UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

Application of Spectral Index-Based Logistic Regression to Detect Inland Water in the Caucasus

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN GEOGRAPHY

By

James Worden

Norman, Oklahoma

2021

Application of Spectral Index-Based
Logistic Regression to Detect Inland Water in the Caucasus

A THESIS APPROVED FOR THE
Department of Geography and Environmental Sustainability

BY THE COMMITTEE CONSISTING OF

Dr. Kirsten de Beurs
Dr. Jennifer Koch
Dr. Mike Wimberly

Table of Contents and Front Matter

Contents and Front Matter iv

Abstract v

Chapter 1.
Introduction 1

Chapter 2.
Surface Water Detection in the Caucasus 7

Chapter 3.
Application of Spectral Index-Based
Logistic Regression to Detect Inland Water in the Caucasus 47

Chapter 4.
Conclusions 74

Acknowledgements 76

References 77

Abstract.

The Caucasus is a diverse region with many climate zones that range from subtropical lowlands to mountainous alpine areas. The Caucasus region also contains irrigated agricultural lands, heavily vegetated wetlands, reservoirs, and canals in various states of repair. This study evaluates logistic regression models based on multiple optical water indices to establish the best-performing water detection method (MNDWI) for three Landsat path/row tiles and introduces the optimal probability cut-off method (OPC). The OPC logistic regression model is applied to the Caucasus region covering 19 Landsat tiles from May to October 2019. The global water product from the European Commission Joint Research Center (ECJRC) (1) was used to generate training and validation data points by stratified random sampling. We used 6745 manually classified points (3261 non-water, 3484 water) to validate the OPC and ECJRC max extent water products' performance using an estimated proportion of area error matrix to evaluate accuracy. This approach produced Max extent water maps with higher accuracy (89.2%) and detected 392 km^2 more water than the ECJRC Max extent product (86.7%) for the entire year. Despite high levels of overall accuracy, the error matrix is evaluated using the proportion of area to create user's and producer's accuracy by class (2); we witness a significant loss in performance in detecting accuracy for the water class (15.6% OPC, 11.5% ECJRC). Evaluation of the accuracy of land cover products solely by overall accuracy can be misleading to the user, emphasizing the need for a more rigorous assessment of water detection performance metrics beyond the traditional methods of evaluating accuracy.

Key Words:

Water Index, Logistic Regression, Landsat, Caucasus, Global Land Cover Datasets

Chapter 1

1. Introduction

Water is arguably the most essential compound related to carbon-based life (3). However, our relationship with water can change with the quantity present in a system. Too much water Flooding can cause loss of life and disease prevalence, while drought can cause famine and dehydration (4). With satellite imagery availability starting in the 1970s, applying satellite data to detect water bodies has been an effective means to inventory and quantify the amount of surface water present at a location (5).

The inception of the EROS (Earth Resources Observation Satellites) project was announced in September of 1966 by Department of Interior (DOI) secretary Stewart L. Udall (6). This program's design was, and still is, to observe and collect data about the Earth's natural resources using orbiting remote sensing satellite platforms. Udall believed that this type of observational data would influence future resource management policies, stating that,

"Facts on the distribution of needed minerals, our water supplies and the extent of water pollution, agricultural crops, and forests, and human habitations, can be obtained on a global basis and used for regional and continental long-range planning" (6).

The DOI, in collaboration with the National Aerospace and Space Administration (NASA), the US Geological Survey (USGS), and other agencies launched the Earth-observing satellites in July of 1972, referred to as the Earth Resources Technology Satellite (ERTS-1) or, starting in 1975, as Landsat 1 (7). Landsat 1 consisted of two systems, the Return Beam Vidicon (RBV), the primary, and the Multispectral Scanner (MSS), the secondary (8). Since then, there have been seven more Landsat satellites, with Landsat 8 (OLI) being the most recent and Landsat 9 scheduled for launch around September 2021(9). The Landsat satellite missions have provided

the world's most continuous Earth observation data for almost fifty years, generating its nine millionth image as of September 19, 2020 (9,10).

The Landsat 8 satellite launched into orbit on February 11, 2013, from the Vandenberg Air force Base in California (11). This satellite maintains a near-polar, sun-synchronous orbit at an inclination of 98.2 degrees (12). At an altitude of 705 km, Landsat 8 orbits the Earth every 99 minutes, with a 16 day revisit time(12). Landsat 8 carries two sensors, the Operational Land Imager and the Thermal Infrared Sensor (TIRS). The OLI has a nine-band capacity with a bandwidth range of 0.43µm to 2.29 µm at 30 m resolution. The panchromatic band (Band 8) has a resolution of 15 meters. Two thermal bands make up the TIRS, ranging from 10.60 µm to 12.51 µm at a resolution of 100 meters(11).

In 2008, the USGS announced a change to its data distribution policy. This new policy's goals were to provide up-to-date data products to all users and retain ownership of the data created by the Landsat Satellite series, including ensuring that all products are freely available or shall not exceed the cost of fulfilling user requests (13). By 2016, after the launch of the Sentinel 1&2 satellites, the European Space Agency also made its Earth observation data freely available.

Following the release of such large quantities of data distributed freely to the public, data acquisition's monetary barrier was effectively removed. Unfortunately, the amount of data available was disproportional to the general user’s ability to store and process data. Using all these resources' full potential requires a high level of storage and processing capability which can be a significant barrier to many regular users (14). The Google Earth Engine platform was developed to address this type of technological barrier to access. Available to users at no monetary cost, the Google Earth Engine is a cloud-based platform designed to grant access to computing and storage resources to the general public, while paid commercial licenses are

available for commercial endeavors. The Earth Engine data catalog contains multi-petabyte analysis-ready geospatial datasets and eliminating the user's need to gain additional storage capacity. This data can be obtained and manipulated by an application programming interface (API) available through the internet, including an associated web-based interactive development environment (IDE) that provides visualizations of the data (14). Since 2017, several studies have been conducted to delineate surface water using Google Earth Engine (15–18).

Numerous remotely sensed water detection indices had been developed by applying algebraic operations to specific wavelength combinations to take advantage of multispectral imagery and the unique spectral reflective qualities of water (16,19–27). McFeeters (19) used a band ratio method derived from the normalized difference vegetation index (NDVI). NDVI calculated by dividing the red (0.6 -0.7µm) and near-infrared (NIR) (0.7- 1.1µm) band's difference by its sum takes advantage of the contrast between the reflectivity of various land cover types to highlight vegetation and suppress all other land cover types present in multispectral imagery (28). McFeeters (19) found that this method could detect water by reversing the band order in the numerator and replacing the red band with the green. This new band combination suppresses vegetation's spectral reflectance patterns while highlighting those of water, referred to as the normalized difference water index (NDWI) (19).

The NDWI method for water detection, however, is not without its shortcomings. The NDWI method tends to perform poorly in detecting water around areas that contain built-up or urban areas (20) and needs to be more resilient to misclassifying built-up areas as water; thus, a different band was selected. Instead of the NIR band, Xu (2006) replaced this band with a larger bandwidth, the mid or shortwave infrared (SWIR) (1.55- 1.75µm) band creating the modified normalized difference water index (MNDWI). Changing the NIR band with the SWIR band

results in lower index values for water than the NDWI method, suggesting a default threshold value of zero to delineate water features. The generic threshold value of zero to detect water features is more of a starting point. An adjustment to the threshold value is needed to accommodate the variations in chemical and biological components and atmospheric conditions present during image acquisition (29,30). Manual adjustment to the threshold value is a time-consuming and subjective process that is not interchangeable between images (30,31). Our method removes the need for the researcher to determine a threshold by allowing the individual logistic regression model for the scene to determine the optimal threshold.

The South Caucasus region is made up of three countries: Armenia, Azerbaijan, and Georgia. This region hosts many different climate zones ranging from mountainous alpine climates to dry subtropical lowland plains (32–34). From the 1800s, malaria has been prevalent in the Caucasus Region, with a recorded 600,000 cases in Azerbaijan in 1934 (35). By the 1950s, the number of documented malaria cases totaled 781,239 (36). After a comprehensive effort from the Global Malaria Eradication Campaign in the 1960s, the number of malaria incidents declined, and two malaria species were eradicated. A third malaria species (P. vivax) escaped elimination and precipitated a surge of malaria cases in the mid-1990s (37). Each of the three countries experienced land reform and privatization of the agricultural sector in the 1990s, which led to the segmentation of large agricultural plots into smaller private and commercially owned farms, leaving irrigation systems in disrepair or inadequate in some agricultural areas (32–34).

The P.vivax species of malaria is vector-borne through the anopheles mosquito. The anopheles mosquito requires persistent breeding pools caused by intraseasonal rainfall to proliferate. A relationship can be observed between the increased surface area of persistent pools of water and the abundance of mosquitoes in mechanistic models (38). It is essential to be able to inventory

the spatial distribution and area of standing surface water to understand the magnitude of potential breeding environments for mosquitoes. Due to limitations of flight range and survival rate, mosquitoes are restricted to areas that contain persistent pools of water (39,40). Human interactions with mosquitoes are most likely to occur in close proximity to mosquito habitats. An increase of mosquito breeding areas may contribute to an outbreak of malaria, making the detection and quantification of surface water in malaria-prone regions of great importance in predicting malaria-based endemics.

This study aims to improve the band ratio water index's detective capability by establishing a relationship between water's spectral characteristics, the index value, and threshold selection using a logistic regression model to create probability maps and the optimal probability cut-off (OPC) method to generate water maps. I will further this technique by applying the OPC method to the entire Caucasus region for the period May to October 2019. Manual evaluation of training points can be a time-consuming, challenging process producing temporally static data that is less representative of the conditions as time increases from the acquisition date. This study will be using open-source global water products through progressive time steps to train the logistic regression water model and OPC value. I will show that it is possible to improve water mapping capabilities by taking advantage of existing land cover detection datasets to train and classify more accurate surface water maps.

Chapter 2 begins with evaluating the performance of logistic water models and probability threshold using different water indexes to establish the index with the best performance using three Landsat tiles over two dates in Georgia, Armenia, and Azerbaijan. I compare the accuracy of water detection of the best performing logistical model to a global water dataset from the European Commission Joint Research Center (ECJRC).

Chapter 3 expands this technique across 19 Landsat tiles for May-October 2019 in Georgia, Armenia, and Azerbaijan, using stratified random sampling of a global water dataset to generate training points for the logistic water model.

Chapter 4 will consist of synthesis and general conclusions from the previous two chapters.

Chapter 2. Surface Water Detection in the Caucasus

Worden, James, and Kirsten M. de Beurs. "Surface water detection in the Caucasus." *International Journal of Applied Earth Observation and Geoinformation* 91 (2020): 102159

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(41). 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 (36). In 1934, there were almost 600,000 reported incidences of malaria in Azerbaijan (35). After the completion of the Global Malaria Eradication Campaign, there were no reported cases of malaria by 1960 (37). 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 (41). After the outbreak of malaria was observed, control and prevention measures were subsequently implemented on a large scale between 1997 and 2011 (41).

Malaria is a vector-borne disease that is caused by Plasmodium and distributed by Anopheles mosquito (42). 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 (43).

Remotely sensed data are often used to identify surface water in a landscape (5,26,44–46). 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 (19–21,45). 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 (30,47,48). 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 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 (1).

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 (Figure 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 (49). In Azerbaijan, 80% of agricultural production comes from irrigated lands, and agriculture employed 39.7% of the country's labor force in 2010 (50). 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, the croplands were re-distributed to rural families after 1992,

and about 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 (33). 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 (33). Poor management in an aging infrastructure led to the majority of the irrigation systems being in poor condition in Azerbaijan (50). 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 (50).

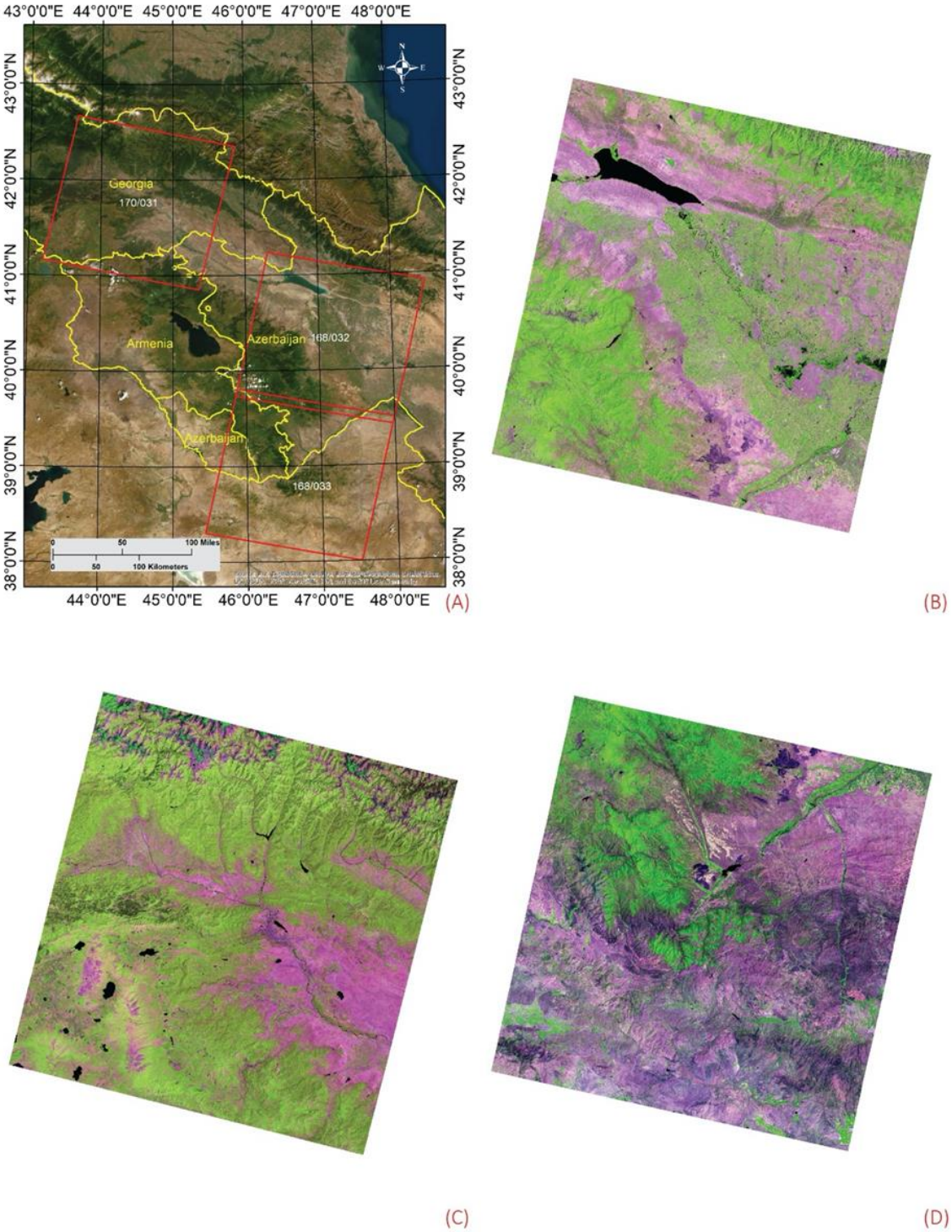

Figure 1: (A) Overview image of the Caucasus study region, (B) the Azerbaijan Landsat 8 tile 168/032 (6,5,7 RGB), (C) the Georgia Landsat 8 tile 170/031 (6,5,7 RGB), (D) The Armenia Landsat 8 tile 168/033 (6,5,7 RGB).

