Planetary-scale surface water detection from space

Donchyts, Gennadii

DOI
10.4233/uuid:510bd39f-407d-4bb6-958e-dea363c5e2a8

Publication date
2018

Document Version
Final published version

Citation (APA)
Donchyts, G. (2018). *Planetary-scale surface water detection from space*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:510bd39f-407d-4bb6-958e-dea363c5e2a8

PLANETARY-SCALE SURFACE WATER DETECTION FROM SPACE

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 11 januari 2018 om 10:00 uur

Gennadii DONCHYTS

Master of Science in Radiophysics and Electronics,
Taras Shevchenko National University, Kiev, Oekraïne,
geboren in Lobachivka, Oekraïne.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. ir. N. C. van de Giesen

Samenstelling promotiecomissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. E. Eisemann	Delft University of Technology

Ohafhankelijke leden:

Prof. dr. J.C.J. Kwadijk	University of Twente
Prof. dr. ir. R.F. Hanssen	Delft University of Technology
Prof. dr. ir. W.G.M. Bastiaanssen	Delft University of Technology, UNESCO-IHE
Prof. dr. M. Herold	Wageningen University & Research
dr. ir. H. C. Winsemius	Vrije Universiteit Amsterdam, Deltares

Keywords: surface water, multi-spectral satellite imagery, unsupervised machine learning, k-means, Canny's edge filter, NDWI, LANDSAT, SENTINEL, SRTM, HAND, Google Earth Engine

Printed by: Gildeprint, Enschede

ISBN 978-94-6233-862-3

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

“Entia non sunt multiplicanda praeter necessitatem.“
“Entities should not be multiplied beyond necessity.“

— William of Occam

CONTENTS

Acronyms

AWEI Automated Water Extraction Index. 12, 30

BN Bayesian Network. 52, 53, 55, 57

CART Classification And Regression Tree. 5, 96, 99, 102

CPD Conditional Probability Distribution. 52, 57

CRF Conditional Random Field. 52

DAG Directed Acyclic Graph. 52

EM Expectation Minimization. 53

EO Earth Observation. 1, 8, 23, 24, 26, 115, 116

HAND Height above the nearest drainage. 3, 5, 22, 23, 37, 78, 82, 95–97, 99, 102–104, 107, 111, 116

HSV Hue Saturation Value. 12

InSAR Interferometric Synthetic Aperture Radar. 24

IPCC Intergovernmental Panel on Climate Change. 1

KDE Kernel Density Estimation. 58

LOWESS Locally Weighted Scatterplot Smoothing. 45, 46

MCMC Markov Chain Monte Carlo. 53

MNDWI Modified Normalized Difference Water Index. 12, 19, 20, 30, 36, 40, 41, 82, 95, 98, 99, 101–103

MRF Markov Random Field. 52

NDSI Normalized Difference Snow Index. 19, 20

NDVI Normalized Difference Vegetation Index. 76, 95, 99

NDWI Normalized Difference Water Index. 11–13, 20, 22, 30, 34, 35, 39–41, 48, 69, 72–76, 95, 102, 103

NED National Elevation Dataset. 21, 22, 56

OSM OpenStreetMap. 3, 5, 24–26, 88, 91, 92, 96, 107, 110

PDF Probability Distribution Functions. 52, 63

PGM Probabilistic Graphical Model. 52, 53

SAR Synthetic Aperture Radar. 3, 4, 9, 15, 24, 25, 30, 36, 47, 48, 63, 68, 69, 113

SDGs Sustainable Development Goals. 1

SRTM Shuttle Radar Topography Mission. 3, 5, 56, 78, 91, 92, 94–97, 102, 105–107, 110, 111, 114

VGI Volunteered Geographic Information. 7, 92

SUMMARY

THIS thesis studies automated methods of surface water detection from satellite imagery. Multiple existing methods are explored, discussed, and some new algorithms are introduced to allow very accurate detection of surface water and surface water changes.

The methods range in applicability from the local level to global, and from detecting high-frequency changes to low-frequency changes. Their trade-offs regarding the accuracy and applicability of the surface water detection methods are also discussed.

Several applications are presented to test the introduced methods. One of the studies focuses on a long-term global surface water change detection over the past 30 years at 30m resolution. The other application looks at the generation of a permanent surface water mask for Murray-Darling River Basin in Australia. Additionally, an in-depth validation for a small reservoir in California, USA is presented, to demonstrate performance of the new methods.

The algorithms discussed in the thesis were applied and tested to process both passive optical multispectral and active Synthetic Aperture Radar (SAR) satellite data. Combining data from all freely available satellite sensors requires harmonizations of the satellite data, but also, significant computing resources. In this thesis, Google Earth Engine parallel processing platform was used to perform most of the experiments.

We will see, that when studying surface water dynamics, the best results can be achieved by combining discriminative and generative methods of surface water detection. This way, the surface water can also be detected from satellite images where surface water is only partially visible.

In the thesis, top-of-atmosphere reflectance images are used to detect surface water. The atmospheric correction is not required when dynamic local thresholding methods are used to detect surface water.

1

INTRODUCTION

ACCURATE, efficient and high-resolution methods of surface water detection are needed for a better water management. Datasets on surface water extent and dynamics are crucial for a better understanding of natural and human-made processes, and as input data for hydrological and hydraulic models. In spite of great progress in the harmonization of freely available satellite data, efficient processing of higher level products remains inherently non-automatable.

Monitoring of surface water extent and dynamics is essential for a better understanding of natural processes, anthropogenic factors as well as climate change. This is becoming more important because water resources are under pressure from economic sectors such as industry, agriculture, energy, tourism, as well as domestic use. Furthermore, water availability is decreasing, driving more regions into water insecurity FAO [2006], McDonald et al. [2011].

Recent international agendas on climate change and environment demand objective information on planetary land and water surface conditions and changes, to be able to study the drivers behind them. The Sustainable Development Goals (SDGs) UN [2015] define seventeen challenges to be achieved by 2030, and almost half of them will directly or indirectly require up to date and high-resolution information on surface water. To name some, sustainable management for water and its use for food production, the role of the surface water for the spreading of deceases, and climate change.

The Intergovernmental Panel on Climate Change (IPCC) in the Technical Paper VI on Climate Change and Water defines knowledge gaps in observations and our understanding of the behavior of water. Moreover, with the variety and volumes of Earth Observation (EO) data available today, these gaps frequently mean the lack of proper algorithms to process existing data. The raw data need to be processed properly to derive higher level variables easily interpretable by a broader, frequently non-academic, community.

1.1. RESEARCH QUESTIONS

Many attempts have been made during the last decade to establish global scale surface water coverage and occurrence NASA/NGA [2003], Lehner and Döll [2004], Feng et al. [2016], Yamazaki et al. [2015], but most of the studies so far were limited in spatial or

temporal resolution, and accuracy. Many questions remain open, and the most important one is the lack of our understanding of how surface water has been changing during the last decades.

Accurate and fully automated detection of surface water from multi-temporal satellite data at high spatial resolution is hard. For efficient and accurate surface water detection from EO data, we need to address issues of global objectivity, accuracy, applicability, as well as provide access for a broad range of users. In the present study, this will be done by answering the following questions:

- **What are the limits of automated surface water detection methods?**
- **How to deal with typical issues such as accounting for clouds, hill shadows, snow, ice and mixed urban and rural areas?**
- **How to extract the maximum information from very noisy images, where surface water can be only partially visible?**
- **How to upscale methods to the global scale?**
- **How to streamline the use of satellite images from different passive optical and active radar satellite sensors and what observation frequency can be achieved in this case?**

1.2. CONTRIBUTIONS

The contributions can be summarized as follows:

- A new method (M_1) for accurate **surface water** detection has been developed, based on local thresholding of spectral indices computed from multispectral satellite datasets. The method is demonstrated to perform better than existing methods to discriminate surface water from noisy satellite images. The limitations and the prospects are discussed.
- A probabilistic method (M_2) is developed to **reconstruct surface water** from satellite images where surface water is only partially observed (due to limited swath, atmospheric noise or snow/ice cover). It is shown that the new method can be used to provide accurate estimates of reservoir surface water area from noisy satellite images.
- A statistical method (M_3) was developed to estimate global-scale **surface water changes** from medium-resolution multitemporal and multispectral satellite data. The method was applied to process more than a petabyte of Landsat data to identify surface water changes globally.
- An in-depth **small reservoir study** was conducted, aiming at the reconstruction of surface water area dynamics from satellite data. Here, we make use of the above two methods (M_1 and M_2) to process satellite data from multiple passive multispectral optical and radar sensors (Landsat, ASTER, Sentinel-2 and Sentinel-1). The resulting water masks were validated against in-situ water level observations. A very high correlation was obtained for both cloud-free and noisy images.

- To address global water challenges, the first planetary-scale analysis of three decades of satellite images has been performed, quantifying Earth's surface water changes at the 30m spatial resolution and occurring during the last three decades. Two areas were identified that contribute most to surface water gains (Tibetan Plateau) and losses (Aral Sea). The results of the study are made freely available in the form of a website (`http://aqua-monitor.deltares.nl`) with the help of the parallel satellite data processing platform Google Earth Engine. The study was performed by applying method M_3.

