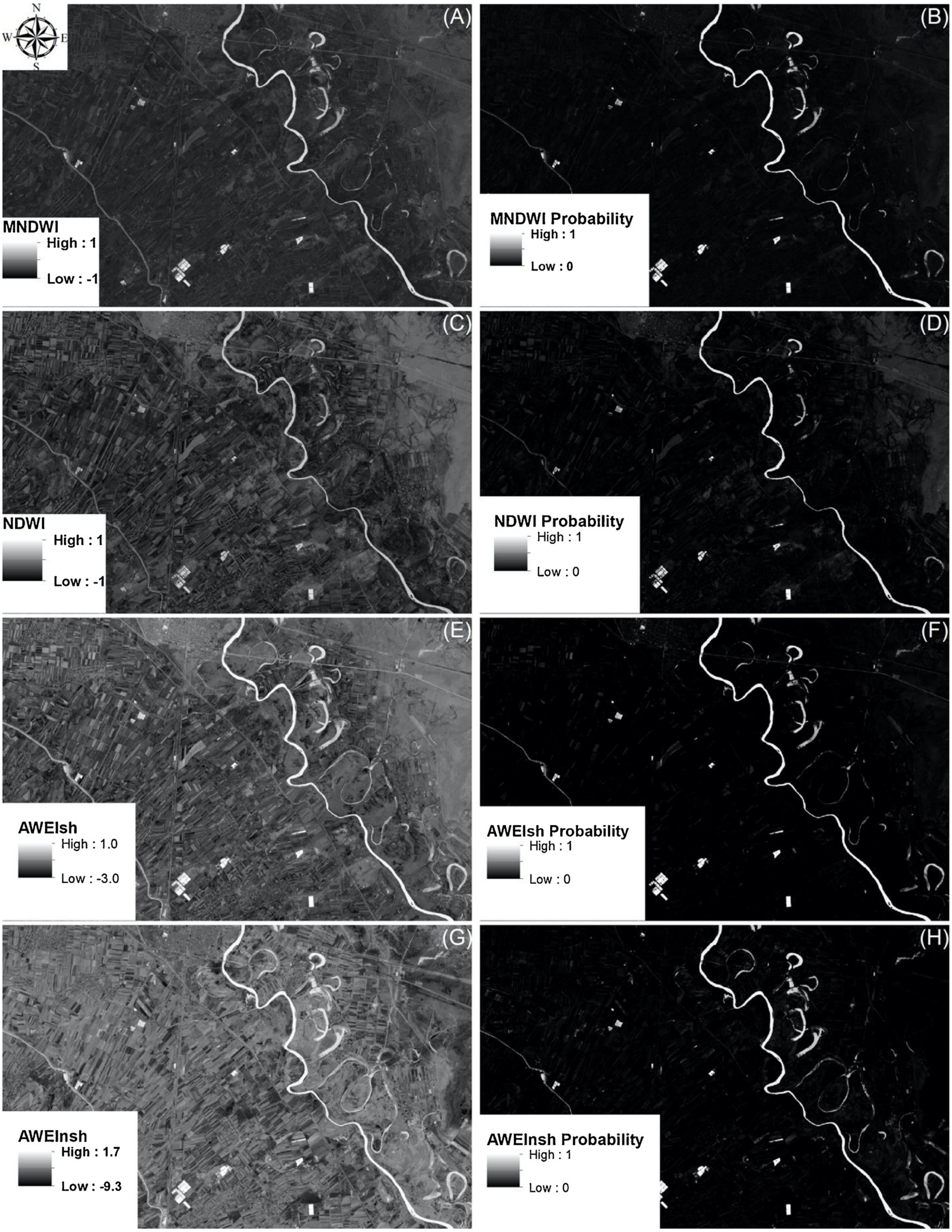


Fig. 7. This figure highlights the increased separation of water and vegetation values between the water index and water index probability maps. The figure includes the MNDWI water map (A) and the MNDWI water probability map (B); the NDWI water map (C) and NDWI water probability map (D); the AWEIsh water map (E) and AWEIsh water probability map (F), and the AWEInsh water map (G) and AWEInsh water probability map (H). Both AWEI water map values were normalized for comparison between the different indices.

if we are using an impervious surface and elevation mask to prevent apparent urban and elevation based confusions. This approach achieved an overall accuracy of 93%, which was similar to what was found for freely available global surface water products. However, we demonstrate that while the global surface water product has a high overall detection accuracy, it was not as reliable in detecting small water features, such as irrigation channels. This result agrees with the findings from Ogilvie et al. (2018). In this study, the authors observed that the ECJRC Monthly History water maps performed well in large bodies of water, but had substantial water omission errors in water bodies of 5 hectares and less; which correlates with the inability of the ECJRC water product to detect surface water in small water canals.