Data

3.1 Google Earth Training and Validation

We manually identified 1000 validation and training (30m by30m) 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 30m. 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.

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

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 (27). The NDWI uses the reflected visible green (0.52-0.60µm) and the near-infrared (NIR, 0.7-1.4µm) 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 (19).

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,1.55-1.75µm) 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 (21). 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). 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.

In some cases, the AWEInsh equation may not effectively eliminate all shadow types and surfaces of low albedo. With these limitations in mind, the second equation, "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 (21). 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 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).

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: (20); **b**: (30); **c**: (47); **d**: (21); **e**: (51); **f**: (52)

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 (1). 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 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 – 2.35µm), 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 (53). The SRTM utilized C band (5.6cm) and X band (3.1cm) 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 (53). 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) (54). 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 (55). To produce training data to train the cubist regression tree algorithm, 1,800 high-resolution scenes were obtained from the NGA WARP 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. Figure 2 provides an overview of the applied methodology.

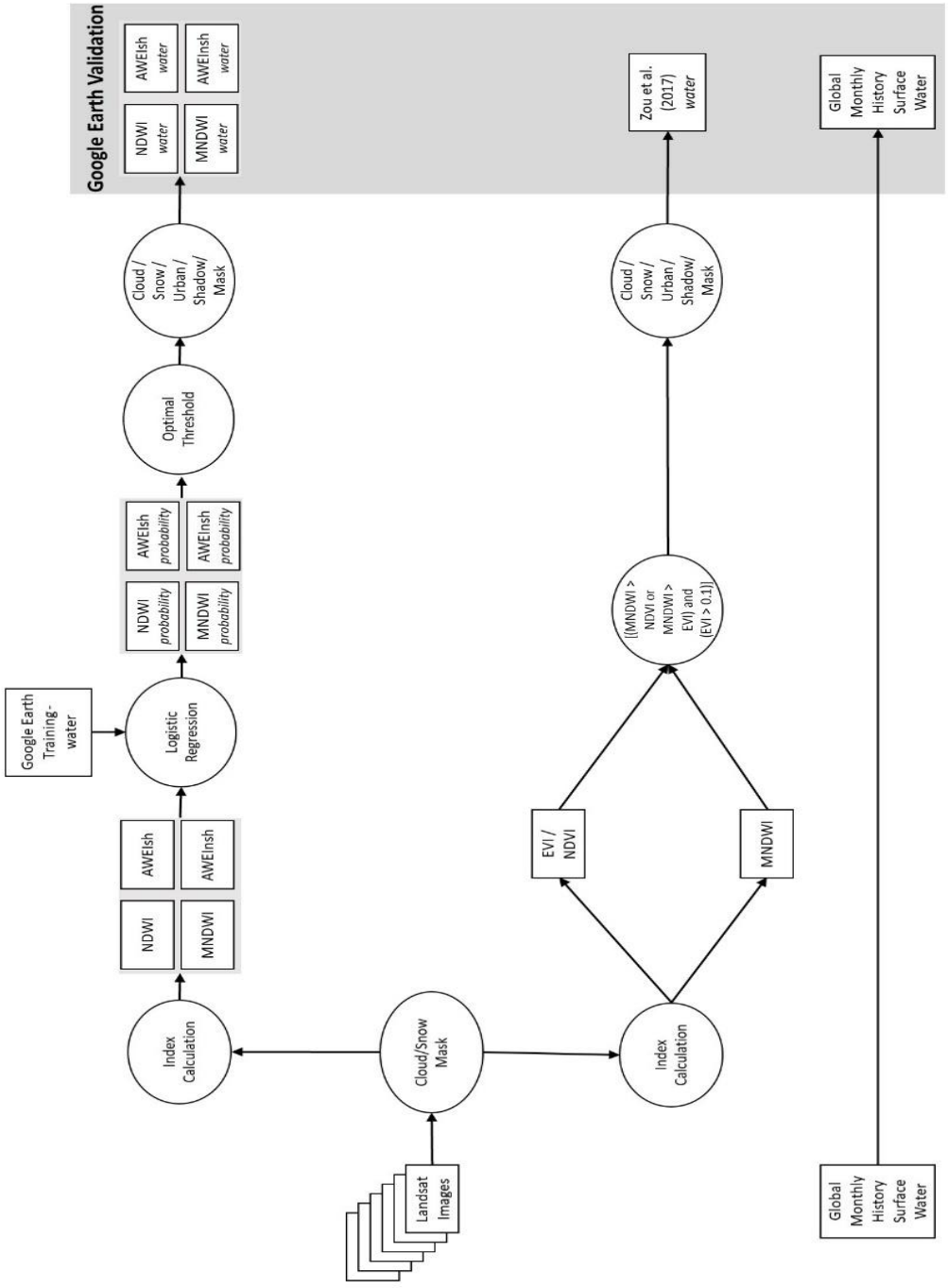

Figure 2. This figure displays the workflow process for the extraction of surface water body area from water index probability maps. For consistency and continuity, the Cloud/Snow/Urban/Shadow mask was applied when comparing the surface water area of Zou et al. 2017 and the probability maps for the wet/dry 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 (44,46,56,57). Others stated that the logistic regression performs well for testing hypotheses and describing the relationship between predictor variables and a categorical outcome variable (58). 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 (57). The probability equation for the logistic regression is described as follows:

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

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 (30). 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™ 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 (20). 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 (51).

The AWEI water extraction index considered multiple thresholds by calculating commission and omission errors and plotting the percentage errors against threshold values (21). 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 (21). 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 (59). In this study, we will be focusing on the former. The 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 (60). Others disclosed the advantage of using a ROC curve to measure the performance (accuracy) of the different water detection methods (61). 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 (62). 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 (63). 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 (64). 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 (64). 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. 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.

Table 3: Results of Logistic Regression and the OPC **values. An optimum probability cut-off 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	Optimum Probability cut-off
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

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 1). 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, 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.

Table 4: Overall accuracy for each water index by seasons 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%*	

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 European 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.

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%

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.

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

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.

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 consistent with the method's tendency to underestimate the percentage of total water pixels in the study area (Table 5).

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

Evaluating the OPC method, we find that there are strengths and weaknesses between the performance of the different water probability maps. Figure 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.

Figure 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.

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. Figures 4 and 5 highlight the misclassification in areas that contain urban and mountain shadow land features. In Figure 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. Figure 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 AWEInsh performed the worst as it relates to mountain shadow misclassification.

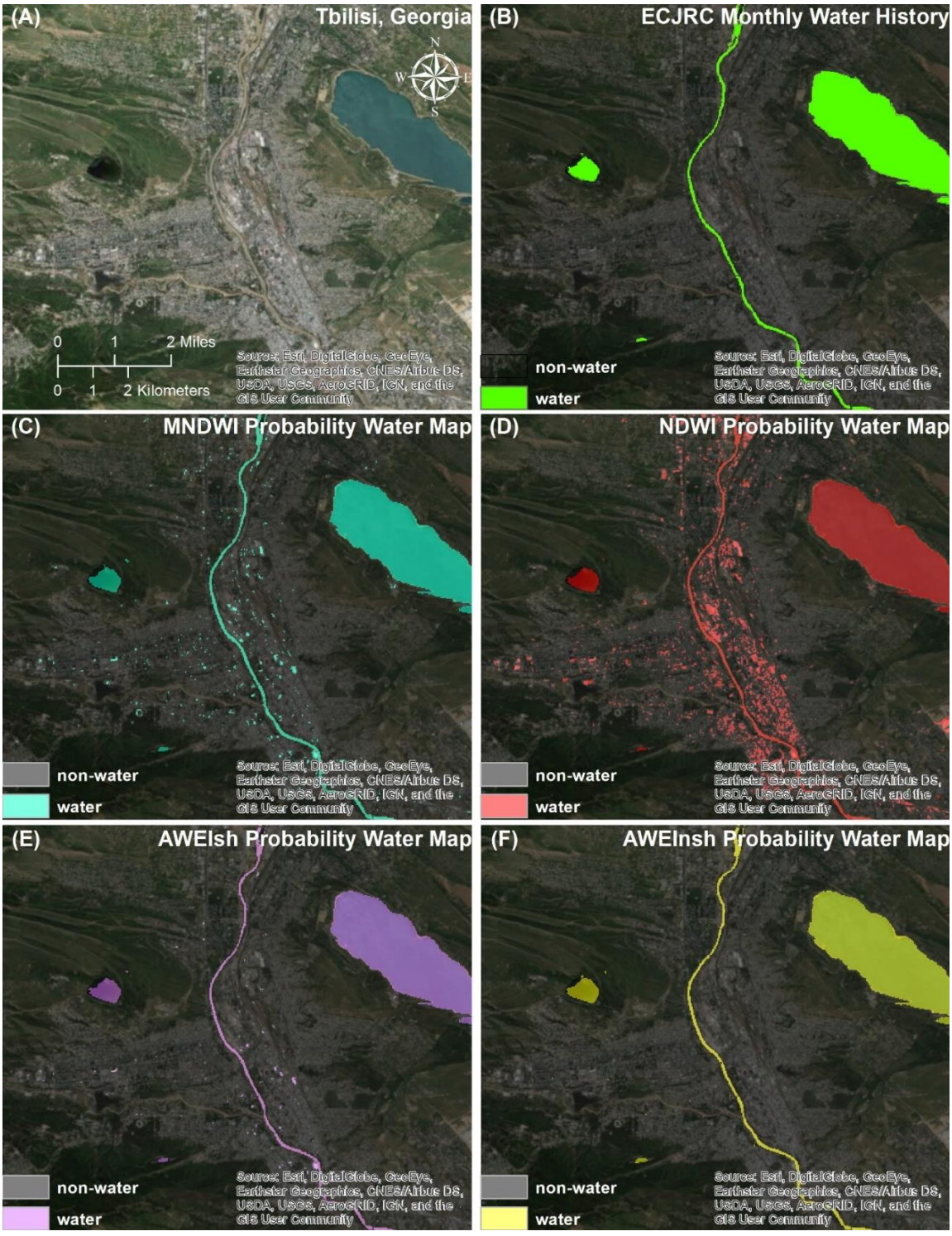

Figure 4: This figure shows the difficulties of urban pixel confusion of the OPC method between the selected water index probability maps and the ECJRC dataset for the month of July. (A) shows a true-color image of Tbilisi, Georgia. (B) the ECJRC Monthly History water map. (C) the MNDWI water probability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

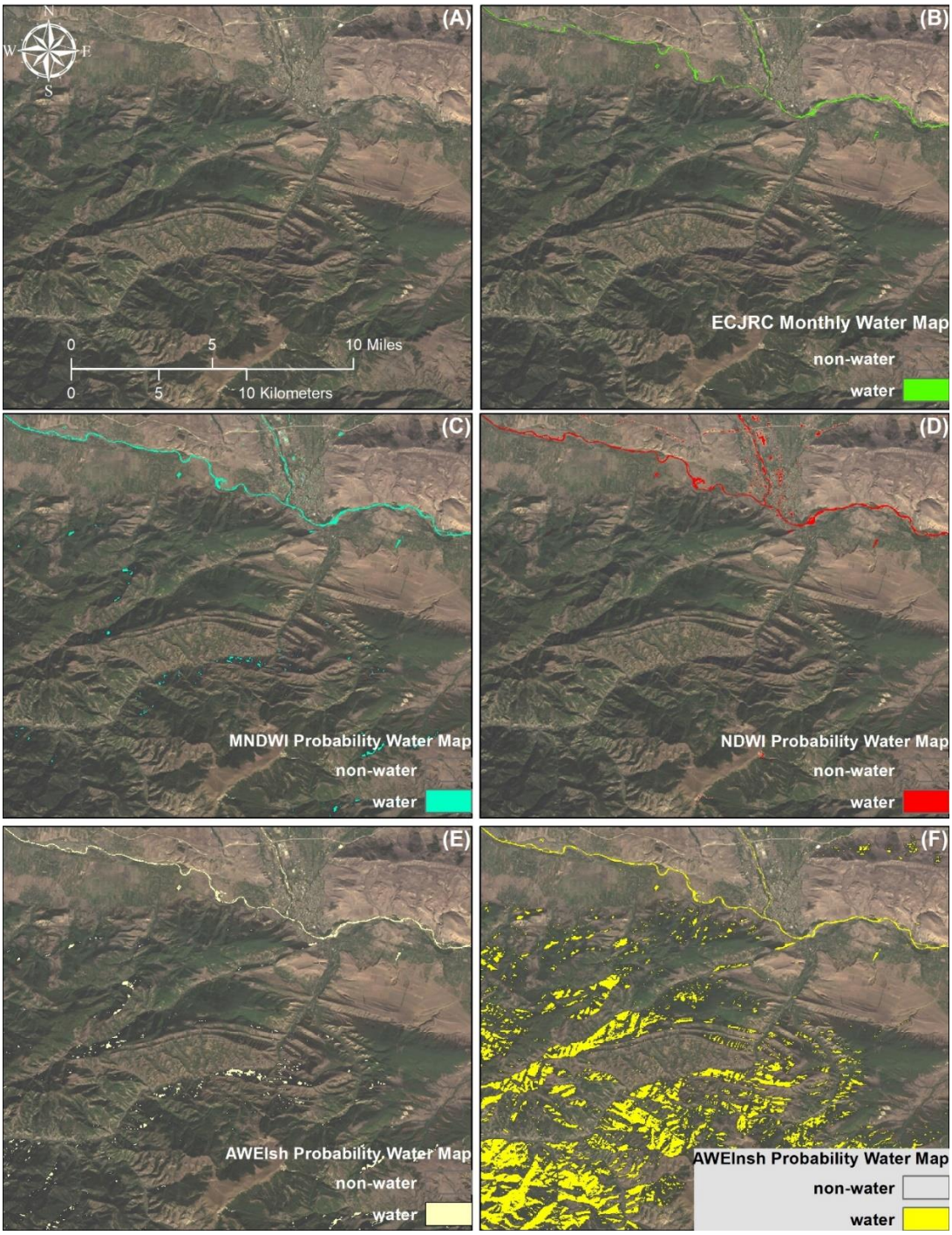

Figure 5. This figure shows the difficulties of mountain shadow pixel confusion of the OPC value method between the selected water index probability maps. (A) shows a true-color image of the mountain region Northwest of Tbilisi, Georgia. (B) Displays water detected by the ECJRC Monthly History water 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

Evaluating Figures 3, 4, and 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. Figure 6 shows the water index probability maps before and after the urban areas mask was applied.