- Additionally, the same study analyzes surface water changes along the 40km coastal buffer zone globally. The Chinese coast was identified as the largest contributor to coastal changes, when aggregated by country, mainly due to land-reclamation projects.

- Permanent surface water mask for the Murray-Darling basin has been estimated using the method M_1, Height above the nearest drainage (HAND), and supervised classification for topographically difficult areas. The resulting surface water mask has been made available for inspection in the form of a website (`http://osm-water.appspot.com`). The water mask developed for Murray-Darling River Basin was compared to the surface water vector dataset extracted from OpenStreetMap (OSM) and a potential water mask derived from the 30m digital elevation model (Shuttle Radar Topography Mission (SRTM)). The positional accuracy of the rivers is estimated for three river datasets, and the results are discussed with regard to overall surface water coverage and positional differences.

Eventhough most of the methods developed during this research were applied to process multispectral satellite imagery, some of them are also applicable to process other types of imagery, such as backscatter information generated by Synthetic Aperture Radar (SAR).

The algorithms presented in this research were successfully applied to develop various surface water datasets and software tools, allowing more accurate detection of surface water and contributing to a better understanding of the Earth's surface water extent and dynamics.

A crucial aspect is also related to simplicity and reproducibility of the methods so they can be easily extended to new datasets and will optimally use existing technical infrastructure. In the present research, all methods are shared with the community with an open-source license, to ensure they can be easily reused and extended.

1.3. OUTLINE

The thesis is organized as follows. Chapter 2 reviews the relevant literature and existing methods used to detect surface water from freely-available multispectral satellite sensors. Additionally, it mentions Google Earth Engine - a parallel processing platform used to perform most of the analysis used for this research. The results discussed in the present thesis would be impossible to achieve without the adoption of this platform, which revolutionized satellite data processing, enabling planetary scale analysis for remote sensing researchers around the world.

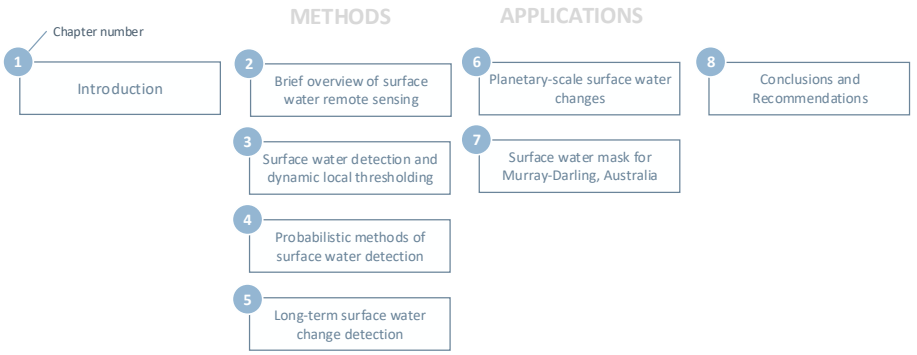


Figure 1.1: Thesis visual map

The next four chapters focus on the development of new methods for automated surface water detection. Chapter 3 studies in detail various issues related to surface water detection. Here, the method M_1 is introduced, based on the Canny edge detector and Otsu thresholding to allow very accurate detection of surface water.

We will see how surface water can be detected very accurately even for very noisy images. The method has ben validated using in-situ data available for Prosser Creek Reservoir - a reservoir in California, USA. The method was also extended to process images from multiple passive sensor multispectral satellite missions: Landsat 4, Landsat 5, Landsat 7, Landsat 8, ASTER, and Sentinel-2. It was also applied to process SAR imagery from Sentinel-1. We also constructed a simple regression model using the highest quality cloud-free images. The model was used to evaluate performance of the method in Chapter 4.

Chapter 4 focuses on the use of probabilistic methods to further improve surface water area estimates (M_2). Here we will see how a surface water mask can be reconstructed even with a small number of noise-free pixels available. This is achieved by a two-step approach, where during the first step, a high-resolution bivariate probability density function is constructed from cloud-free images. Then, a method was developed to infer the final surface water mask from the partially-observed water masks. We also validate the results by comparing it to the model constructed in the previous Chapter, to demonstrate that the method performance is improved.

In Chapter 5, a simple, yet powerful statistical method (M_3) will be introduced to analyze long-term surface water changes from a long series of images. We will see how simple surface reflectance percentile composite images can be combined with a linear regression to detect surface water changes and how these surface water changes can be distinguished from other processes, such as cloud cover or snow. The main advantage of this approach is that it allows to significantly speed-up parallel processing while preserving information about surface water changes. The method was then applied to process about 1.5 petabytes of images from medium resolution (<30m) Landsat sensors as shown in Chapter 6. The results of the study are also summarized in the form of surface water

changes aggregated per river basin globally, as well as along the coastline, aggregated per country.

Chapter 7 demonstrates the use of the methods developed in Chapter 3 to derive permanent surface water mask for 2013-2015 the for Murray-Darling River Basin in Australia. Three surface water masks will be compared, derived from three alternative sources: 30m multispectral satellite imagery (Landsat 8), OSM and SRTM. Here, a stepwise approach to detect permanent surface water from percentile composite images will be applied to analyze images measured by Landsat 8 mission. We will also see how the HAND dataset can be used as a topographic mask to detect mountainous areas where commission error of surface water detection is present due to shadowing. The resulting water mask is then refined using a supervised classification method based on Classification And Regression Tree (CART), to perform the final clean-up of the dataset.

Chapter 8 summarizes the main findings, existing challenges and trends related to methods of surface water detection as well as to the Earth Observation (EO) in general.

2

BRIEF OVERVIEW OF REMOTE SENSING METHODS AND DATASETS

This chapter provides an overview of existing remote sensing methods used to process multispectral satellite imagery in general, and to detect surface water in particular. Their main challenges, limitations, and trends will be addressed. Additionally, a short overview of the freely available medium resolution multi-spectral satellite datasets, used in the current work, is provided. A short discussion addresses the main challenges faced when analyzing satellite imagery at a global scale and high spatiotemporal resolution. To conclude, a short overview of higher level water datasets, as well as alternative water datasets, such as Volunteered Geographic Information (VGI), will be given.

Keywords: *literature overview, datasets, multi-spectral satellite imagery, spectral water indices, outline*

2.1. INTRODUCTION

WATER is one of the key sources of life on Earth, and is constantly changing. Waves batter the coast causing erosion, river banks are transformed under the flow of water, glaciers melt, resulting in new lakes and wetlands, reservoirs and harbors are constructed, and much more. Accurate estimation of these surface water changes is crucial for better understanding and management of the natural and anthropogenic processes causing them.

For decades, satellites were used to accumulate large amounts of information, resulting in multi-petabyte archives of images collected. However, only during the last decade, due to recent developments in cloud computing, we have started to transform these massive amounts of data into valuable knowledge. The main limitations that prevented analysis of these large archives of satellite data were their commercial nature. A major change that has enabled exploration of surface water at high spatial and temporal resolution was the opening of access to the Landsat mission data by NASA in 2008. Moreover, it took five more years to ensure that the first global-scale research results would be produced where these datasets could be explored to their full extent Hansen et al. [2013].

The number of satellites orbiting Earth is constantly increasing, with their technical characteristics constantly improving, resulting in better spatial, temporal, spectral, and radiometric resolutions. In the last few years, the number of EO satellites launched into the Earth's orbit, and the number of images observed has increased exponentially, mainly due to the reduction of satellite sizes and the costs required for their delivery to orbit. This rise of the satellite industry presents new scientific challenges, demanding more efficient and robust methods to process massive amounts of data. With the technological and methodological developments, the focus of the surface water analysis shifts towards planetary-scale analysis, enabling development of a more complete understanding of the Earth's natural and anthropogenic processes. Until recently Hansen et al. [2013], Pekel et al. [2016], such planetary scale monitoring and long-term estimates of land use changes with high spatial resolution were not feasible.

2.2. CLOUD COMPUTING - A NEW ERA FOR SATELLITE IMAGE PROCESSING

The massive growth in volumes of satellite data has resulted in a large demand in storage, computation and smart analytics to enable the analysis of planetary-scale data. Until recently, such analyses were performed by highly specialized scientists and engineers, and on a case-by-case basis. New cloud platforms for large satellite data analysis, such as Google Earth Engine Gorelick [2012], are rapidly removing barriers from using planetary-scale data.