Our results also correspond with a study focused on shoreline detection (Kelly & Gontz, 2018) but disagrees with a study in Australia

Fig. 8. Compares the ability to detect the Upper Garabakh Canal, Azerbaijan, between the selected water indices and the MNDWI probability water map, using the OPC method. A threshold value of 0 was applied to each water index. (A) Shows a true-color image of the Upper Garabakh Canal. (B) Canal water detected by the MNDWI probability water map. (C) Canal water identified by MNDWI water index (D) Canal water identified by the NDWI water index (E) Canal water detected by the AWEIsh water index. (F) Canal water identified by the AWEInsh water index.

that did not find MNDWI to perform the best (Fisher et al., 2016). It is important to note that we did not evaluate the best performing index from that study.

Threshold values established for the extraction of surface water bodies using any type of water index are based on the individual conditions and characteristics of the specific scene being evaluated. The logistic regression, when applied to a water index and classified training points, generates a slope and intercept value that best fits the model of regression. The probability map derived from this logistic regression creates a scale for the selected water index, unique in its relationship to the physical characteristics of the area. We demonstrate that the use of flexible OPC values adapted to individual images allow

Fig. 9. Compares the ability to detect the Upper Garabakh Canal, Azerbaijan, between the selected water indices, the ECJRC Monthly History water map, and the MNDWI probability water map, using the OPC value method. The thresholds chosen for the selected water indices were determined by applying the lowest potential threshold value for each water index, as listed from previous studies. (A) shows a true-color image of the Upper Garabakh Canal. (B) Canal water detected by the MNDWI probability water map. (C) Canal water identified by the MNDWI water index (-0.05 threshold). (D) Canal water identified by the NDWI water index (-0.19 threshold). (E) Canal water detected by the AWEIsh water index (-0.15 threshold). (F) Canal water identified by the AWEInsh water index (-0.15 threshold).

for the accurate detection of water in complex landscapes without the need to establish rigid thresholds which are unlikely transferable to different study regions.

8. Author statement

Conceptualization – James Worden (JW) and Kirsten de Beurs (KdeB). Development of methodology (JW). Validation (JW). Writing and editing (JW and KdeB). Preparation of figures and tables (JW).

Funding acquisition (KdeB).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the NASA: Land-Cover/Land-Use Change project entitled “Land use patterns and political instability as predictors for the re-emergence of malaria in the Caucasus” to KMB. Project number: 16-LCLUC16-2-0017.

References

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

Alsharif, A.A.A., Pradhan, B., 2014. Urban sprawl analysis of Tripoli Metropolitan city (Libya) using remote sensing data and multivariate logistic regression model. Journal of the Indian Society of Remote Sensing 42 (1), 149–163.

Austin, P.C., Steyerberg, E.W., 2012. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. BMC Medical Research Methodology 12 (1), 82.

Bomblies, A., 2012. Modeling the role of rainfall patterns in seasonal malaria transmission. Climatic Change 112 (3–4), 673–685. https://doi.org/10.1007/s10584-011-0230-6.

Brown de Colstoun, E.C., Huang, C., Wang, P., Tilton, J.C., Tan, B., Phillips, J., Niemczura, S., Ling, P.Y., Wolfe, R.E., 2017. Global Man-Made Impervious Surface (GMIS) Dataset from Landsat. NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, USA.

Bruce-Chwatt, L.J., 1959. Malaria research and eradication in the USSR. A review of Soviet achievements in the field of malariology. Bulletin of the World Health Organization 21 (6), 737–772.

Cheng, Q., Varshney, P.K., Arora, M.K., 2006. Logistic regression for feature selection and soft classification of remote sensing data. IEEE Geoscience and Remote Sensing Letters 3 (4), 491–494. https://doi.org/10.1109/LGRS.2006.877949.

Di Rocco, M., Picco, P., Arslanian, A., Restagno, G., Perfumo, F., Buoncompagni, A., Gattorno, M., Borrone, C., 1997. Retinitis pigmentosa, hypopituitarism, nephronophthisis, and mild skeletal dysplasia (RHYNS): A new syndrome? American Journal of Medical Genetics 73 (1), 1–4. https://doi.org/10.1002/(SICI)1096-8628(19971128)73:1<1::AID-AJMG1>3.0.CO;2-Y.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., 2007. The shuttle radar topography mission. Reviews of Geophysics 45 (2).

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27 (8), 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 Sensing of Environment 140, 23–35. https://doi.org/10.1016/j.rse.2013.08.029.

Fisher, A., Flood, N., Danaher, T., 2016. Comparing Landsat water index methods for automated water classification in eastern Australia. Remote Sensing of Environment 175, 167–182.

Global Health Group, 2015. Eliminating Malaria in Azerbaijan. https://globalhealthsciences.ucsf.edu/pub/2015-eliminating-malaria-azerbaijan.