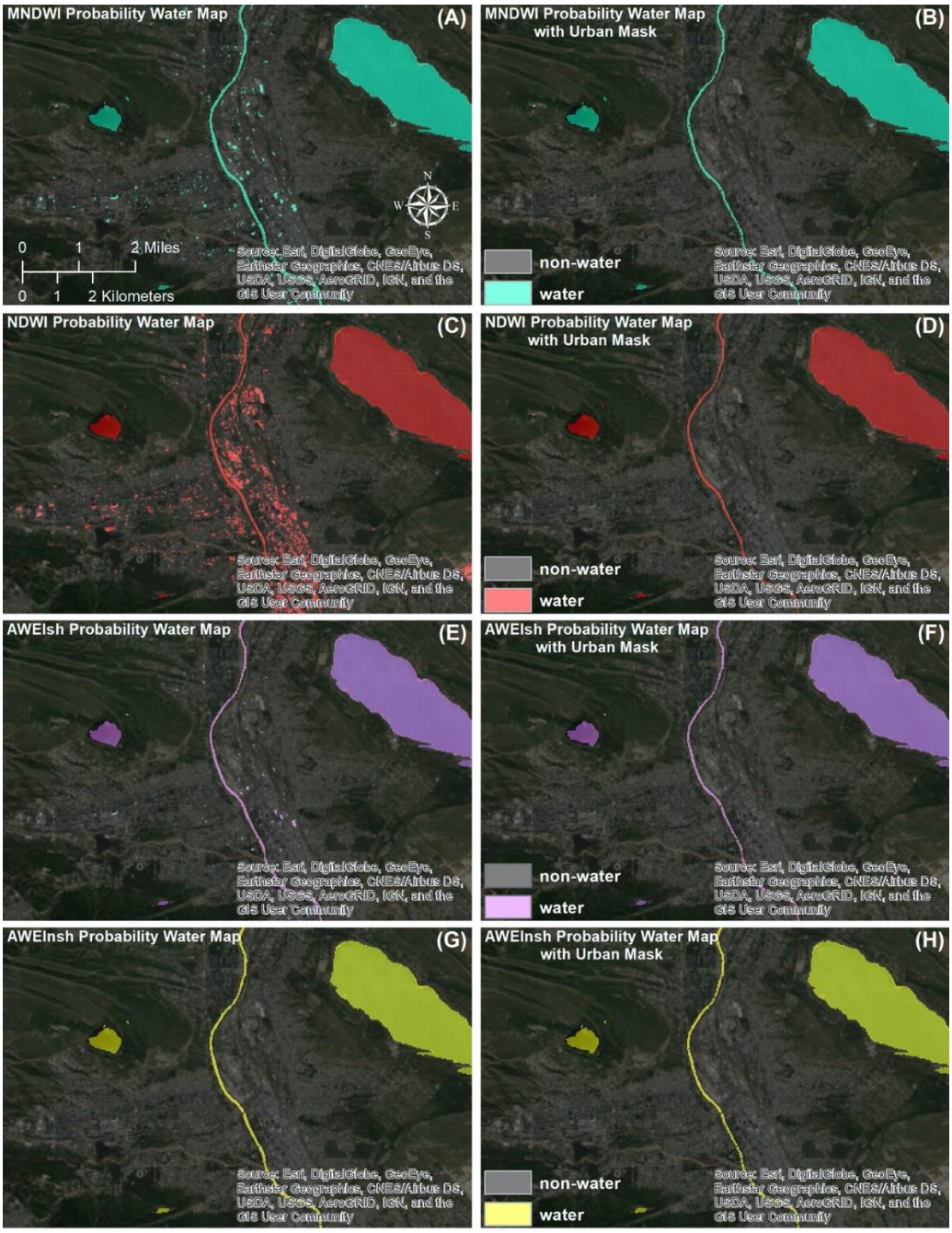

Figure 6: Displays the misclassification of urban pixels when OPC water map (left) and the application of the urban area mask for the resulting water probability index (right). MNDWI water probability map (A) and the MNDWI water probability map with the urban mask applied (B); the NDWI water probability 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).

Many studies have used water indices with other methods to reduce the misclassification of shallow water bodies (26,27,47). 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 (29). 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 (Figure 7).

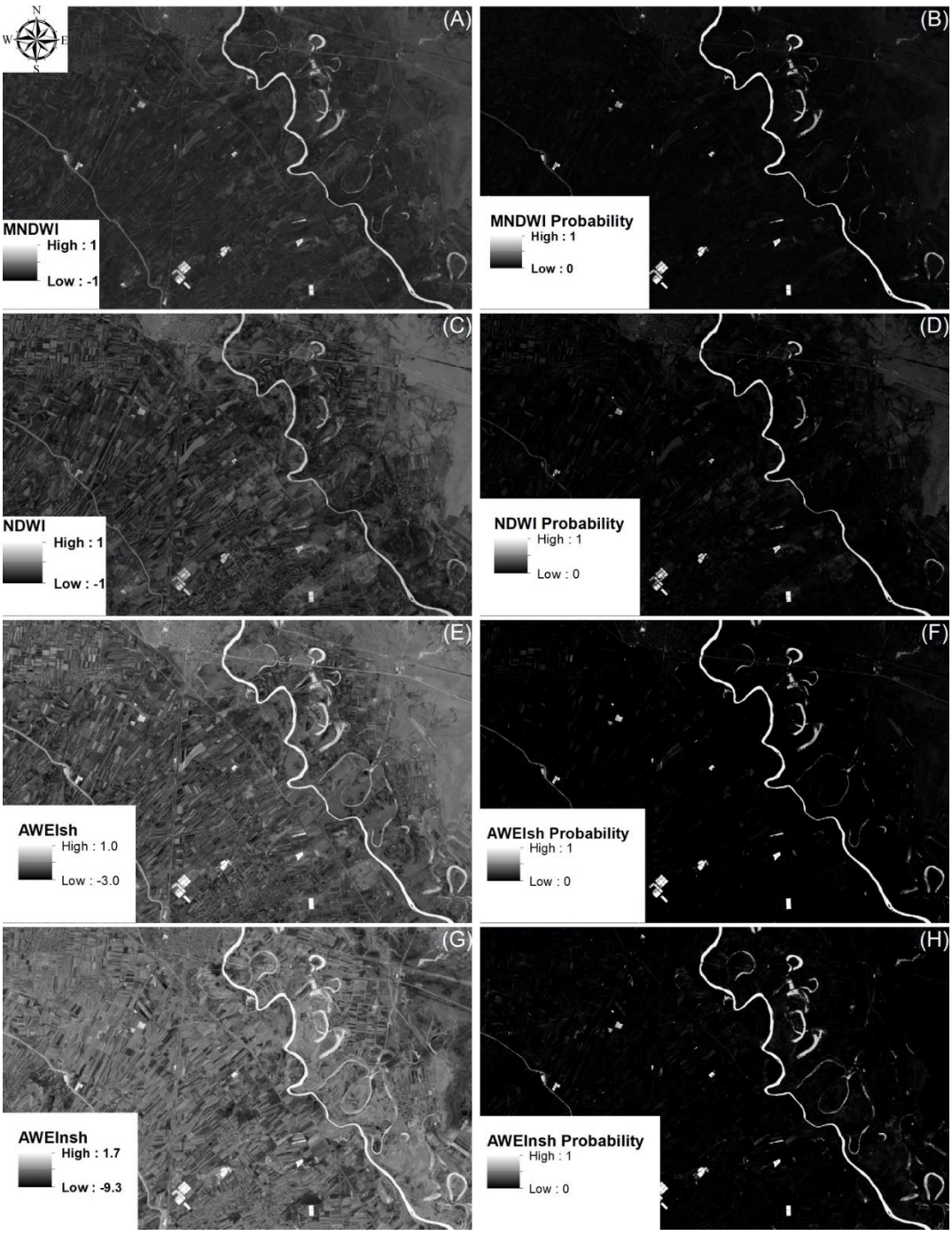

Figure 7: This figure highlights the increased separation of water and vegetation values between the water index and water index probability maps. The figure includes the MNDWI water map (A) and the MNDWI water probability map (B); the NDWI water map (C) and NDWI water probability map (D); the AWEIsh 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.

Looking at the spectral reflectance tendencies of water, MNDWI values of water are generally greater than zero (20). 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 (21). 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 Figure 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 is due to the water in this canal having negative index values.

Figure 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. The thresholds chosen for the selected water indices were determined by applying a threshold value of zero for 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.

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. 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.

Figure 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).

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 (30). 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 (65). 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 (31). 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 (66). 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(26). 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 (1). 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., 2017, Turbule 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 detection in the type of

water body being evaluated. In Figures 4 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 Figures 4 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 vegetation to be evaluated as dry land (67). 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 (30). As technology progresses and the spatial resolution gets finer, the number of mixed pixels present in the scene will decrease, resulting 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 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 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 (48) but disagrees with a study in Australia that did not find MNDWI to perform the best (52). 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 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. 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.

.

Chapter 3. Application of Spectral Index-Based Logistic Regression to Detect Water in the Caucasus

1. Introduction

Since 1972, the Landsat sat program has been instrumental in resource mapping and land cover change and has fostered a new generation of academic research and analytic methods in the remote sensing community (7). Many studies around the world have utilized Landsat imagery in the detection of surface water dynamics (26,51,67–69)

There are many ways to detect water from the images obtained by remote sensing satellites, such as single band, spectral index, machine learning, and spectral mixture analysis-based methods (70). Single-band procedures apply a single bandwidth to describe land cover features. Such single band techniques include density slicing, contour model-based segmentation, and simple linear iterative clustering (SLIC) (71–73). Combining spectral bands using band ratio methods improves the discriminative ability of the single band method. Using the normalized difference of bands or applying coefficients to highlight or suppress land cover features allows for a default threshold value of zero as a starting point for water landcover delineation. Commonly used water indexes are the NDWI, MNDWI, AWEI, and the enhanced water index (EWI) (20,21,74,75). Machine learning (ML) algorithms attempt to train models to learn by association. The machine learning approach falls into two categories of regression and classification algorithms, while classification algorithms can be classified into two sub-categories: supervised and unsupervised (70). Artificial neural network (ANN), maximum likelihood classification, and support vector machine (SVM) are commonly used ML approaches. Spectral unmixing consists of breaking down mixed pixels into smaller spectral components to increase water detection accuracy (70).

The advantage of choosing a logistic regression model is establishing causal relationships and understanding which variables have the most substantial influence on the predicted outcome (76). ML methods that require multiple variables in their decision-making can make identifying these relationships difficult, and the internal logic of the models are not as transparent to the user (76,77).

Training data quality can significantly impact a classification model's accuracy or effectiveness and should be derived from *in situ* data (78). Obtaining quality training data can be a consumptive process as requirements of cost, time, and processing power are a common barrier to access (79). Landsat imagery has been used to create training data used in different land cover classifiers as an alternative to in situ training (80,81). For regions with high landcover class homogeneity, fewer training pixels may suffice. Still, heterogeneous areas require a large number of training points to be considered representative of the scene's conditions (78).

Currently, many global landcover datasets are distributed freely to the public, such as the Global Man-made Impervious Surface (GMIS), the Shuttle Radar Topography Mission (SRTM), and the European Commission's Joint Research Centre (ECJRC) Global Surface Water datasets (1,53,54).

The ECJRC is a freely available global water dataset based on over thirty years of Landsat data from 1984-2015 and is updated frequently with current data for its products. The water products offered by the ECJRC are highly accurate in the detection of surface water but struggles in detecting small or vegetated water bodies as a result of a flooded landscape (82,83)

Past studies have shown that it is possible to improve upon existing global land cover products by using these datasets to generate training samples(84,85). Combining training samples from

existing land cover products with other land cover classification techniques such as random forests (RF) has led to improvements in accuracy for detecting vegetation compared to the original training dataset(86).

For this study, we will train an MNDWI logistical regression water model for the Caucasus region using a global surface water product distributed by the ECJRC.

2 Data

2.1 Landsat

For this study, we used level 2 images from the Landsat 8 (OLI) mission, consisting of 19 tiles covering the Caucasus region from May to October 2019. These level 2 images are atmospherically corrected and have a spatial resolution of 30 m. The Landsat images must meet the criteria of having less than 30% cloud cover present in the scene to be included in the study.

The quality assessment pixel_qa band associated with each Landsat 8 image is used to create a cloud mask. The cloud mask removes clouds, cloud shadows, and snow pixels from the training/validation dataset.