The initiative to provide free access to supercomputer power for non-profit organizations and researchers was mentioned in 2009 for the first time[1], followed by the official release of the platform in 2010[2]. This platform provides access to a plethora of satellite information in three ways: (1) storage of satellite data in the cloud; (2) provision of computational resources; and (3) availability of analytical tools to process data into a clear end-user product. Since then, it has resulted in numerous academic achievements, which have helped to analyze and better understand Earth's surface changes that have

[1] http://blog.google.org/2009/12/seeing-forest-through-cloud.html
[2] http://blog.google.org/2010/12/introducing-google-earth-engine.html

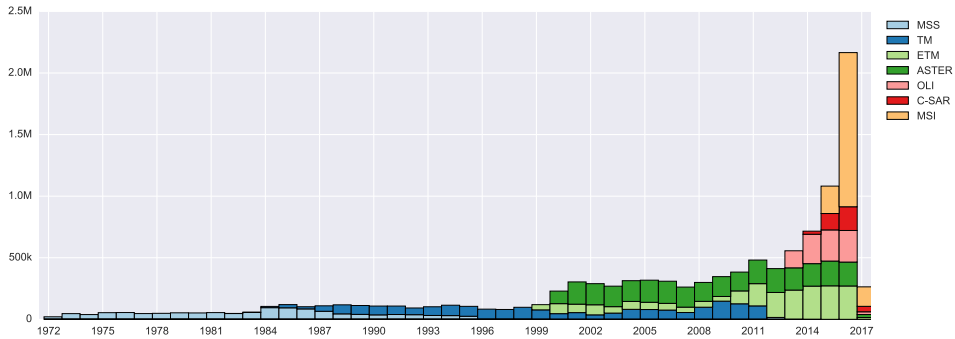


Figure 2.1: The number of Landsat and Sentinel scenes measured during the last four decades and used in the analysis

happened in the last few decades Hansen et al. [2013], Pekel et al. [2016].

The datasets used in the current research account for more than 11 million freely available medium resolution images observed by LANDSAT, ASTER, Sentinel-1, and Sentinel-2 missions from the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) during the last four decades. For several examples, images from the HYPERION hyperspectral sensor from the EO-1 mission will be used. There are only about 86,000 HYPERION images available globally, which have been acquired in the last fifteen years, but their spectral resolution is more than 20 times higher than those of other sensors. The resolution of the sensors varies from 10-30m for optical and SAR bands, and 90-120m for thermal bands. Figure 2.1 shows the total number of multispectral and SAR scenes acquired by these sensors during the past decades. The total size of these datasets as of 2017 is almost two petabytes large. Analysis of a dataset this large, requires the use of large-scale hardware and software infrastructure capable of processing it.

Using a calculation similar to Wagner and Fritsch [2015], it would take **two years** to download this dataset on a 100Mbit/s network, and then another **126 years** to process it from raw formats to Level 1 products on a single machine assuming 4Mbit/s processing speed.

2.3. MULTISPECTRAL SATELLITE SENSORS AND FREELY AVAILABLE DATA

Multispectral satellite sensors and freely available data in raw form satellite images require the use of specialized algorithms, frequently adjusted to detect specific features, quantify land use changes, or identify anomalies. The main reasons why satellite data processing is hard, is that satellite images represent a complex mix of natural and human-made processes, which are frequently very dynamic and interfere with each other. The resulting observations are often limited by technical limitations of the sensors and data processing pipelines. The presence of clouds, aerosols, and complex topographic conditions, as well as technical sensor limitations, result in multiple types of noise or gaps in the measurements, which need to be addressed before any valuable information can be extracted from the satellite data.

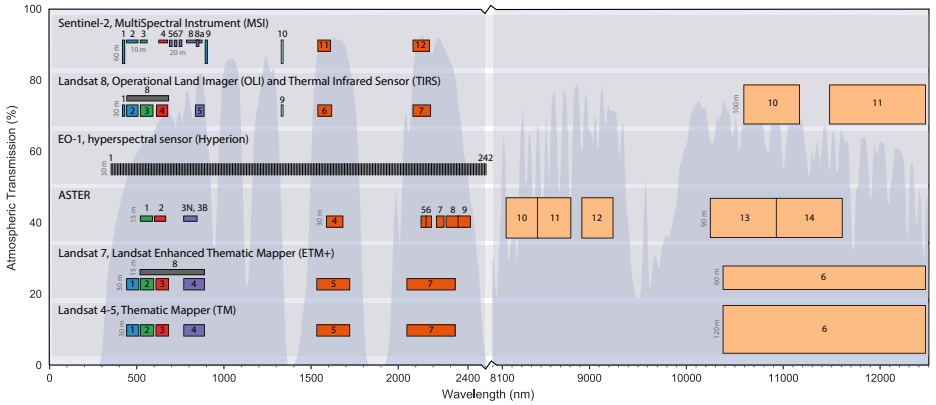


Figure 2.2: Spectral bands of the Landsat, ASTER, and Sentinel 2 satellite sensors. Atmospheric transmittance is estimated using the MODTRAN4 model as discussed in Verhoef and Bach [2003]

The present research focuses mainly on the use of multispectral satellite imagery. This includes 30m multispectral imagery measured using a Thematic Mapper (TM) sensor, used for Landsat 4 and Landsat 5 missions, an Enhanced Thematic Mapper Plus (ETM+) sensor, used for Landsat 7 mission, and an Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) sensors on board of Landsat 8. The spatial resolution of most of these sensors is 30m, except for 15m panchromatic and 90-120m for thermal sensors. For some parts of the present research, the data measured by a new 10m resolution satellite from ESA, Sentinel 2A, launched in 2015, as well as 15m resolution imagery from the ASTER mission were used. All of the sensors mentioned above are capable of measuring light in multiple spectral bands, comprising visible and infrared parts of the electromagnetic spectrum as shown in Figure 2.2.

The processing of multispectral satellite imagery usually involves many steps, including geometric and radiometric calibration. As for the Landsat and ASTER missions, the raw images show uncalibrated radiance values for every spectral band, stored as at-sensor measured digital numbers (DN_λ), which need to be converted into top-of-atmosphere (TOA) spectral radiances (L_λ) and reflectances (ρ_λ). See Chander et al. [2009] for details on the required steps. Generating spectral radiance values uses a linear formula:

$$L_\lambda = \alpha_\lambda \times Q_\lambda + \beta_\lambda \tag{2.1}$$

where

L_λ - spectral radiance at the sensor's aperture $[W/(m^2 \times sr \times \mu m]$
α_λ - band-specific rescaling gain factor $[W/(m^2 \times sr \times \mu m/DN]$
β_λ - band-specific rescaling bias factor $[W/(m^2 \times sr \times \mu m/DN]$
Q_λ - quantized calibrated pixel value $[DN]$

For multi-temporal analysis, TOA reflectance needs to be calculated to compensate

for image-to-image solar irradiance differences. These calculations can be done using the equation:

$$\rho_\lambda = \frac{\pi \times L_\lambda \times d^2}{ESUN_\lambda \times \cos\theta_s} \tag{2.2}$$

where

ρ_λ - spectral reflectance, indicating amount of solar energy reflected from the Earth's surface of wavelength λ

d - Earth-Sun distance in astronomical units [AU]

$ESUN_\lambda$ - mean spectral solar irradiance [$W/(m^2 \times \mu m$] of wavelength λ

θ_s - solar zenith angle [$degrees$]

A recent Global Monitoring for Environment and Security (GMES) initiative from the European Commission (EC) and ESA resulted in the launch of two Sentinel-2 satellites Drusch et al. [2012]. The data products provided by these satellites already include steps shown by equations (2.1) and (2.2) and represent TOA reflectances for 13 spectral bands.

For higher-level data products, additional steps may be performed, such as atmospheric corrections, usually required for a more accurate spectral analysis. Atmospheric correction allows for compensating effects caused by aerosols, for more information see Gao et al. [2006]. Additional steps may involve topographic correction, where local effects of relief are corrected for as well, such as in Tan et al. [2013].

In this study, if not otherwise specified, most of the analysis is based on the top-of-atmosphere (TOA) spectral reflectance values. As will be seen later, this provides a good compromise for water detection applications and allows the use of methods to combine products generated by different satellite missions.

2.4. METHODS OF SURFACE WATER DETECTION FROM MULTISPECTRAL IMAGES

Luckily, spectral signatures of water in most clear-sky observations are very distinctive and can easily be detected even when using only TOA data products, which are not corrected for atmospheric effects.

However, limited spectral, spatial and radiometric resolution of the sensor and many other factors may significantly influence the accuracy of the detected surface water Ji et al. [2009].

Existing methods for surface water detection from multispectral satellite data use the fact that water significantly absorbs most radiation at near-infrared wavelengths and beyond. This fact makes it easy to detect clear water employing spectral indices, such as the Normalized Difference Water Index (NDWI), McFeeters [1996]:

$$NDWI = \frac{\rho_{green} - \rho_{nir}}{\rho_{green} + \rho_{nir}} \tag{2.3}$$

where ρ_{green} and ρ_{nir} correspond to the spectral reflectance of green and near-infrared bands. By design, the index values (similar to normalized difference vegetation index (Rouse Jr et al. [1974]) vary between -1 and 1, with water appearing mostly when the index value is greater than zero.