Harrell Jr, F.E., 2015. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering & Remote Sensing 70 (7), 829–840.

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. https://doi.org/10.1029/2018RG000598.

Ji, L., Zhang, L., Wylie, B., 2009. Analysis of dynamic thresholds for the normalized difference water index. Photogrammetric Engineering and Remote Sensing 75 (11), 1307–1317. https://doi.org/10.14358/PERS.75.11.1307.

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 Sensing 6 (6), 5067–5089. https://doi.org/10.3390/rs6065067.

Jones, J., 2015. Efficient wetland surface water detection and monitoring via landsat: Comparison with in situ data from the everglades depth estimation network. Remote Sensing 7 (9), 12503–12538.

Kelly, J.T., Gontz, A.M., 2018. Using GPS-surveyed intertidal zones to determine the validity of shorelines automatically mapped by Landsat water indices. International Journal of Applied Earth Observation and Geoinformation 65, 92–104.

Lacaux, J.P., Tourre, Y.M., Vignolles, C., Ndione, J.A., Lafaye, M., 2007. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. Remote Sensing of Environment 106 (1), 66–74. https://doi.org/10.1016/j.rse.2006.07.012.

Lee, S., 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. International Journal of Remote Sensing 26 (7), 1477–1491.

López-Ratón, M., Rodríguez-Álvarez, M.X., Cadarso-Suárez, C., Gude-Sampedro, F., 2014. Optimalcutpoints: An R package for selecting optimal cutpoints in diagnostic tests. Journal of Statistical Software 61 (8), 1–36. https://doi.org/10.18637/jss.v061.i08.

Mammadov, S., Gasimov, E., Kurdova-Mintcheva, R., Wongsrichanalai, C., 2016. Elimination of Plasmodium vivax malaria in Azerbaijan. American Journal of Tropical Medicine and Hygiene. https://doi.org/10.4269/ajtmh.16-0173.

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.

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, L., McIntyre, A., Tan, P., Curnow, S., Ip, A., 2016. Water observations from space: Mapping surface water from 25years of Landsat imagery across Australia. Remote Sensing of Environment 174, 341–352. https://doi.org/10.1016/j.rse.2015.11.003.

Ogilvie, A., Belaud, G., Massuel, S., Mulligan, M., Le Goulven, P., Calvez, R., 2018. Surface water monitoring in small water bodies: potential and limits of multi-sensor Landsat time series. Hydrology and Earth System Sciences 22 (8), 43.

Oklahoma Water Resources Board, 2018. Water Facts. https://www.owrb.ok.gov/util/waterfact.php.

Okorie, P.N., Popoola, K.O., Awobifa, O.M., Ibrahim, K.T., Ademowo, G.O., 2014. Species composition and temporal distribution of mosquito populations in Ibadan, Southwestern Nigeria. Journal of Entomology and Zoology Studies 2 (4), 164–169. https://doi.org/10.1016/j.micinf.2011.07.011.Innate.

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. https://doi.org/10.1038/nature20584.

Peng, C.-Y.J., Lee, K.L., Ingersoll, G.M., 2002. An introduction to logistic regression analysis and reporting. The Journal of Educational Research 96 (1), 3–14.

Sivanpillai, R., Miller, S.N., 2010. Improvements in mapping water bodies using ASTER data. Ecological Informatics 5 (1), 73–78.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240 (4857), 1285–1293.

Tulbure, M.G., Broich, M., Stehman, S.V., Kommareddy, A., 2016. Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-arid region. Remote Sensing of Environment 178, 142–157. https://doi.org/10.1016/j.rse.2016.02.034.

United States Agency International Development, 2017a. Climate Change Risk in Armenia: Country Risk Profile.

Work, E.A., Gilmer, D.S., 1976. Utilization of satellite data for inventorying prairie ponds and lakes. Photogrammetric Engineering and Remote Sensing 42 (5), 685–694.

World Bank Group, 2012a. Armenia: Climate Change and Agriculture Country Note.

World Bank Group, 2012c. Republic of Azerbaijan: Climate Change and Agriculture Country Note.

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. https://doi.org/10.1080/01431160600589179.

Zhang, Y., Feng, L., Li, J., Luo, L., Yin, Y., Liu, M., Li, Y., 2010. Seasonal–spatial variation and remote sensing of phytoplankton absorption in Lake Taihu, a large eutrophic and shallow lake in China. Journal of Plankton Research 32 (7), 1023–1037.

Zou, Z., Dong, J., Menarguez, M.A., Xiao, X., Qin, Y., Doughty, R.B., Hooker, K.V., David Hambright, K., 2017. Continued decrease of open surface water body area in Oklahoma during 1984–2015. Science of the Total Environment 595, 451–460. https://doi.org/10.1016/j.scitotenv.2017.03.259.

Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39 (4), 561–577.