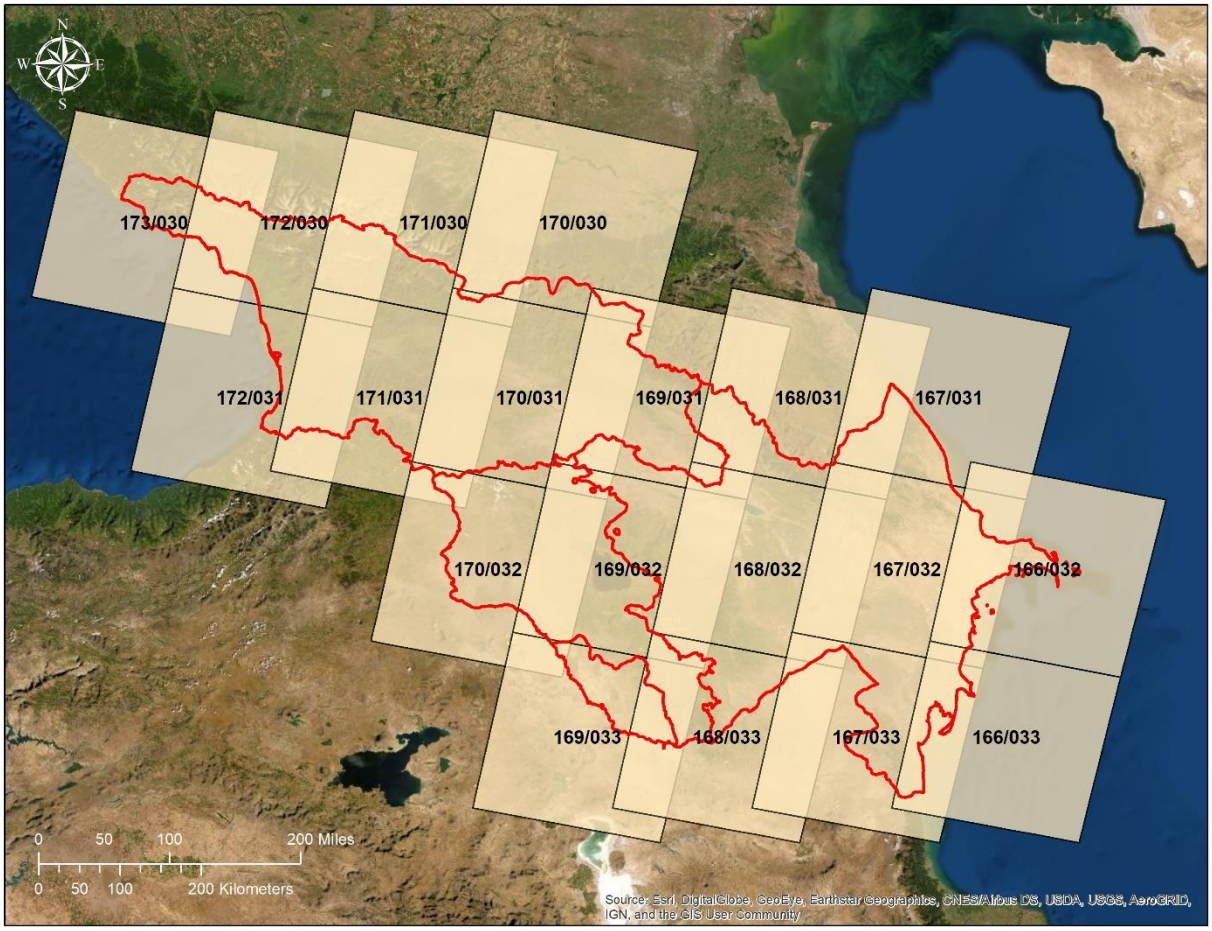

Figure 1. All Landsat 8 Path/Row tiles covering Armenia, Azerbaijan, and Georga in the Caucasus region.

2.2 Training Dataset

We developed the training points as follows. First, the ECJRC Monthly Water History map was selected using the path/row set's corresponding month. We then selected a stratified random sample of 1500 points (750 water, 750 non-water) based on the ECJRC water map using the Landsat tile extent of the path/row set as the region of interest. Using the sampled points, the MNDWI values were extracted from each Landsat path/row set to train the logistic regression

model. The validation dataset to determine the performance metrics and the OPC value was validated using Google Earth imagery.

2.3 Validation Dataset

Within the country boundaries of Georgia, Armenia, and Azerbaijan, we randomly selected and then evaluated and classified 6491 stratified validation points (3238 non-water and 3253 water) using Google Earth Imagery. We used the ECJRC Yearly Max Extent water map for 2019 to generate the validation by stratified random sampling. Each point was manually evaluated and classified using Google Earth Imagery. The validation point locations were limited to the Landsat tile paths' overlapping sections to increase the number of uses in validating the model across the Caucasus region (Fig. 2). With the validation points located in the overlapping areas of the path/row sets, each validation point can be used twice to validate the logistic regression model in adjacent path/row sets.

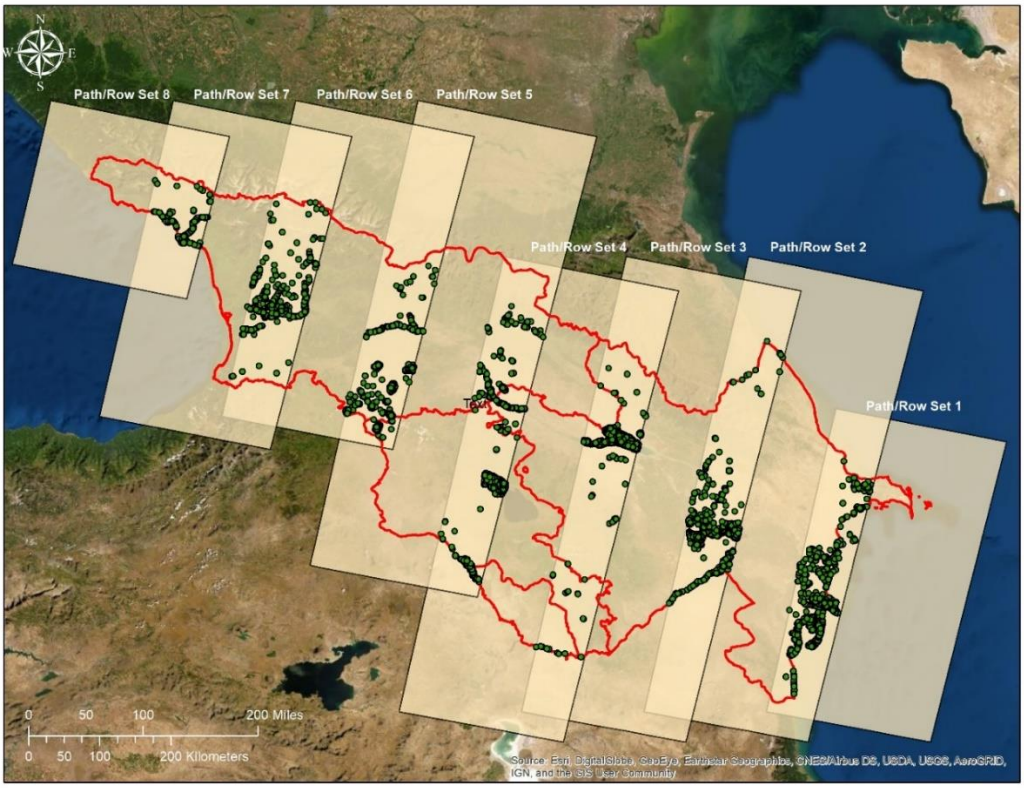

Figure 2. This figure displays the Path/Row sets and validation points located in the overlapping sections of the Path/Rows.

Besides identifying the random points based on the ECJRC water dataset, we also added 500 additional water validation points focusing on points that showed a divergence between the ECJRC data and the data generated in this study. These post-classification validation points were sampled from a difference water map derived from the OPC and ECJRC extent maps, using grid cells where the OPC method detected water and ECJRC did not. The 500 additional points were manually classified using Google Earth Imagery, separating the points into four classes: water, not water, unclear, and seasonal/ water fraction (SWF). Pixels were labeled as SWF if there was subpixel water present (water fraction less than 50%). This class includes seasonal water, which are pixels that contain no water in the image, but observational evidence suggests that water was present at one time. Points labeled as unclear were either obscured by cloud cover

or were otherwise unable to verify the land cover type from the Google Earth images. Pixels classified as unclear or SWF were omitted from the additional water validation points, leaving 254 points added to the validation dataset. The post-classification validation points' inclusion increases the total number of the validation sample size to 6745 points (3261 non-water, 3484 water).

3 Methods

3.1 Generating OPC water maps

Previously (Worden and de Beurs, 2020), we used the MNDWI water index to train and validate a logistic regression model and produce water probability maps. Using the Optimal Probability Cut-off (OPC) method, we created water maps over three Landsat 8 tiles in the Caucasus Region(82). The previously developed method depended heavily hand selected training and validation points. This study aims to scale this methodology up and apply the OPC method to all Landsat images covering the Caucasus region from May to October 2019. Instead of hand training thousands of validation points, we use training data sampled directly from the ECECJRC Monthly Water History. In Fig. 3., we present an overview of the methodology applied to each Path/Row set date in the time series.

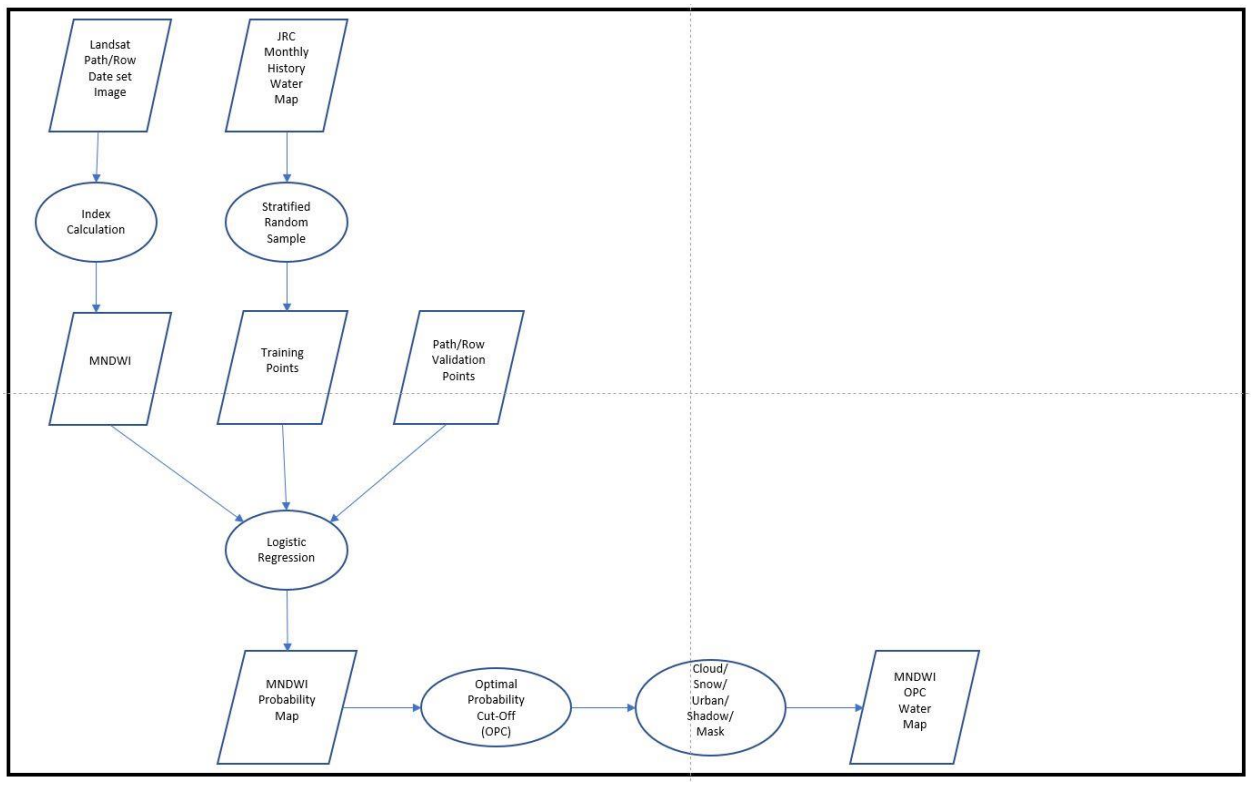

Figure 3. This figure describes the process of generating OPC water maps from multiple path/row sets using the ECJRC Monthly Water History for training and manually classified validation points using Google Earth imagery.

We used the ECJRC Monthly History water product to generate each Path/Row date set's logistic regression training points. After generating the training points, MNDWI values are extracted from the point locations and used to train a logistic regression model, resulting in a slope and intercept value for the water model. The slope and intercept values are then entered into the logistic regression probability equation (1), described below.

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

In this equation, *p* represents the probability value of the presence of water in a pixel, *b* represents the index's slope value, *x* represents the MNDWI value of the pixel, with *a* as the y-intercept. Applying the probability equation from the logistic regression model allows us to create water probability maps, in which the value for each pixel is the probability water is present. The threshold determination between water and non-water landcover is evaluated to

develop the water map using the optimal probability cut-off (OPC). The OPC is an iterative process based on the receiver operator characteristic (ROC), testing all probability cut-off values that produce the most significant degree of accuracy for water classification. A cloud/snow mask derived from the pixel_qa band included in Landsat images was applied to remove this pixel type from the logistic model's training and validation datasets.

To reduce confusion between water pixels and other land-cover classes, we used a cloud/snow, mountain shadow, and urban area mask applied in our previous study to develop the final water map (82).

3.2 Performance Evaluation of Water Maps

The ECJRC and OPC dataset's performance are assessed by comparing the Max Extent water datasets' accuracy for the entire Caucasus Region, including comparing the accuracy between the individual path/row sets of the ECJRC Monthly water history and the OPC water maps. To evaluate the two Max Extent datasets for the Caucasus Region, we have developed an error matrix of sample counts and an error matrix of estimated proportions. Olofsson et al. (2012) warn of calculating overall and producer's accuracy estimations directly from a sample count error matrix, suggesting that such a matrix would not account for the variation in estimation weights based on the mapped classes' proportion. Instead, they argue that an error matrix describing the estimated proportion of the area is more appropriate for evaluating producer's and overall accuracy for land cover classification maps. Since water is a relatively uncommon land surface category, we believe that this is especially relevant in our case. We calculated the overall accuracy from the sample count matrix and compared the result to the overall, user's and producer's accuracy based on the mapped area class proportion (10). We compare the ECJRC

Monthly history and OPC water maps' performance for each path/row set in addition to the Caucasus region.

For each observation, the path/row sets' comparison uses the sample count matrix to calculate overall accuracy. The producers' accuracy for the water class is derived from the estimated proportion of mapped area by class.

3. Results

This study used a random stratified sample of the ECJRC yearly max extent water map to construct a dataset to train a logistic regression water model to classify water across Armenia, Azerbaijan, and Georgia from May to October 2019 using the OPC method.