This index should not be confused with another, also called NDWI and introduced by Gao [1996] to detect water-stressed vegetation. The Modified Normalized Difference Water Index (MNDWI) Xu [2006], appears to be more sensitive, due to the use of the shortwave infrared band instead of the near-infrared in NDWI. The authors claim that the index results in better surface water detection in urban areas when compared to the use of the near-infrared band.

$$MNDWI = \frac{\rho_{green} - \rho_{swir1}}{\rho_{green} + \rho_{swir1}} \tag{2.4}$$

While NDWI and MNDWI are among the most widely used spectral indices for water detection, many other efforts have been made trying to develop a new spectral index for surface water detection, based on a simple linear band combination Automated Water Extraction Index (AWEI) Feyisa et al. [2014], WI_{2015} Fisher et al. [2016] and based on Hue Saturation Value (HSV) transformation Pekel et al. [2014]. An excellent overview and comparison of performance for some of these indices for Australia can be found in Fisher et al. [2016]. Even though the WI_{2015} were reported to perform better than the classical NDWI index, the general conclusion of the authors was that each index was highly dependent on the composition of the validation pixels, and no index performed best across all water and non-water pixel types. The same paper mentions that very few studies were performed that compare the strengths and weaknesses of different spectral water indices.

The detection of surface water from cloud-free images is simple, but doing this for noisy images is hard. Most of existing methods, when applied to real-world satellite imagery, require manual tuning and, in general, need to be significantly adjusted to be used for planetary scale analysis. The task of surface water detection becomes much more challenging in the presence of cloud, haze, snow or ice.

2.5. WHAT ARE THE MAIN CHALLENGES OF SURFACE WATER DETECTION?

While water detection from cloud-free and ice-free multispectral images appears to be trivial, it remains a very challenging task when working with real-world images. In this case, clouds, snow, and ice may frequently be misinterpreted as water. The land surface may be covered by snow for a significant portion of the year, especially in temperate and cold climate zones. Additionally, errors of commission (false positive detection of water) can be observed in areas with shadows due to topographic conditions or the presence of clouds. Recently, Zhu and Woodcock [2014, 2012] introduced a set of methods for clouds and cloud shadow detection which are currently employed by NASA/USGS to process all Landsat images. To detect cloud shadows, these methods make use of information on satellite sensor view angle and solar zenith/azimuth and are also used to produce Surface Reflectance Landsat products[3]. These parameters, combined with elevation data, can also be used to detect hill shadows Tan et al. [2013]. An alternative approach to exclude clouds and shadows is the use of average reflectance composites instead of instantaneous images Potapov et al. [2012], Hansen et al. [2013]. However, this averaging usually damages spectral signatures of the surfaces, limiting the application of some methods, such as spectral unmixing Keshava [2003].

[3] http://landsat.usgs.gov

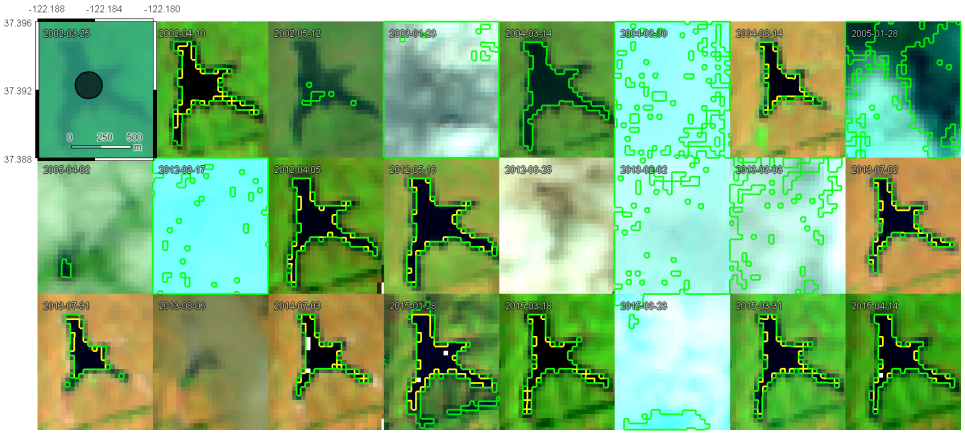


Figure 2.3: Hyperion image mosaic for 2002-2015, showing the variability of land cover, seasonal vegetation changes and hill shadow processes over Felt Lake, USA, CA. Green and yellow edges indicate values of $NDWI = 0$ computed using two band combinations: $\lambda_{b18,b54} = [528nm, 895nm]$ (green) and $\lambda_{b24,b50} = [589nm, 854nm]$ (yellow). Source: https://code.earthengine.google.com/23048e43a9cf677d0b183224ecd1dac8

2.5.1. VARIABILITY OF WATER SPECTRAL SIGNATURES DUE TO NATURAL AND ANTHROPOGENIC PROCESSES

One of the reasons why water classification from multispectral imagery is a very challenging task is that most of the observed surface water pixels are not noise-free. Mostly, it is a mixture of spectral signatures generated by a combination of different materials observed at a given time and location. Additional factors, resulting in spectral mixtures, are the physical limitation of the satellite sensors, such as spectral and spatiotemporal resolution, but also, systematic errors of the data processing pipeline. Many of the factors influencing spectral signatures of observed water bodies are caused by non-stationary events, and are impossible to predict very accurately. These factors include movement of clouds and aerosols in general, snow and ice cover changes, and suspended or diluted substances present in water.

Figure 2.3 shows a few examples of multispectral satellite images, in this case, measured by hyperspectral sensor Hyperion. For this specific area, about 50% of the images are covered or partially covered by clouds, haze or fog. The green and yellow edges show where two variations of the NDWI index are equal to zero. It can be seen that the default threshold fails to detect the water boundary correctly in most examples. This happens mainly due to noise from clouds, shadows, and mixed pixels present in the images. Additionally, significant differences can be observed between two indices constructed using slightly different band combinations. A more detailed analysis of the reasons why this happens and different automated methods to overcome this problem will be discussed in Chapter 3.

Figure 2.4 shows the actual variability of spectral curves observed for this location, converted from DN numbers to radiance, and subsequently, to reflectance values. Multiple interconnected events result in increasing complexity of recovering the actual land

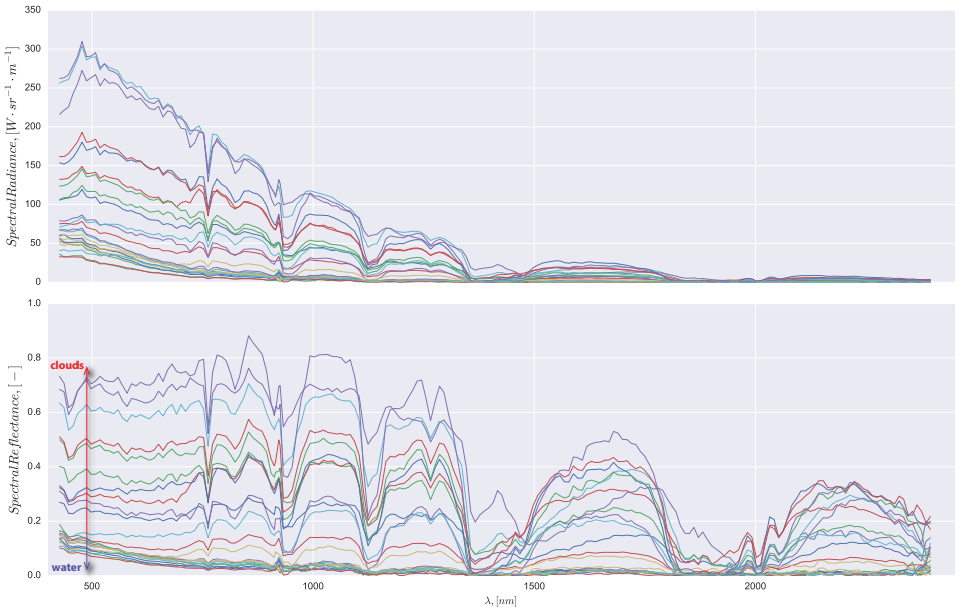


Figure 2.4: Hyperion spectral signatures measured at-sensor for the images on 2.3 sampled as a mean around waterbody center with a buffer of $r = 90m$ radius. Most of the spectral signatures with the values $\rho_{\lambda=400nm} < 0.2$ belong to water.

use classes observed at a given location and time.

Automatic and accurate reconstruction of surface water from very noisy images is, in general, a tough task, and there is no method that works perfectly for all land use types and atmospheric conditions. Most of the processes that occur in the same area and are observed by the satellite sensor are very random, with unknown distribution and are very hard to model using existing methods. However, some recent efforts in applying more advanced statistical learning methods, such as Bayesian Networks Mello et al. [2013], Conditional Random Fields Hoberg et al. [2015], Markov Random Fields Elmi et al. [2016], and Deep Learning Chen et al. [2014], look very promising.

2.5.2. SPECTRAL AND RADIOMETRIC RESOLUTIONS

When working with multi-mission satellite data, the radiometric resolution is another factor that may influence surface water detection. Radiometric resolution refers to the sensitivity of the sensor to incoming radiation, which is characterized by the minimum and maximum radiance values as well as the number of bits used to store measured values. Usually, this value varies between 8 and 16 bits. For Landsat TM, ETM+, and ASTER sensors, the effective radiometric resolution is 8-bit. For Landsat 8 and Sentinel-2, it was improved to 12 bit (stored as 16 bit). For most of the applications, the effect of these subtle changes in radiance is neglected. However, the effect of 8-bit radiometric resolution may influence thresholds used to detect surface water, especially as water is mostly represented in low radiance values.

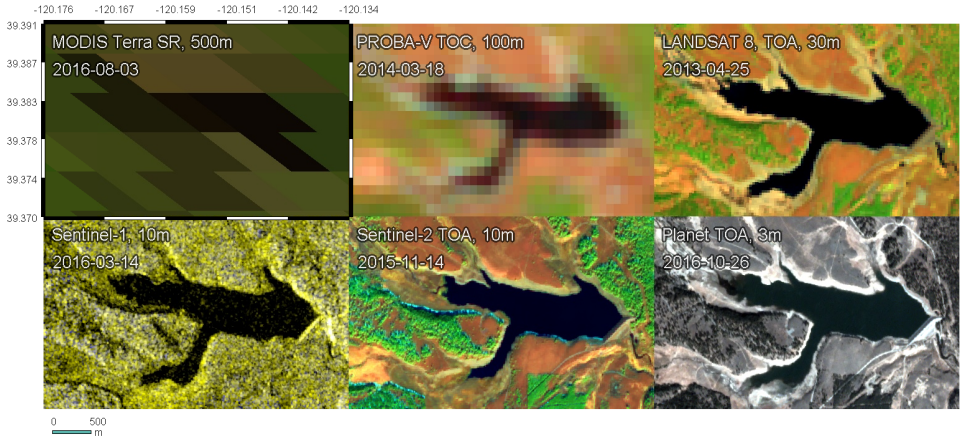


Figure 2.5: Examples of images with different spatial resolutions, from six different multispectral sensors, Prosser Creek Reservoir, Nevada County, California, USA. Source: `https://code.earthengine.google.com/93e8cd71bbc2e58584f04645e8690279`

2.5.3. SPATIAL AND TEMPORAL RESOLUTION

For the sensors studied in the present research, the spatial resolution of the sensors varies between 10m to 30m for optical and SAR bands. For thermal bands, which were used mainly for cloud and snow masking, spatial resolution varies in the range of 60m-120m.

Figure 2.5 shows a few examples of cloud-free satellite images over the same area. As can be seen, while medium (30m) to high-resolution (3m) images can be used to resolve most of the reservoir boundary variations, it may be a more challenging task to accurately estimate surface area using images with coarser resolution. However, recent developments of more advanced super-resolution image processing techniques based on convolutional neural networks can be applied to reconstruct the actual shape of the water boundary using even very coarse and damaged representations Ledig et al. [2016], Shi et al. [2016], Johnson et al. [2016]. For remote sensing applications, a recent overview of different methods based on the super-resolution paradigm can be found in Garzelli [2016]. The use of these methods goes beyond the scope of this thesis.

A temporal resolution mismatch between observations and the actual changes occurring on the land surface may significantly influence the applicability of satellite data for multi-temporal analysis. However, for some specific types of surface water changes, generative methods can be used to predict the actual state of waterbodies, by combining data from historical changes. One of these methods will be discussed in Chapter 4, where a new generative algorithm will be developed to improve the estimation of surface water extent. This method is then applied to fill in water pixels that are missing, either due to the presence of noise, such as clouds or snow or due to the limited swath width of the satellite.

2.5.4. CLOUDS AND CLOUD SHADOWS

One of the recent advancements in single-image cloud and projected cloud shadow detection is FMask[4], Zhu and Woodcock [2012], Zhu et al. [2015]. The method was so successful that it is currently used to reprocess all Landsat images acquired. To fully eliminate the influence of clouds and cloud shadows, more advanced methods are required, focusing on (*a*) detection of pixels partially or completely damaged by atmospheric artifacts (*b*) filling of missing pixels using methods of statistical inference. The FMask algorithm also provides surface water as a side product. However, the accuracy of the resulting water mask can be very inaccurate, especially, when dealing with mixed pixels, where images are only partially covered by cloud cover, as shown in Figure 2.8.

Detection and correction of effects caused by clouds and aerosols are one of the most studied topics of optical remote sensing. Clouds, haze, fog, and other substances cause absorption and scattering of the solar electromagnetic flux on its way to the Earth surface, and then on the way back to the satellite sensor. Many methods exist that provide atmospheric radiometric correction. The simplest approaches are the easiest to implement, such as dark-object subtraction (DOS) Chavez [1996], which is based on searching for the darkest object in the image and subtracting it from the radiances observed at-sensor. The drawback of this correction is that it may generate unrealistic results when the darkest pixels do not represent an actual dark object. More advanced methods involve simulation of the path radiances, using auxiliary data, such as information about aerosols **S**econd **S**imulation of a **S**atellite **S**ignal in the **S**olar **S**pectrum (6S) Vermote et al. [1997], Zhang et al. [2012] or **MOD**erate resolution atmospheric **TRAN**smission (MODTRAN) Berk et al. [1987].

Several satellite missions, such as Landsat, MODIS, and PROBA-V, already provide higher-level surface reflectance (SR) products, where images are corrected for atmospheric effects. Additionally, some of the data providers include a quality assessment (QA) band, which indicates if the pixel is covered by clouds. However, the use of atmospherically corrected images from different satellite missions may result in a mismatch between images, even though the resulting images look more appealing and better match the ideal spectral signatures of different land cover types. At the time of the writing of this thesis, the official NASA/USGS Landsat SR products were still provisional. The documentation also mentions that the efficacy of SR correction is likely to be reduced in areas where atmospheric correction is affected by adverse conditions[5]:

(*a*) SR is not run for a scene with a solar zenith angle greater than 76° (*b*) SR quality reduces for data acquired over high latitudes (> 65°) (*c*) The panchromatic band is not processed to SR (*d*) Hyper-arid or snow-covered regions (*e*) Low sun angle conditions (*f*) Coastal regions where land area is small when compared to adjacent surface water (*g*) Areas with extensive cloud contamination

Furthermore, there are issues with artifacts in the data over certain geographic areas, specifically inland water bodies, areas of high relief, and areas with high aerosol abundance.

The above issues with SR Landsat products, the inconsistency between different SR

[4]FMask algorithm source code: `https://github.com/prs021/fmask`

[5]Google Earth Engine Landsat 5 SR Dataset: `https://code.earthengine.google.com/dataset/LANDSAT/LT5_SR`

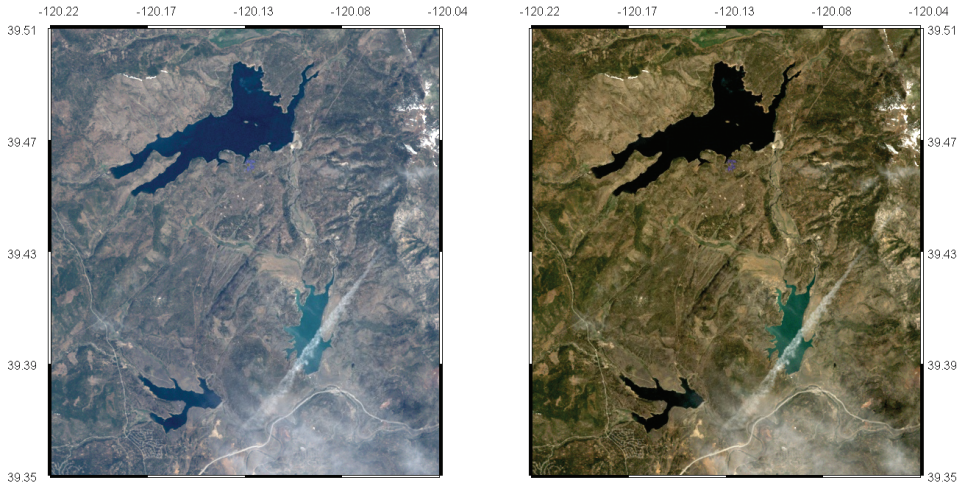


Figure 2.6: TOA reflectance (left) and surface reflectance (right) false-color composite image (swir1, nir, green). Source: `https://code.earthengine.google.com/e09497a0f8cd6d20b1ce3b29600b958d`

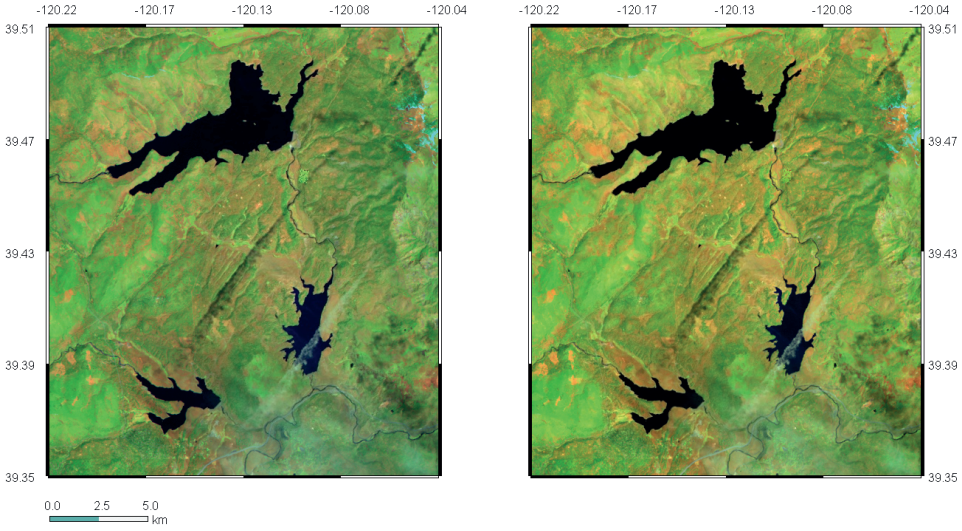


Figure 2.7: TOA reflectance (left) and surface reflectance (right) true-color composite image. Source: `https://code.earthengine.google.com/e09497a0f8cd6d20b1ce3b29600b958d`

methods used for different satellite products (Sentinel-2, Landsat) as well as missing SR products in some sensors (ASTER, EO-1) all resulted in this study mostly focusing on the use of TOA reflectance images. The main reason for this is that most of the higher level products use different atmospheric correction algorithms, which does not necessarily lead to improved classification. Additionally, for single-image based classification of surface water using adaptive methods, based on the maximum likelihood principle, this type of correction may not be required Song et al. [2001].