Below, we compare the overall accuracy between the ECJRC monthly water history and OPC for each observation of the Path/Row date sets. We then evaluate the ECJRC yearly water max extent (12 months) and OPC max extent (6 months) with a traditional error matrix including the overall estimated area of surface water/land detected and an error matrix in terms of an unbiased estimator of the proportion of area.

3.1 Overall accuracy

To thoroughly assess landcover classification maps' performance, we provide an error matrix, the proportion of mapped area for each category (water/non-water), including the user's, producer's, and overall accuracy (2). Table 1 displays an error matrix that includes the area and proportion of the mapped landcover classes for the ECJRC and OPC max extent water map. We find that the OPC 2019 water extent map detects almost 400 km^2 more water in 6 months than the Yearly ECJRC water extent map (4130 km^2 vs. 3738 km^2, Table 1). Calculating the overall accuracy of the two detection methods using the error matrix in Table 1 results in the ECJRC max extent

water map with an overall accuracy of 85.2% and the OPC max extent map with an overall accuracy of 88.6%.

Table 1. Error matrix of sample-based validation points from the ECJRC Yearly Max extent water map for 2019 and the OPC Max extent water map from May to October 2019.

ECJRC Jan - Dec 2019	Non-Water	Water	Total	Mapped Area by Class (km^2)	Proportion of the Mapped Area by Class (Wi)
Non-Water	2675	408	3083	182060	0.9799
Water	586	3076	3662	3738	0.0201
Total	3261	3484	6745	185798	1
OPC May-Oct 2019	Non-Water	Water	Total	Mapped Area by Class (km^2)	Proportion of the Mapped Area by Class (Wi)
Non-Water	2834	342	3176	181668	0.9777
Water	427	3142	3569	4130	0.0222
Total	3261	3484	6745	185798	1

Considering the error matrix using the estimated proportion of area, both datasets are highly accurate, retaining an overall accuracy of 86.7 % (ECJRC) and 89.2% (OPC) (Table 2). In the Caucasus region, the ECJRC and OPC water maps accurately detect the non-water landcover, having a commission/omission error of 13%/0.4% (ECJRC) and 11%/0.3% (OPC). In the detection of the water landcover class, the ECJRC and OPC methods' performance drops

considerably. Examining the producer's accuracy for detecting the water class, we see the ECJRC dataset correctly detects 11.5% of the surface water, omitting a large proportion of water from the map. The OPC water map performs only slightly better, having a producer's accuracy of 15.5%, indicating a loss of performance from the underestimation of surface water area.

Table 2. **The estimated proportion of area error matrix, including user's, producer's, and overall accuracy, between the ECJRC Yearly Max Extent water map for 2019 and the OPC max extent water map from May to October 2019**.

ECJRC Jan - Dec 2019	Non-Water	Water	Total	User's Accuracy	Producers Accuracy	Overall Accuracy
Non-Water	0.8502	0.1297	0.9799	0.8677	0.9962	0.8671
Water	0.0032	0.0169	0.0201	0.8400	0.1153	
Total	0.8534	0.1466	1			
OPC May-Oct 2019	Non-Water	Water	Total	User's Accuracy	Producers Accuracy	Overall Accuracy
Non-Water	0.8725	0.1053	0.9778	0.8923	0.9970	0.8921
Water	0.0027	0.0196	0.0222	0.8804	0.1567	
Total	0.8751	0.1249	1			

3.2 Accuracy Assessment by Region

The previous section evaluated the ECJRC and OPC max extent water products' overall accuracy from all validation points across the Caucasus region. However, because the spatial and temporal distribution of surface water is not uniform across the study region, it is also essential to assess the ECJRC and OPC methods' overall accuracy for each Path/Row set in the study period.

In Fig.4., we describe the overall accuracy (%) between the two datasets across all Path/Row sets from May to October.

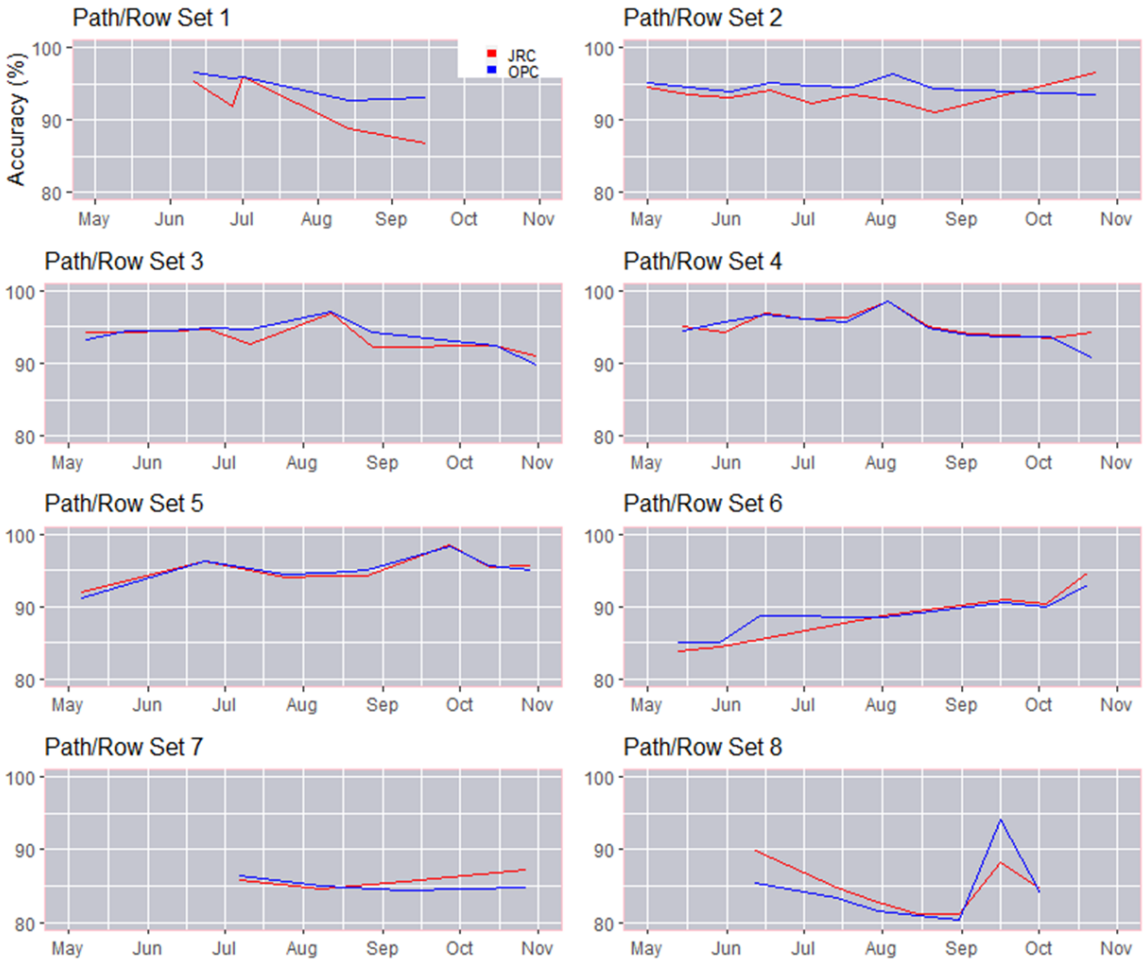

Figure 4. The overall accuracy between the ECJRC monthly water history and the OPC method from May to October 2019 reveals that the most significant differences in accuracy can be found in the far eastern and far western path/rows.

With Path/Row set 6,7 and 8 being the exception, we observe that the overall accuracy for both water detection methods remains high, ranging from 87% to 98%. In each of the Path/Row sets, the ECJRC and OPC max water extent share similar accuracy trends, each outperforming the other in various observations throughout the time series.

As mentioned, the OPC and ECJRC methods' accuracy dropped in Path/Row sets 6, 7, and 8, with the lowest overall accuracy of 80% between the sets. The three most western sets (6-8) cover the western half of Georgia and have the least amount of surface water area out of the eight total Path/Rows that comprise the Caucasus study region. Visually verifying the surface water types in these Path/Row sets with high-resolution imagery, we find that these Path/Row sets contain sinuous and nonsinuous braided river systems, with a few small reservoirs and lakes. Braided river systems can rapidly change due to seasonal flow regimes and sediment transport, causing changes in water location and discharge (87). High discharge events and water channel drift can influence the spectral signature and MNDWI values of a pixel by changing the fraction of subordinate land cover classes within the satellite image's spatial resolution (30). Variations on these river conditions can change the landcover type of a validation point, making it unreliable. A result of confining the validation points to the Landsat tiles' overlapping areas has led to a condensed sampling of the river systems in these Path/Row sets', exhibiting a loss of overall accuracy for these Path/Row sets.

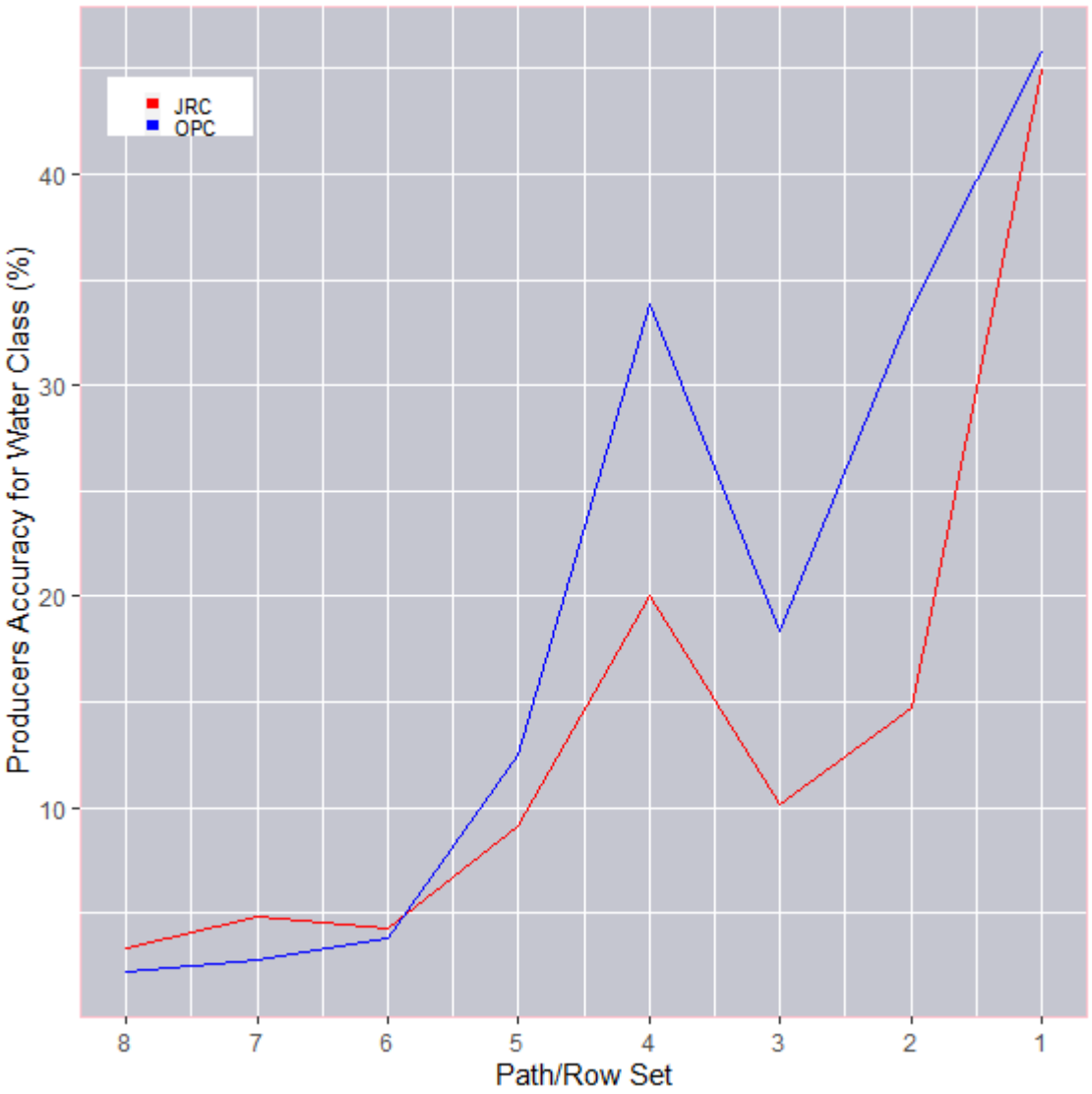

Figure 5. The ECJRC and OPC Max extent water map producer's accuracy for detecting the Caucasus region's water class for 2019. The Path/Row sets order is reversed for the graph (west to east) to assist in the spatial orientation of the paths to the region.

The ECJRC and OPC Max Extent water datasets' high overall accuracy is due to the detection of both water and non-water classes across the study area. However, the focus of these two water detection methods is to detect water. When we focus on just the accuracy of detecting the water class, the performance of these is not necessarily as significant as the high values of overall accuracy would suggest. Fig. 5 gives insight into the water detection difficulties among the

different Path/Row sets. Fig. 5 allows us to visualize the relationship between the type and temporal stability of the surface water that dominates the terrain and the loss of performance of the ECJRC and OPC water maps among the individual Path/Row sets. The path/row sections in the eastern part of our study area (1 and 4) boast the highest producer's accuracy in detecting surface water because they contain the largest and most stable water bodies to extract the MNDWI values for the model's training and validation points. The water bodies in Path/Row set 2 and 3 are temporally unstable due to heavy vegetation growth in the larger water bodies, which can vary throughout the year, causing a loss in performance. As previously mentioned, the drop in performance in the western Path/Row set (6-8) can be attributed to the braided river systems' temporal instability.

3.3 Detection of Small Water Bodies and Irrigation Canals

Evaluating the two water datasets' performance solely by statistical evaluations does not fully represent the method's ability to discriminate between the different sizes and types of water bodies present in the satellite image. Previous studies have shown that the ECJRC water dataset is highly accurate in detecting large bodies of surface water but tends to omit smaller bodies of water, including water bodies that contain significant amounts of vegetation(82,83). Examples of the omission for these types of water bodies are highlighted in Figs. 6 and 7. Fig. 6 shows the difference in performance between the two water detection methods in detecting the small floodwater areas within agricultural plots west of Zəngənə, Azerbaijan. The OPC method can detect this water, whereas the ECJRC method remains insensitive to this water type. In addition to the floodwater areas omitted by the ECJRC water map, small irrigation channels that are present can also go undetected. In contrast, the OPC method has the ability to delineate these water types, giving a more accurate representation of the surface water present in the scene.

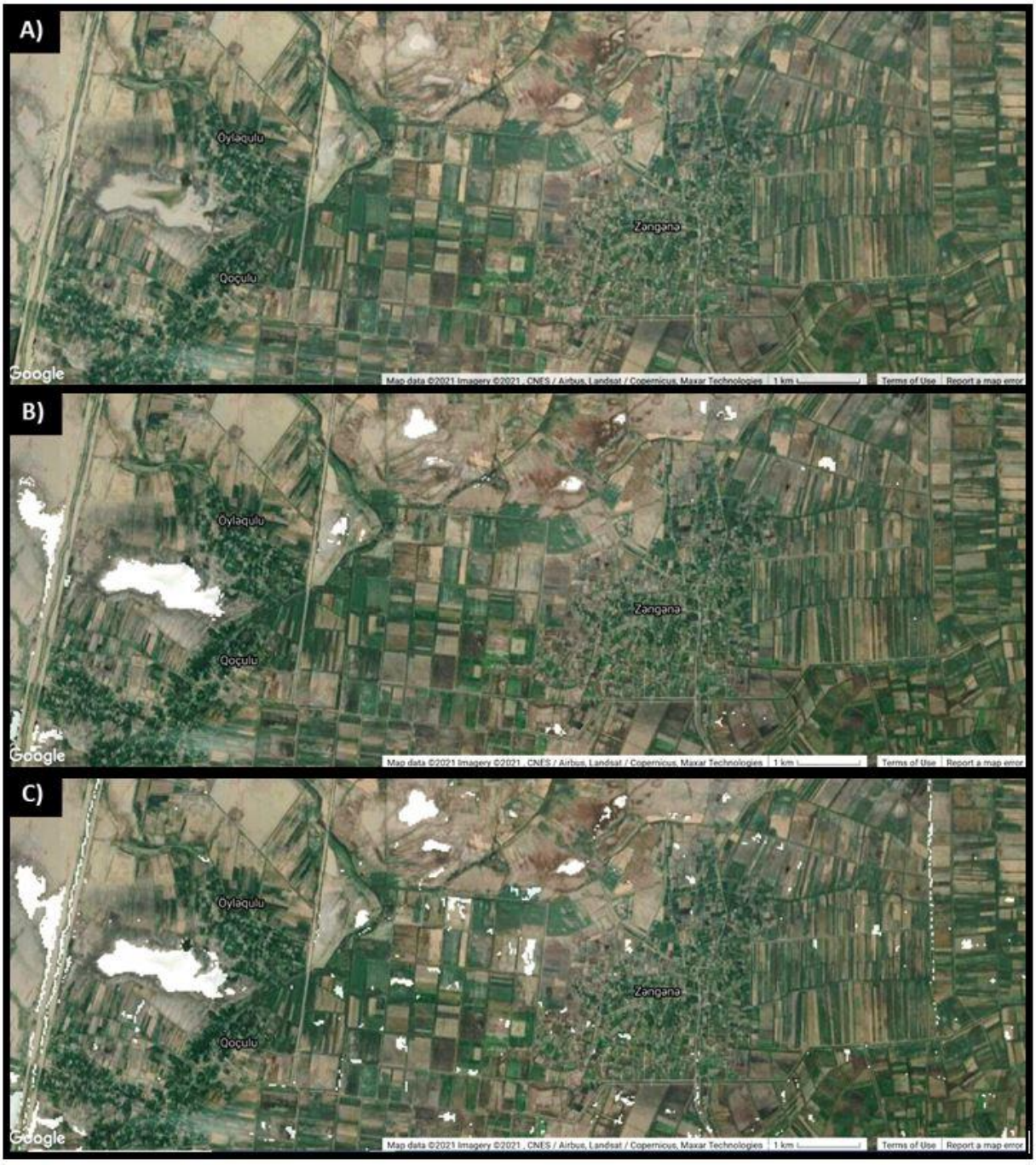

Figure 6. There is a clear difference between the OPC's and ECJRC's ability to detect small water bodies present in the scene. (A) displays the true-color Google Earth image of the area surrounding Zəngənə, in Azerbaijan. (B) displays the water detected by the ECJRC Yearly Max Extent water map for 2019. (C) shows the water detected by the OPC Max Exent water map from May to October 2019.

We find that the ECJRC dataset struggles to represent shallow, vegetated water bodies fully (Figure 7). Vegetated waterbodies are typical in Azerbaijan's agricultural areas. The presence of vegetation in water can change the water's spectral signature, depending on the distribution of its subpixel components(30). The OPC method proves to be more resilient to the spectral deviations from water pixels containing vegetation and various compositions.

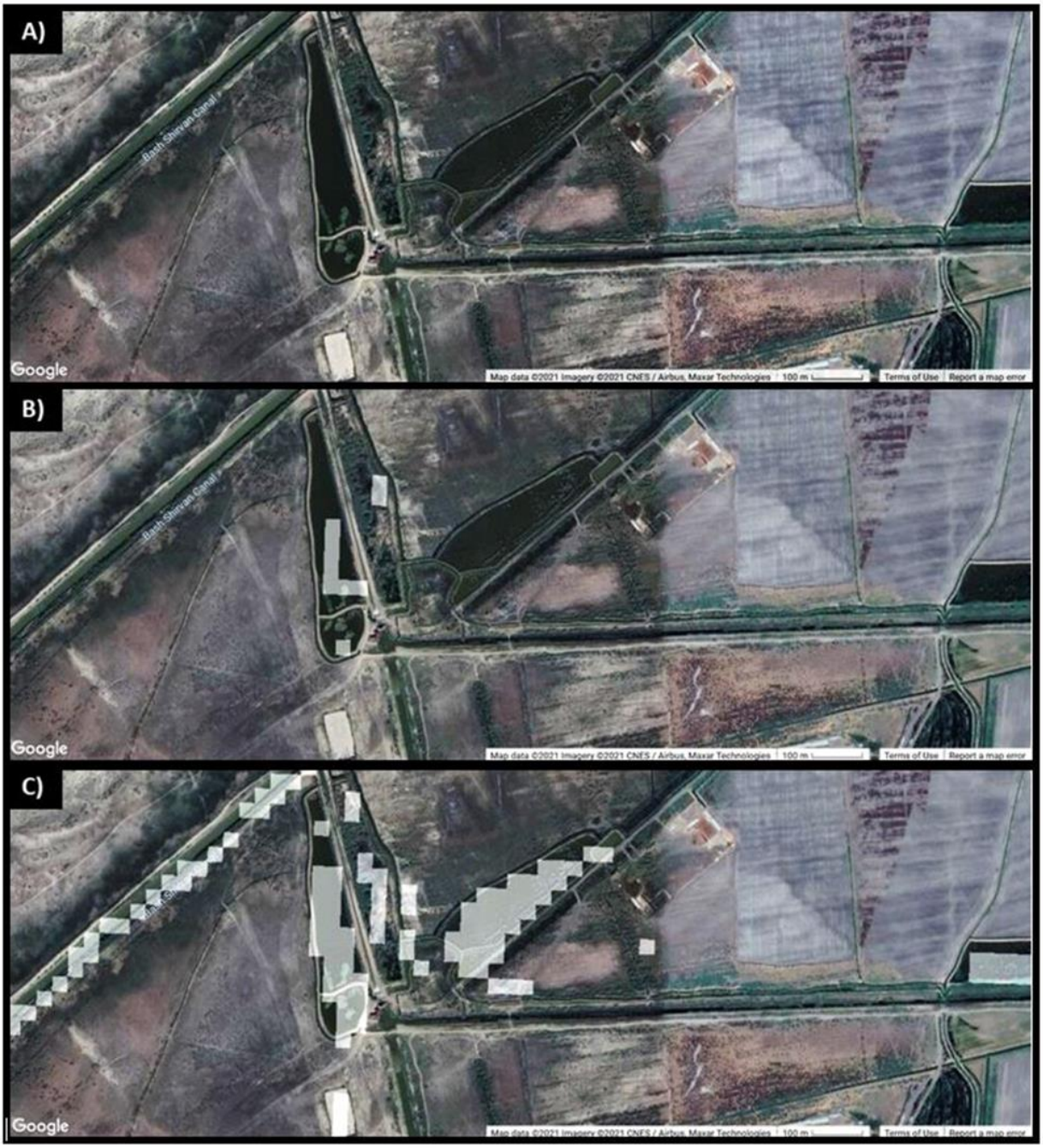

Figure 7. There is a clear difference between the OPC's and ECJRC's ability to detect small and vegetated water bodies. (A) displays the true-color Google Earth image of the area North-East of Zəngənə, Azerbaijan. (B) shows the water detected by the ECJRC Yearly Max Extent water map for 2019. (C) indicates the water detected by the OPC Max Exent water map from May to October 2019.

The difficulties in detecting surface water from the ECJRC dataset are not limited to small and vegetated bodies of water. Major water canals can also be a source of confusion in the ECJRC water map, as shown by the omission of the canal infrastructure in Azerbaijan (Figure 8). Here we can see that the ECJRC Max extent water map almost completely ignores this section of the water canal presented in the figure. Alternatively, the OPC method proves to be more reliable in detecting the water present in the canal, giving an accurate representation of the surface water present in the scene. The omission of shallow, vegetated, and unstable water bodies can significantly affect the reported surface water area and possibly alter the conclusions of studies that use this dataset in their model. Note that some of the very small irrigation channels are also not accurately detected in the OPC data.

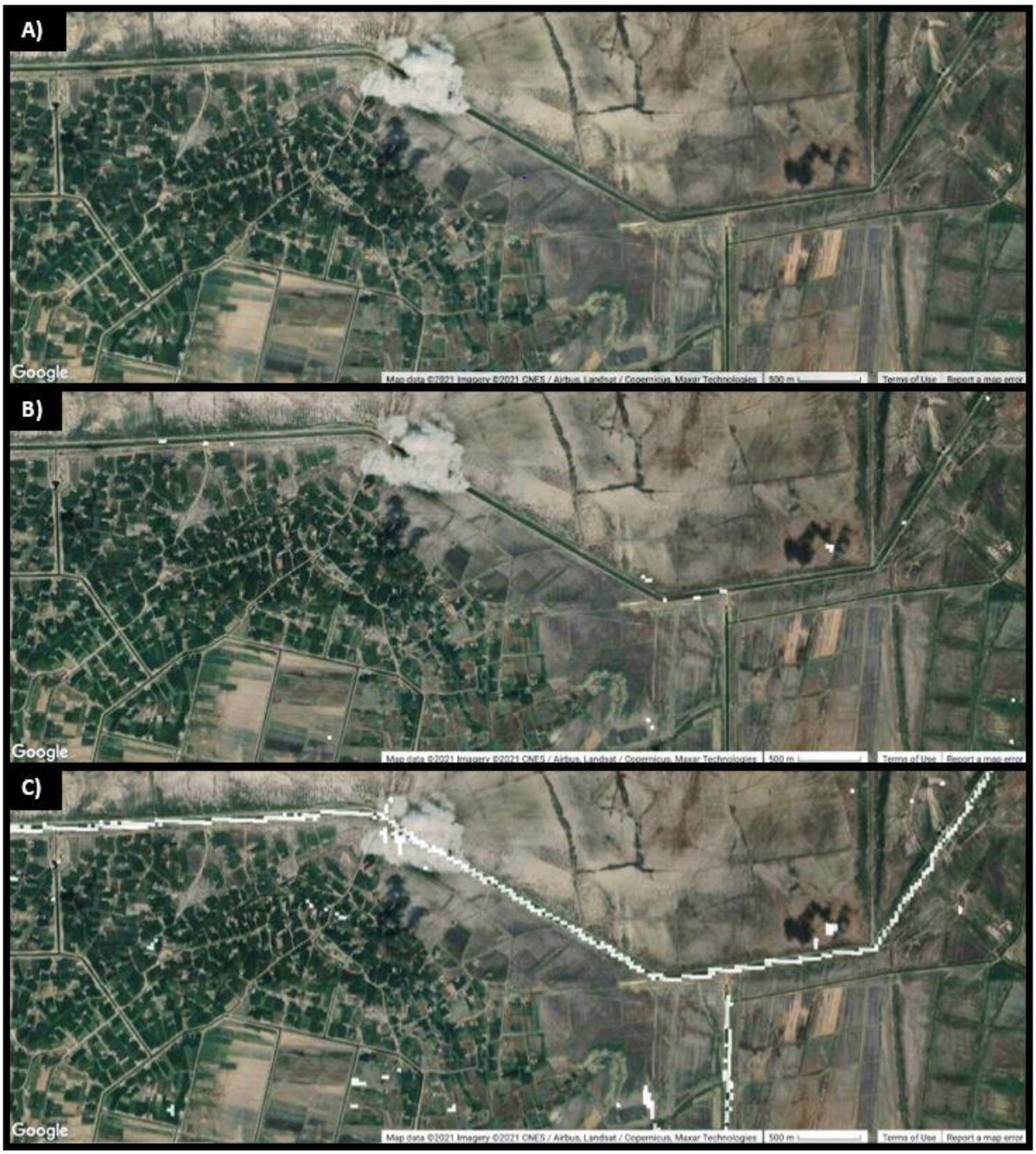

Figure 8. This figure contrasts the OPC and ECJRC methods' ability to detect water in the Bash Shirvan Canal, Azerbaijan. (A) the true-color Google Earth image of the canal. (B) displays the canal water detected by the ECJRC Max extent water Map 2019. (C) displays the canal water detected by the OPC Max extent water map from May to October 2019.

4. Discussion

4.1 Issues with Cloud Masking

Cloud presence and removal adds a level of complexity not easily resolved when conducting time series analysis. In Worden and de Beurs (2020), we selected images with zero to minimal cloud cover, but in this case, we worked in image time series, which included some cloudy images. One of the limitations of the proposed method is our use of the pixel_qa quality assessment band, which occasionally had low quality. In Fig. 9, we provide an example of clouds present in a scene but omitted from the pixel_qa mask. Located around the Yenikend Reservoir, Azerbaijan, we observe a significant amount of cloud cover present in the true-color image (A). These clouds are not present in the pixel_qa mask (B) and which causes confusion by classifying the clouds as water in the final OPC water map (C). The Landsat 8 pixel_qa band is derived using the CFMask algorithm. As reported in the Landsat 8 Land Surface Reflectance Code (LaSRC) Product Guide, there are known issues with the CFMask cloud detection relating to targets with a high albedo, including sand, building rooftops, and ice and snow(88). The CFMask also struggles to identify clouds when there is a significant difference between the cloud's temperature and its surrounding landcover (88).