It is important to note that for water detection the use of TOA can even be preferable, especially when working with multi-mission satellite products. The main reason for this is that the auxiliary datasets used to generate SR products are usually of a coarser spatial resolution, which may introduce additional error to the image, even though the resulting spectral signal may be better.

In the next chapters we will see that, even without SR correction, the TOA images can be used to detect surface water very accurately when automated methods of dynamic thresholding are applied. Additionally, the use of SR products for surface water detection may not be required when working with multi-spectral satellite imagery, because most existing methods of surface water detection may use near-infrared and infrared bands, which are less sensitive to aerosols. The effect of SR correction is much more evident for the visible part of the spectrum (Figure 2.6) when compared to false-color composite using infrared and near-infrared bands (Figure 2.7).

Eventhough existing methods of SR correction can significantly improve radiometric properties, most of the methods are based on a single image only (ACCA, Irish et al. [2006]), which assumes that significant amounts of aerosols remain uncorrected, as well as secondary effects, such as cloud shadows.

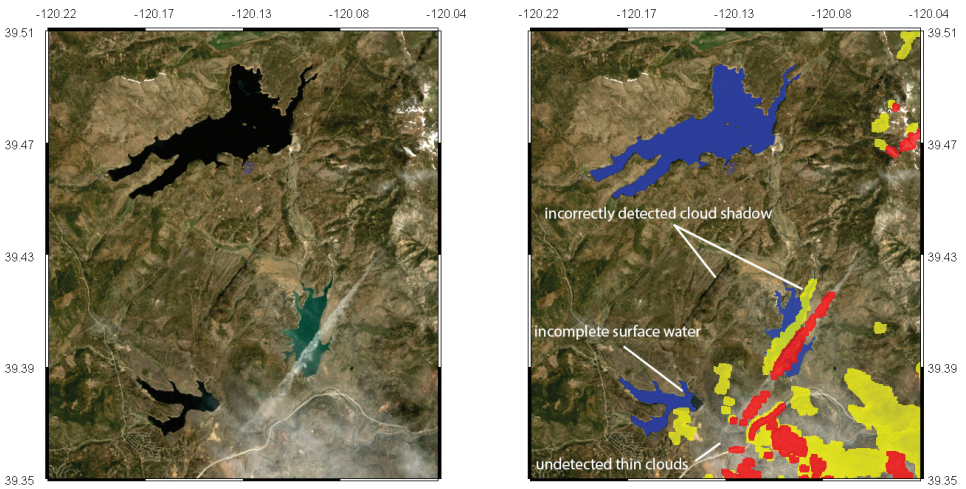


Figure 2.8: (red), cloud shadows (yellow) and surface water mask (blue) as detected by FMask algorithm. Source: `https://code.earthengine.google.com/b74dc921e6725cffd1b29bca5b6d3c43`

Some of the recent enhancements of the methods include the use of multi-temporal

methods to further improve the accuracy of cloud and cloud shadow detection methods, TMask[6], Zhu and Woodcock [2014]. The main idea of the method is to extend the multi-step FMask algorithm with a new step, where a harmonic model is used to simulate the most probable reflectance values. The method is based on the seasonal changes of the reflectance values for pixels covered by vegetation. However, the authors mention that the method may falsely identify ephemeral pixels as clouds or cloud shadows. In our experience, the applicability of the method is limited in areas with significant surface water dynamics. Additionally, the authors mention that the method performs extremely slowly, making it nearly impossible to scale for global analysis where millions of satellite images need to be processed. We will discuss in Chapter 4 how probabilistic models can be used to identify clouds and cloud shadows as an alternative to this approach.

A comprehensive study on the cloud cover variability has been performed recently by Wilson and Jetz [2016][7]. Even though the study focuses on a 1km spatial resolution only, it provides detailed intra- and inter-annual analysis of cloud cover frequency. These parameters can be crucial when developing data-driven models for cloud mask and cloud shadow detection. However, methods of high-resolution cloud and cloud shadow detection are still being developed, and no universal method yet exists that can provide an automated way to detect clouds and cloud shadows at high spatial resolution.

2.5.5. SNOW AND ICE

Snow and ice can cover a significant portion of waterbodies in temperate and cold climates and need to be detected accurately to prevent misinterpretation of detected surface water. In general, snow/ice can be easily detected. First, by using the thermal band, because most waterbodies freeze at zero degree Celsius under normal conditions. In addition, methods of snow and cloud detection have been studied for a long time using a Normalized Difference Snow Index (NDSI), Valovcin [1976], Hall et al. [1995], defined as:

$$NDSI = \frac{\rho_{green} - \rho_{swir}}{\rho_{green} + \rho_{swir}} \tag{2.5}$$

where swir and green bands correspond to wavelengths of $\lambda_{green} = 660nm$, $\lambda_{swir} = 1600nm$. It must be noted that the MDNWI spectral index in 7.2 exactly repeats the definition of the NDSI index. This fact makes it more difficult to discriminate water from snow when using the MNDWI index. In general, a good separation still can be achieved by using near-infrared as well, which is less sensitive to snow/ice content.

[6] TMask algorithm source code: `https://github.com/prs021/tmask-algorithm`
[7] `http://www.earthenv.org/cloud`

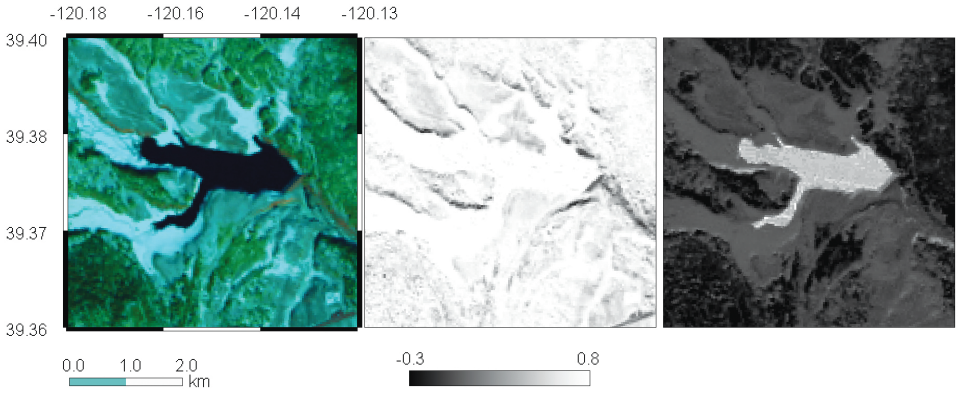


Figure 2.9: An example of a Landsat 5 image from December 17, 2000 TOA composite image (left: swir1, nir, green) demonstrating how the NDSI / MNDWI and NDWI spectral index values vary over a snow-covered waterbody. Source: `https://code.earthengine.google.com/3b5de713c26202cad46e5ddfc5deb7c2`

As can be seen in Figure 2.9, the NDWI spectral index provides a more reliable way to discriminate between surface water and land pixels when snow and ice are present. This significantly limits the applicability of the MNDWI spectral index, to only the multispectral images measured during warm seasons of the year, or to temporal composite images where snow cover is eliminated.

2.5.6. TOPOGRAPHIC EFFECTS

Topographic illumination effects, such as shadows from hills or buildings, may significantly disturb the signal measured by satellite sensors, causing false-positive surface water detection. Additional illumination correction, known as topographic correction, is required to compensate for these effects. Illumination effects can be corrected by compensating for solar radiance due to topographic effects, usually described by the Bidirectional Reflectance Disturbance Function (BRDF), which defines how light is reflected at the surface. In remote sensing, simplified formulas were introduced to compensate topographic effects for Lambertian and non-Lambertian surfaces. Some of the best known are the cosine model and C correction models Teillet et al. [1982]. Other models were developed later, such as the Minnaert, and SCS+C Smith et al. [1980], Teillet et al. [1982]. A good overview of some of the methods can be found in Gao and Zhang [2009], Soenen et al. [2005].