Even if all cloud masks were perfect in optical imagery, the presence of clouds results in data loss and reduces land cover classification effectiveness. Using active sensors would assist in this process by recovering data loss from clouds

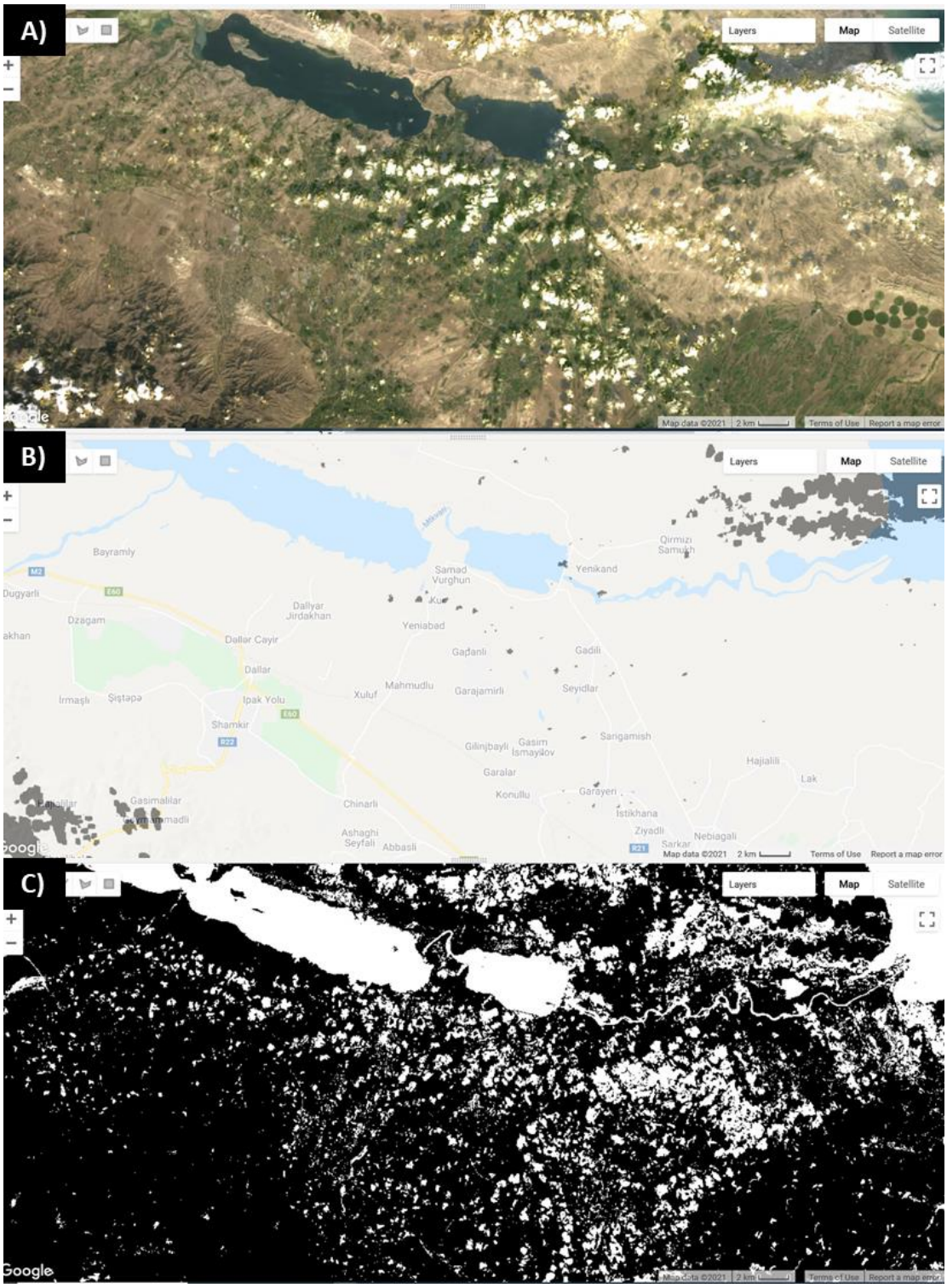

Figure 9. This figure shows the undetected cloud/cloud shadows in the pixel_qa cloud mask around the Yenikend Reservoir, Azerbaijan. (A) True -color Landsat 8 image of Path/Row set 4 acquired on October 22, 2019. (B) The pixel_qa cloud mask, where the clouds detected are represented by grey areas. (C) Final OPC water map containing cloud/cloud shadows present in the scene but omitted from the pixel_qa cloud mask.

4.3 Challenges in Water detection

Olofsson et al. (2013) found that a map can be highly accurate but can still have low accuracy in detecting individual classes due to bias, suggesting that commonly used accuracy metrics such as overall accuracy and kappa coefficient do not take full advantage of the accuracy assessment data. To utilize this data to its full potential, we must include user's, producers', and overall accuracy, area adjusted map classification error, and generating confidence intervals for the adjusted area estimates (2).

In this study, we evaluated the overall, user's and producer's accuracy, along with an area adjusted accuracy from the estimated proportion of the area. We applied these accuracy metrics to the max extent water maps from the ECJRC and OPC methods for 2019. Due to clouds' presence (masked or unmasked) in the imagery, the area calculation for each observation date can not be considered representative of surface water area from the loss of data from cloud obscuration. With a lack of complete representation of both classes' areas in the OPC water maps, we did not evaluate the accuracy of the estimated proportion of area by class for each observation in the path/row sets. As an area adjusted accuracy metric was not an option, we only included the two water detection products' overall accuracy for each path/row set date.

In comparing the overall accuracy between the two max extent water maps for the time series, we see that both datasets are highly accurate when evaluating both classes together (86.7% ECJRC, 89.2%). Despite retaining high overall accuracy across the study area, assessing the two water maps' accuracy by class tells quite a different story about the datasets' effectiveness related to water classification. When evaluating the estimated proportion producer's accuracy for water, the accuracy drops for the ECJRC and OPC max extent water maps (11.5%/15.7%), displaying a

substantial underestimation of the surface water present. Across the path/row sets (Fig. 5), we observe a significant reduction in the producer's accuracy for detecting the water class. The water maps' overall accuracy is misleading in evaluating performance because each class is weighted equally in the calculation directly from the sample count error matrix (2,89). In this region, inland water is a rare class making up around 2% of the total landcover in both water maps. The unweighted accuracy of both water datasets suggests that the water detection methods accurately detect water. However, a proportionally weighted accuracy assessment tells us the water maps are accurate in distinguishing non-water land cover, which is great if we were making non-water land cover maps. By weighing the error matrix by the proportion of estimated area, we observe a very low producer's accuracy suggesting that a large portion of water is omitted from both water maps, contradicting the viability of traditional unweighted accuracy assessments. Such a loss in the performance between the ECJRC and OPC in detecting the water class shows that it is necessary to use an unbiased estimator of the area's proportion to properly weigh each class in a sample count error matrix to avoid bias and strengthen accuracy assessments.

4.4 Building Upon Global Datasets

Global land cover datasets are becoming essential in land cover detection in remote sensing communities, with several products freely available to users (1,90–92). These datasets are especially useful as reference data in areas with little opportunity to collect sample data (86). A study focused on urban regions improved urban area maps using the European Space Agency's GlobCover product to train a classifier based on a multinomial logistic regression (86). Global datasets can be used to overcome an insufficient number of samples to train classifiers (85). For this study, we used the ECJRC monthly water history product to generate a new stratified

training dataset for each path/row to create an equal number of training points representing the water and non-water classes for each time step. Using these training points and the validation points derived from the ECJRC max extent dataset, we created a water map with improved performance compared to the parent dataset.

5. Conclusions

This study aimed to apply the OPC method across the entire Caucasus region for May – October 2019 using training and validation points (manually classified) generated from the ECJRC monthly history and yearly max extent water datasets.

Comparing the ECJRC and OPC max extent water maps, both maps are highly accurate, with an overall accuracy of 86.7% / 89.2% when applied over the Caucasus region for the time series. The OPC method proved to be more sensitive to small water body detection, detecting 392 km^2 more water in 5 months than the ECJRC max extent water map for the entire year and proving the viability of using existing global datasets to train and improve upon existing datasets to create maps with greater degrees of accuracy, giving a better representation of surface water areas.

Using the logistic probability map and OPC method, we can improve the performance of existing water indexes and existing global datasets by allowing a dynamic thresholding process that responds to the environmental conditions present to create more accurate water detection products. Arne Bomblies (2012) concluded that changes in mosquito populations could be better explained by pooled water rather than the magnitude of precipitation. Moreover, mosquitoes are also limited to areas that surround their preferred breeding habitat (39). We can use remotely sensed data to inventory the magnitude and distribution of surface water available for mosquito breeding areas. By knowing where conditions of standing water exist, we can better anticipate

the potential for human-mosquito interaction. Understanding where this interaction can occur will grant early insight into possible scenarios in the resurgence of malaria. This result agrees with the findings of Lacaux et al. (2007), stating that the use of remotely sensed data could improve predictions of malaria-based models in preventing epidemics.

In this study, there was a significant loss in water detection performance from accuracy based on the class's estimated area proportion. We are reminded that water as a landcover class remains challenging to detect from satellite imagery. Caution must be observed when establishing overall accuracy for land cover delineation. These values can be misleading, emphasizing the need for a more rigorous assessment of water detection performance metrics beyond the traditional methods of evaluating accuracy.

Chapter 4

The importance of mapping water resources is essential for anticipating trends in availability and managing human-environmental interactions. In the second chapter, we demonstrated that we could make accurate water maps by applying a logistic regression water model to the region's best-performing water index (82). Combining logistic regression with the appropriate water index, we capture the temporal characteristics of the region's water bodies to create probability maps with a dynamic threshold that responds to the environmental conditions that give an area's water its unique spectral identity. The ECJRC Global water history products are highly accurate but tend to experience confusion when detecting small bodies of water (82,83)**.** Our water detection approach applies a flexible optimal threshold method to a basic classification and regression model that produced better results than an existing global water product.

Chapter 3 expands the OPC water detection method to evaluate the Caucasus Region (Armenia, Georgia, and Azerbaijan) in a time series from May to October 2019. Global landcover datasets have been used to train and calibrate other land cover classification methods without the need to collect measurements in the field, which can be costly or inaccessible from terrain (84–86,90). The ECJRC monthly history water product was used as training data for a logistic regression water model from a stratified random sample using water and non-water as strata. Chapter 3 shows that it is possible to improve water detection accuracy by applying a dynamic probability threshold generated from training data derived from a global landcover dataset.

Both the ECJRC and OPC max extent water maps displayed a high percentage for overall accuracy and exhibited a loss of accuracy from omission error when evaluating accuracy by class. As traditional pixel counting accuracy metrics of kappa and overall accuracy can be misleading to the user, we find it necessary to apply the estimated proportion of area to sample

count matrixes to give an unbiased assessment of accuracy that provides a complete performance evaluation (2).

The OPC method's biggest weakness lay in the data loss from cloud obscuration and misclassification of clouds in the cloud mask. The issue of cloud presence in satellite imagery is two-fold. The presence of clouds affects the area calculation by either removing cloud data over water bodies from masking or the inclusion of cloud pixels in area calculation.

In the future, using a more robust cloud masking method is needed to obtain accurate max extent water maps; and the inclusion of multiple inputs from different observation platforms (passive and active) to overcome the data loss from clouds in water maps aggregated into smaller time steps.

Global land cover datasets have given us great insight into the presence, distribution, and temporal behaviors of land cover (1,91,92). These global datasets' effectiveness builds the foundation for future detection methods by providing training data for other classifiers to improve surface water detection and inventory. By understanding the temporal behaviors in quantity and distribution of water, we can better prepare for the management and allocation of water resources for potential water scarcity scenarios in the future.

Acknowledgements

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. I would also like to thank my thesis committee members for their assistance and direction in the production of my research.

References

1. Pekel JF, Cottam A, Gorelick N, Belward AS. High-resolution mapping of global surface water and its long-term changes. Nature [Internet]. 2016;540(7633):418–22. Available from: http://dx.doi.org/10.1038/nature20584

2. Olofsson P, Foody GM, Stehman S V, Woodcock CE. Remote Sensing of Environment Making better use of accuracy data in land change studies : Estimating accuracy and area and quantifying uncertainty using strati fi ed estimation. Remote Sens Environ [Internet]. 2013;129:122–31. Available from: http://dx.doi.org/10.1016/j.rse.2012.10.031

3. Westall F, Brack A. The Importance of Water for Life. Sp Sci Rev [Internet]. 2018;1–23. Available from: http://dx.doi.org/10.1007/s11214-018-0476-7

4. Chaplin MF. Water : its importance to life. 2001;5.

5. Work EA, Gilmer DS. Utilization of satellite data for inventorying prairie ponds and lakes. Photogramm Eng Remote Sensing. 1976;42(5):685–94.

6. DOI. United States Department of Interior News Release. Department of the Interior; 1966. p. 3.

7. Lauer DT, Morain SA, Salomonson V V. The landsat program: Its origins, evolution, and impacts. Photogramm Eng Remote Sensing. 1997;63(7):831–8.