Most of the topographic correction models are based on the use of the relative solar incidence angle, or illumination conditions, defined as:

$$IC = \cos(\theta)\cos(\alpha) + \sin(\theta)\sin(\alpha)\cos\left(\phi_\theta - \phi_\alpha\right) \tag{2.6}$$

where θ is the solar zenith angle, α is the topographic slope angle (0 = horizontal), ϕ_θ is the solar azimuth angle, and ϕ_α is the aspect angle of the topographic surface (0 = north). The resulting illumination variable IC varies between -1 and 1. The simplest topographic correction methods assume that the surface is Lambertian (ideal surface, diffusely reflects light in all directions, isotropic). The cosine model normalizes the re-

flectance of any pixel based on the assumption that the total irradiance received at a given location is directly proportional to the cosine of the incidence angle, the angle between the normal to the pixel surface and the solar zenith direction Teillet et al. [1982]:

$$L_n = L\frac{cos(\theta)}{IC} \tag{2.7}$$

where L_n is the corrected reflectance and L if the observed reflectance on the inclined surface. The cosine model assumes that the surface is Lambertian and is independent of wavelength. Further development of topographic correction algorithms resulted in the introduction of more advanced, semi-empirical models, such as C correction Teillet et al. [1982]. A variation of the empirical C model was then introduced by Tan et al. [2013], which assumes non-Lambertian surfaces and is based on empirical rotation in the reflectance/illumination space. In topographically complex areas, the use of this model can significantly reduce commission errors in surface water detection applications. To demonstrate the effects of the topographic correction on surface water detection, an empirical rotation model developed by Tan et al. [2013] is applied to process a Landsat image focused on a reservoir located in a mountainous region (figures 2.10 and 2.11).

$$L_n(\lambda) = L(\lambda)\frac{cos(\theta) + C(\lambda)}{IC + C(\lambda)} \tag{2.8}$$

where λ is a band-specific wavelength and the variable $C = b/a$ is an empirical coefficient equal to the ratio of slope a and intercept b of a linear regression performed on two images with different illumination conditions:

$$L(\lambda) = a(\lambda) \cdot IC + b(\lambda) \tag{2.9}$$

To demonstrate results of the application of this model, we have applied it to correct satellite images in a hilly area around a small reservoir in California, USA, Lake Piru.

The results of the topographic correction can be seen in figure 2.10, and corrected reflectance values for the near-infrared band are shown in figure 2.11. As we can see, the topographic correction step significantly corrects original reflectance values. The correction was performed with the help of the 3m National Elevation Dataset (NED) Gesch et al. [2002].

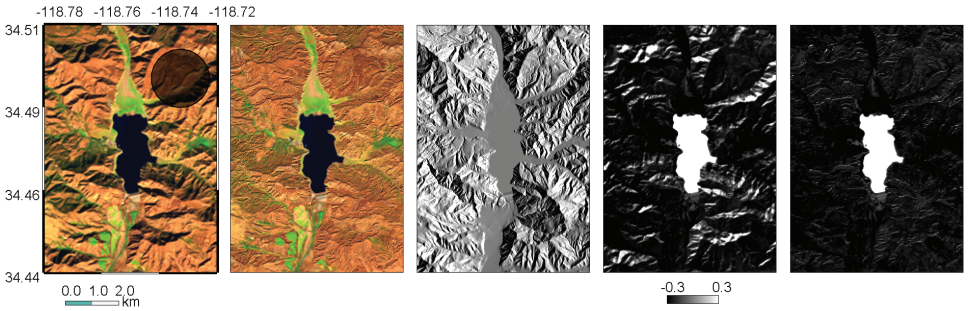


Figure 2.10: An example of the topographic correction algorithm Tan et al. [2013] applied to a Landsat 8 false-color composite image (swir1, nir, green) acquired on November 5, 2013. Original image (A), corrected image (B), hill shade from NED (C), NDWI computed from the uncorrected image (D), and NDWI computed from corrected reflectances (E). The dark circle indicates pixels used to produce the chart in figure 2.11. Source: `https://code.earthengine.google.com/90a3782edde7c0ef5ca895c800c1a688`

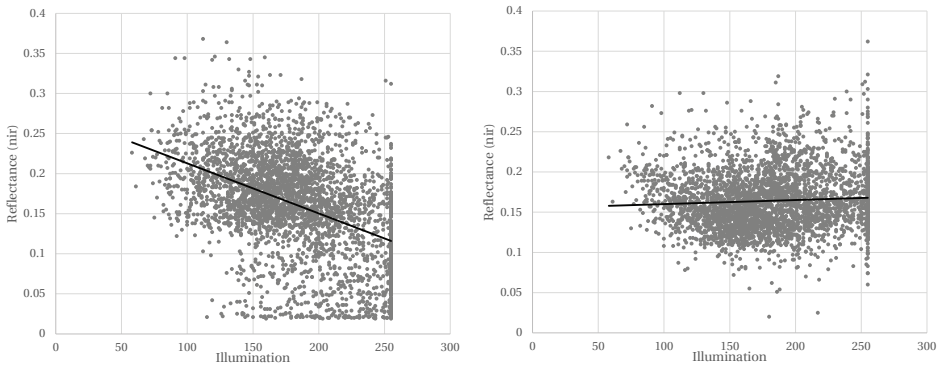


Figure 2.11: Original (left) and corrected (right) values for near-infrared band aggregated over the marked circle area in 2.10

The above model provides better correction of topographic effect for multispectral images (Tan et al. [2010]). However, the corrected image may still contain small artifacts caused by errors in the elevation model and more complex light interactions. It is important to note that topographic correction may need to be avoided if a significant variability of surface water mask between images takes place. In this case, topographic correction may introduce local artifacts for the pixels where a mismatch occurs between the digital elevation model and the surface water edge.

For surface water detection, an alternative method is to exclude pixels where the topographic correction may be required. One way to perform this kind of masking is by using the HAND topographic index Rennó et al. [2008], Nobre et al. [2011], following Westerhoff et al. [2013].

Even though the empirical illumination correction method described above was hardly used for the research presented in this thesis, it can be very useful to generate a high-frequency surface water dynamics for waterbodies locate in hilly areas in addition to the

methods discussed in chapters 3, 4 and 5. As an alternative, in chapters 5 and 6 we will use HAND to mask out errors in hilly areas for a global surface water change detection study. Additionally, HAND is utilized in the Chapter 7 to identify hilly areas where an additional unsupervised classification step is required to refine the surface water mask.

Even though topographic correction algorithms may significantly improve the overall image illumination for hilly areas, they may also introduce local artifacts when detecting surface water. The main reason for this is that the digital elevation models are mostly measured at a specific time only, but the actual waterbody boundary may vary a lot for ephemeral water. To avoid these errors, additional steps are required, for example those described in Chapter 4, where density-based methods will be used to infer the actual water mask.

Figure 2.12: Lake Piru, California, USA. [8]

In some cases, it is possible to avoid the topographic correction step, for example, when analyzing surface water changes using multi-temporal composite images, based on a large number of observations. However this is only possible under the assumption that the images are generated using image collections where illumination parameters (sun elevation and azimuth) are equally distributed. This will be discussed in more detail in Chapter 5 and 6.

2.6. GLOBAL STATIC AND DYNAMIC SURFACE WATER DATASETS

Global-scale surface water and its dynamics have been very actively studied in the past few years. This includes estimation of different static surface water masks, such as 250m SWBD Farr et al. [2007] and MOD44W Carroll et al. [2009], as well as 30m GLCF Feng et al. [2016] and 90m G3WBM Yamazaki et al. [2015]. An overview and comparison of some of these datasets can be found in Lamarche et al. [2017]. Advances in big data technologies and the massive growth of available satellite data have resulted in the appearance of new high-resolution datasets, with a significant increase in both temporal (from static to monthly) and spatial resolutions (30m) Pekel et al. [2016]. An overview of some of the static and dynamic surface water products can be found in Yamazaki and Trigg [2016].

The upcoming SWOT mission (2021) promises to deliver twice every two weeks, monitoring of 90% of global water bodies smaller than 100m. However, this will not answer the question as to how surface water has changed during the last four decades. Moreover, the accurate knowledge of historical surface water dynamics is very important to fill in gaps in the trending hyper-resolution hydrological activities Wood et al. [2011], Bierkens et al. [2015]. A good overview of the recent advances and trends related to satellite-based EO for hydrology, in general, can be found in McCabe et al. [2017].

In addition to optical and radar surface water mapping, numerous studies were per-

[8] Source: Flickr, http://www.flickr.com/photos/tomsaint/3281932913

formed focusing on the reconstruction of surface water elevation changes using satellite altimetry: HydroWeb [9] Cretaux et al. [2005], DAHITI [10] Schwatke et al. [2015], G-REALM [11] Birkett et al. [2011].

Recently, multiple global studies were performed where a hybrid approach was implemented to reconstruct reservoir and lake storage dynamics from satellite altimetry combined with optical multispectral satellite data Khandelwal et al. [2017], GRanD Viewer [12]. An overview of some of the methods used to monitor large lakes and reservoirs as well as some challenges can be found in Karpatne et al. [2016]. Integration of the data for both water level and surface area changes can be used to generate high-resolution storage changes Duan and Bastiaanssen [2013]. In many cases, having only surface water extent changes should be sufficient for the reconstruction of the waterbody storage dynamics Liebe et al. [2005], Annor et al. [2009]. However, for accurate estimates of water storage dynamics, relations between volume and area need to be estimated very accurately. Some of the recently developed methods include estimation of reservoir bathymetry from SAR data, through Interferometric Synthetic Aperture Radar (InSAR) algorithms Amitrano et al. [2014] or by estimating volume/area curves based on topographic similarities Bemmelen et al. [2016]. Estimating area/volume curves is currently one of the gaps to be filled using EO methods. This information is crucial for improving the local relevance of global hydrological studies.