8. NASA. Landsat Science: Landsat 1 [Internet]. NASA; 2021 [cited 2021 Feb 3]. Available from: https://landsat.gsfc.nasa.gov/landsat-1-3/landsat-1

9. U.S. Geological Survey (USGS). Landsat Satellite Missions [Internet]. 2021. Available from: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-satellite-missions?qt-science_support_page_related_con=0#qt-science_support_page_related_con

10. U.S. Geological Survey (USGS). Landsat Archive Hits Nine Million Scenes [Internet]. 2020. Available from: https://eros.usgs.gov/image-gallery/image-of-the-week/landsat-archive-hits-nine-million-scenes

11. Roy DP, Wulder MA, Loveland TR, C.E. W, Allen RG, Anderson MC, et al. Landsat-8: Science and product vision for terrestrial global change research. Remote Sens Environ [Internet]. 2014;145:154–72. Available from: http://dx.doi.org/10.1016/j.rse.2014.02.001

12. USGS. Landsat Missions: Landsat 8 [Internet]. [cited 2021 Nov 3]. Available from: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-8?qt-science_support_page_related_con=0#qt-science_support_page_related_con

13. U.S. Geological Survey (USGS). Landsat_Data_Policy [Internet]. 2008. p. 3. Available from: https://www.usgs.gov/media/files/landsat-data-distribution-policy-2008

14. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ [Internet]. 2017;202:18–27. Available from: https://doi.org/10.1016/j.rse.2017.06.031

15. Xia H, Zhao J, Qin Y, Yang J, Cui Y, Song H, et al. Changes in water surface area during 1989-2017 in the Huai River Basin using Landsat data and Google earth engine. Remote

Sens. 2019;11(15):1–18.

16. Wang X, Xiao X, Zou Z, Hou L, Qin Y, Dong J, et al. ISPRS Journal of Photogrammetry and Remote Sensing Mapping coastal wetlands of China using time series Landsat images in 2018 and Google Earth Engine. 2020;163(April):312–26.

17. Wang Y, Ma J, Xiao X, Wang X, Dai S, Zhao B. Long-term dynamic of Poyang Lake surface water: A mapping work based on the Google Earth Engine cloud platform. Remote Sens. 2019;11(3).

18. Deng Y, Jiang W, Tang Z, Ling Z, Wu Z. Long-Term Changes of Open-Surface Water Bodies in the Yangtze River Basin Based on the Google Earth Engine Cloud Platform.

19. McFeeters SK. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int J Remote Sens. 1996;17(7):1425–32.

20. Xu H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int J Remote Sens. 2006;27(14):3025–33.

21. Feyisa GL, Meilby H, Fensholt R, Proud SR. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sens Environ [Internet]. 2014;140:23–35. Available from: http://dx.doi.org/10.1016/j.rse.2013.08.029

22. Acharya TD, Lee DH, Yang IT, Lee JK. Identification of water bodies in a landsat 8 OLI image using a J48 decision tree. Sensors (Switzerland). 2016;16(7):1–16.

23. Ali Baig MH. COMPARISON OF MNDWI AND DFI FOR WATER MAPPING IN FLOODING SEASON Muhammad Hasan Ali Baig , Lifu Zhang , Shudong Wang , Gaozhen Jiang , Shanlong Lu , Qingxi Tong Institute of Remote Sensing Applications , Chinese Academy of Sciences , Beijing-100101 , Chin. 2013;2876–9.

24. Wang Z, Liu J, Li J, Zhang DD. Multi-SpectralWater Index (MuWI): A Native 10-m Multi-SpectralWater Index for accuratewater mapping on sentinel-2. Remote Sens. 2018;10(10):1–21.

25. Özelkan E. Water body detection analysis using NDWI indices derived from landsat-8 OLI. Polish J Environ Stud. 2020;29(2):1759–69.

26. Tulbure MG, Broich M, Stehman S V., Kommareddy A. Surface water extent dynamics from three decades of seasonally continuous Landsat time series at subcontinental scale in a semi-arid region. Remote Sens Environ. 2016;178:142–57.

27. Zou Z, Dong J, Menarguez MA, Xiao X, Qin Y, Doughty RB, et al. Continued decrease of open surface water body area in Oklahoma during 1984–2015. Sci Total Environ [Internet]. 2017;595:451–60. Available from: http://dx.doi.org/10.1016/j.scitotenv.2017.03.259

28. Townshend JRG, Justice CO. Analysis of the dynamics of african vegetation using the normalized difference vegetation index. Int J Remote Sens. 1986;7(11):1435–45.

29. Zhang Y, Feng L, Li J, Luo L, Yin Y, Liu M, et al. Seasonal–spatial variation and remote sensing of phytoplankton absorption in Lake Taihu, a large eutrophic and shallow lake in

China. J Plankton Res. 2010;32(7):1023–37.

30. Ji L, Zhang L, Wylie B. Analysis of dynamic thresholds for the normalized difference water index. Photogramm Eng Remote Sensing. 2009;75(11):1307–17.

31. Huang C, Chen Y, Zhang S, Wu J. Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. Rev Geophys. 2018;56(2):333–60.

32. World Bank Group. The Republic of Azerbaijan Climate Change and Agriculture Country Note [Internet]. 2012. Available from: www.worldbank.org/eca/climateandagriculture

33. World Bank Group. Armenia: Climate Change and Agriculture Country Note. Washington DC; 2012.

34. World Bank Group. Georgia Climate Change and Agriculture Country Note [Internet]. 2012. Available from: www.worldbank.org/eca/climateandagriculture

35. Global Health Group. Eliminating Malaria in Azerbaijan [Internet]. San Francisco; 2015. Available from: https://globalhealthsciences.ucsf.edu/pub/2015-eliminating-malaria-azerbaijan

36. Bruce-Chwatt LJ. Malaria research and eradication in the USSR. A review of Soviet achievements in the field of malariology. Bull World Health Organ. 1959;21(6):737–72.

37. Mammadov S, Gasimov E, Kurdova-Mintcheva R, Wongsrichanalai C. Elimination of Plasmodium vivax malaria in Azerbaijan. American Journal of Tropical Medicine and Hygiene. 2016.

38. Bomblies A. Modeling the role of rainfall patterns in seasonal malaria transmission. Clim Change. 2012;112(3–4):673–85.

39. Ba Y, Diallo D, Kebe CMF, Dia I, Diallo M. Aspects of bioecology of two rift valley fever virus vectors in Senegal (West Africa): Aedes vexans and Culex poicilipes (Diptera: Culicidae). J Med Entomol. 2005;42(5):739–50.

40. Costantini C, Li SG, Della Torre A, Sagnon N, Coluzzi M, Taylor CE. Density, survival and dispersal of Anopheles gambiae complex mosquitoes in a West African Sudan savanna village. Med Vet Entomol. 1996;10(3):203–19.

41. Di Rocco M, Picco P, Arslanian A, Restagno G, Perfumo F, Buoncompagni A, et al. Retinitis pigmentosa, hypopituitarism, nephronophthisis, and mild skeletal dysplasia (RHYNS): A new syndrome? Am J Med Genet. 1997;73(1):1–4.

42. Okorie PN, Popoola KO, Awobifa OM, Ibrahim KT, Ademowo GO. Species composition and temporal distribution of mosquito populations in Ibadan, Southwestern Nigeria. J Entomol Zool Stud. 2014;2(4):164–9.

43. Yves M. Tourre, Jean-Pierre Lacaux, Cecile Vignolles J-AN, Lafaye M. Mapping of zones potentially occupied by Aedes vexans and Culex poicilipes mosquitoes, the main vectors of Rift Valley fever in Senegal. Geospat Health. 2008;pp.69-79.

44. Cheng Q, Varshney PK, Arora MK. Logistic regression for feature selection and soft classification of remote sensing data. IEEE Geosci Remote Sens Lett. 2006;3(4):491–4.

45. Lacaux JP, Tourre YM, Vignolles C, Ndione JA, Lafaye M. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. Remote Sens Environ. 2007;106(1):66–74.

46. Mueller N, Lewis A, Roberts D, Ring S, Melrose R, Sixsmith J, et al. Water observations from space: Mapping surface water from 25years of Landsat imagery across Australia. Remote Sens Environ. 2016;174:341–52.

47. Jiang H, Feng M, Zhu Y, Lu N, Huang J, Xiao T. An automated method for extracting rivers and lakes from Landsat imagery. Remote Sens. 2014;6(6):5067–89.

48. Kelly JT, Gontz AM. Using GPS-surveyed intertidal zones to determine the validity of shorelines automatically mapped by Landsat water indices. Int J Appl earth Obs Geoinf. 2018;65:92–104.

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

50. World Bank Group. Republic of Azerbaijan: Climate Change and Agriculture Country Note. Washington DC; 2012.

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

52. Fisher A, Flood N, Danaher T. Comparing Landsat water index methods for automated water classification in eastern Australia. Remote Sens Environ. 2016;175:167–82.

53. Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, et al. The shuttle radar topography mission. Rev Geophys. 2007;45(2).

54. Brown de Colstoun EC, Huang C, Wang P, Tilton JC, Tan B, Phillips J, et al. Global Man-Made Impervious Surface (GMIS) Dataset from Landsat. NASA Socioecon Data Appl Cent Palisades, NY, USA. 2017;

55. Homer C, Huang C, Yang L, Wylie B, Coan M. Development of a 2001 national land-cover database for the United States. Photogramm Eng Remote Sens. 2004;70(7):829–40.

56. Alsharif AAA, Pradhan B. Urban sprawl analysis of Tripoli Metropolitan city (Libya) using remote sensing data and multivariate logistic regression model. J Indian Soc Remote Sens. 2014;42(1):149–63.

57. Lee S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. Int J Remote Sens. 2005;26(7):1477–91.

58. Peng CYJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. J Educ Res. 2002;96(1):3–14.

59. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. Optimalcutpoints: An R package for selecting optimal cutpoints in diagnostic tests. J Stat Softw. 2014;61(8):1–36.

60. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental

evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561–77.

61. Swets JA. Measuring the accuracy of diagnostic systems. Science (80-). 1988;240(4857):1285–93.

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

63. Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. BMC Med Res Methodol. 2012;12(1):82.

64. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006;27(8):861–74.

65. Sivanpillai R, Miller SN. Improvements in mapping water bodies using ASTER data. Ecol Inform. 2010;5(1):73–8.

66. Oklahoma Water Resources Board. Water Facts [Internet]. Oklahoma City; 2018. Available from: https://www.owrb.ok.gov/util/waterfact.php

67. Jones J. Efficient wetland surface water detection and monitoring via landsat: Comparison with in situ data from the everglades depth estimation network. Remote Sens. 2015;7(9):12503–38.

68. Frazier PS, Page KJ. Water body detection and delineation with Landsat TM data. Photogramm Eng Remote Sensing. 2000;66(12):1461–7.

69. Nguyen UNT, Pham LTH, Dang TD. An automatic water detection approach using Landsat 8 OLI and Google Earth Engine cloud computing to map lakes and reservoirs in New Zealand. Environ Monit Assess. 2019;191(4).

70. Bijeesh T V., Narasimhamurthy KN. Surface water detection and delineation using remote sensing images: A review of methods and algorithms. Sustain Water Resour Manag. 2020;6.4:1–23.

71. Pervez W, Khan SA, Hussain E, Amir F, Maud MA. Evaluate the capability of landsat8 operational land imager for shoreline change detection /inland water studies. Int Arch Photogramm Remote Sens Spat Inf Sci - ISPRS Arch. 2017;42(5W1):145–52.

72. Hahmann T, Wessel B. Surface water body detection in high-resolution terraSAR-X data using active contour models. Proc Eur Conf Synth Aperture Radar, EUSAR. 2010;897–900.

73. Behnamian A, Banks S, White L, Brisco B, Milard K, Pasher J, et al. Semi-automated surfacewater detection with synthetic aperture radar data: A wetland case study. Remote Sens. 2017;9(12):1–21.

74. McFeeters SK. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int J Remote Sens. 1996;17(7):1425–32.

75. Trochim ED, Jorgenson MT, Prakash A, Kane DL. Earth and Space Science tracks in northern Alaska. 2016;123–41.

76. Tu J V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol. 1996;49(11):1225–31.

77. Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: A review of classification and combining techniques. Artif Intell Rev. 2006;26(3):159–90.

78. Chen DM, Stow D. The effect of training strategies on supervised classification at different spatial resolutions. Photogramm Eng Remote Sensing. 2002;68(11):1155–61.

79. Weiss GM, Provost F. Learning when training data are costly: The effect of class distribution on tree induction. J Artif Intell Res. 2003;19:315–54.

80. DeFries RS, Hansen M, Townshend JRG, Sohlberg R. Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers. Int J Remote Sens. 1998;19(16):3141–68.

81. Pax-Lenney M, Woodcock CE, Macomber SA, Gopal S, Song C. Forest mapping with a generalized classifier and Landsat TM data. Remote Sens Environ. 2001;77(3):241–50.

82. Worden J, de Beurs KM. Surface water detection in the Caucasus. Int J Appl Earth Obs Geoinf [Internet]. 2020;91(May):102159. Available from: https://doi.org/10.1016/j.jag.2020.102159

83. Ogilvie A, Belaud G, Massuel S, Mulligan M, Goulven P Le, Calvez R. Surface water monitoring in small water bodies : potential and limits of multi-sensor Landsat time series. 2018;4349–80.

84. Inglada J, Vincent A, Arias M, Tardy B, Morin D, Rodes I. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. Remote Sens. 2017;9(1):95.

85. Yu L, Wang J, Gong P. Improving 30 m global land-cover map FROM-GLC with time series MODIS and auxiliary data sets: A segmentation-based approach. Int J Remote Sens. 2013;34(16):5851–67.

86. Li J, Gamba P, Plaza A. A novel semi-supervised method for obtaining finer resolution urban extents exploiting coarser resolution maps. IEEE J Sel Top Appl Earth Obs Remote Sens. 2014;7(10):4276–87.

87. Miall AD. A Review of the Braided-River Depositional Environment. 1977;13.

88. Landsat U.S.G.S. Landsat 8 Collection 1 (C1) Land Surface Reflectance Code (LaSRC) Product Guide [Internet]. Vol. 1. 2020. p. 1–38. Available from: https://www.usgs.gov/media/files/landsat-8-collection-1-land-surface-reflectance-code-product-guide

89. Stehman S V., Foody GM. Key issues in rigorous accuracy assessment of land cover products. Remote Sens Environ [Internet]. 2019;231(April):111199. Available from: https://doi.org/10.1016/j.rse.2019.05.018

90. Friedl MA, McIver DK, Hodges JCF, Zhang XY, Muchoney D, Strahler AH, et al. Global land cover mapping from MODIS: Algorithms and early results. Remote Sens Environ.

2002;83(1–2):287–302.

91. Hansen MC, DeFries RS, Townshend JRG, Carroll M, Dimiceli C, Sohlberg RA. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Vol. 7, Earth Interactions. 2003. p. 1–15.

92. Bartholomé E, Belward AS. GLC2000: A new approach to global land cover mapping from earth observation data. Int J Remote Sens. 2005;26(9):1959–77.