2.7. EARTH OBSERVATIONS AND VOLUNTEERED GEOGRAPHIC INFORMATION

Many global surface water mapping efforts rely on the information derived from EO datasets, such as optical multispectral imagery, SAR imagery or radar altimetry. A combination of remote sensing with existing GIS vector data sets that contain information on water occurrence gets much less attention. GIS vector data sets are frequently measured using high-resolution GPS devices or by manually digitizing aerial or satellite imagery. As a result, these datasets usually show much better precision and contain semantic information, such as feature names, types, and other attributes. Their quality and completeness are not uniform across the globe. One key global vector data set containing water information is OSM Haklay [2010], initiated in 2004 and currently including more than 3 billion objects. From these, more than 8 million objects relate to water OpenStreetMap [2016b]. Many environmental applications use OSM, including the extraction of paved areas and surface water coverage for hydrological applications Schellekens et al. [2014].

The volunteered nature of OSM is the main factor making it less adopted by GIS professionals Mooney et al. [2010] stressing the importance of the development of automated methods and tools to validate its quality in comparison to other datasets. A good example of how OSM quality can be assessed can be found in Girres and Touya [2010], showing how positional differences between linear and polygonal features can be addressed. Furthermore, the "increasing buffer" method, Goodchild and Hunter [1997] can be used to estimate the quality of linear features. An excellent overview of papers focusing on OSM quality can be found in Barron et al. [2014]. The latter also suggests us-

[9] HydroWeb: http://www.legos.obs-mip.fr/soa/hydrologie/hydroweb/
[10] DAHITI: http://dahiti.dgfi.tum.de/en/
[11] G-REALM: https://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/
[12] GRanD Viewer: http://umnlcc.cs.umn.edu/GrandViewer/

ing elements of ISO 19113:2002 [13], such as completeness, error of commission/omission or positional differences, to evaluate the OSM quality. Unfortunately, to our knowledge, no academic literature exists focusing solely on the quality of water features present in OSM and using global or nearly global remote sensing datasets.

In Chapter 7, we will compare surface water masks derived from OSM to see if it can be used as a complementary dataset to generate global coverage, eventhough its local coverage and quality may vary.

2.7.1. RESERVOIRS AND THEIR STORAGE DYNAMICS

Eventhough many of the global waterbodies may change over time, most of the existing surface water vector maps are static, not providing any information on temporal spans where these features are valid. However, many of these vector datasets are incomplete in terms of spatial coverage. On the other hand, satellite data provide a way to extract spatiotemporal variability of these waterbodies. However, semantics is missing in satellite-based products that is available in vector topographic maps. Semantics includes many attributes, such as the names of the waterbodies, the dam construction date, administrative information and much more. Some of these attributes, like construction time or surface area variability, can be extracted from satellite data. Others, like names or technical characteristics, need to be collected manually. At present, a hybrid approach to generate the best quality surface water maps seems to be the most promising, combining multi-temporal satellite information with existing vector maps, while also including results from numerical models to estimate the variability of water-related parameters.

A lot of effort has been made in trying to map water bodies at a global scale. However, existing databases are still scarce and fragmented, as can be seen in Figure 2.13[14]. Here we show the location of reservoirs collected by different databases, including 6859 reservoirs and lakes in GRaND Lehner et al. [2011], 4668 dams collected on Wikipedia[15], more than 150000 reservoir and dam locations available in OSM OpenStreetMap [2016b], and 33684 global dams collected by King's College London Mulligan et al. [2009].

For many of these reservoirs, detailed data on storage changes are not available, yet are required for an automated setup, generation and calibration, of surface and subsurface water models. In turn, these models are essential to optimize water management and to answer challenges related to floods and water stress in the 21^{st} century Hirabayashi et al. [2013].

Most of the existing reservoir datasets are still incomplete, either regarding quality or coverage, with the result that existing global hydrological models, such as eWaterCycle[16] and GloFAS[17], still make use of only coarse databases like GRaND, while missing many small to medium size reservoirs Emerton et al. [2016]. A number of efforts were made recently trying to develop methods to map and reconstruct storage of reservoirs from optical multispectral and SAR imagery (Duan and Bastiaanssen [2013], Eilander et al. [2014], Amitrano et al. [2014], Khandelwal et al. [2017]). However, applying these meth-

[13] http://www.iso.org/iso/catalogue{_}detail.htm?csnumber=26018
[14] http://bit.ly/global-dam-locations
[15] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Dams
[16] http://ewatercycle.nl
[17] http://globalfloods.eu

ods across multiple satellite sensors in an automated manner is still challenging. Hence, the introduction of parallel processing platforms such as GEE and freely available access to higher resolution data such as Sentinel-2 open new possibilities to improve these datasets and to reconstruct historical surface water dynamics. The actual number of currently mapped small reservoirs and lakes is still unknown, as well as their surface area extent and storage variability.

2.8. CONCLUSIONS

Many artifacts present in satellite images require the use of a multi-step approach when the actual signal needs to be filtered from the noise, usually present in almost every satellite image.

Remote Sensing (RS), in general, and EO in particular is a rapidly growing field. During the last decade, these fields have produced an enormous amount of new datasets, with many datasets freely available. The amount of free data available is more than can be digested by the research community in the short-term, so many multidisciplinary research questions, on both technological and algorithmic sides, remain open.

One of the largest challenges is related to the harmonization of satellite data, simplifying processing to extract valuable information from multiple satellite data products, resulting in a higher temporal resolution. Fortunately, Google has solved many of these technical issues by introducing the parallel satellite data processing platform Earth Engine.

Together with the remote sensing datasets, many other high-resolution Earth-scale datasets have become available. One of the fastest-growing datasets is OSM, but in general any other geo-referenced dataset can be applied. The main advantage of this type of dataset is the presence of semantic information (river names, administrative areas, property values and so on). This kind of information becomes crucial in performing higher-level studies, focusing on the impact assessment of short or long-term events, which can be observed from space. Integration, but also cross-validation of the datasets with different origins will be crucial in maximizing the value of these datasets and in improving their overall quality.

Technological developments and outreach efforts undertaken by large companies, have resulted in the development of parallel processing platforms like Google Earth Engine. This platform has truly revolutionized the processing of satellite imagery and has already led to a significant number of successful research efforts that would have been unimaginable to perform otherwise because of the large volumes of data that need to be stored and processed.

While many methods exist to process multispectral passive sensor satellite imagery, the methods listed in this chapter should be, in my opinion, the first to try when processing optical satellite imagery to address the task of surface water detection.

For surface water detection, the use of surface reflectance imagery will not necessarily result in better quality water masks, while the process of the generation of high-quality surface reflectance products is rather complicated.

With the growing volumes of data, the need for automated methods for surface water detection from satellite images is higher than ever. One of the primary goals of this research was to develop a set of fully automated algorithms and software tools to enable

processing of multi-spectral satellite imagery for surface water detection at high spatial and temporal resolutions. In the next three chapters 3, 4 and 5, more advanced unsupervised methods will be introduced to address this.

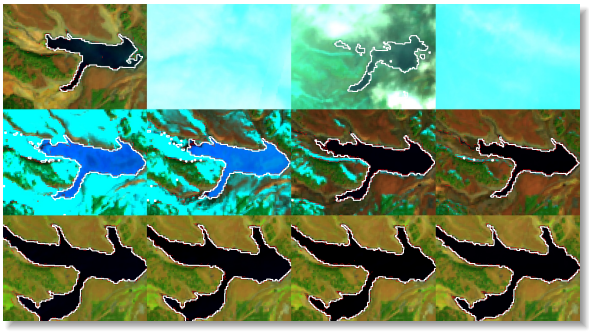
For long-term studies, as well as for permanent surface water detection (Chapter 7), the use of composite images may be sufficient, while for others, the highest temporal and spatial resolution data may be required. As an example, to study rapid surface water changes, such as floods, where the temporal resolution of physical processes is high and may require processing of all possible data sources to quantify the actual surface water dynamics, the highest temporal and spatial resolution data would be necessary.

Figure 2.13: Mapping of about 150000 global dam locations from multiple databases

3

SURFACE WATER DETECTION AND DYNAMIC LOCAL THRESHOLDING

The chapter discusses an automated surface water detection method, which can be applied globally and without the need for manual adjustments. A new method, based on dynamic local thresholding of spectral water indices, is presented. The method allows accurate detection of surface water from multispectral imagery even in the case of noisy data. The simplicity of the method makes it attractive for large scale applications and therefore can be used to address our research questions focusing on the accurate and high spatiotemporal resolution surface water detection. Applicability of the method for images with significant cloud and snow cover is tested. The method was applied to process both top of atmosphere (TOA) and reflectance percentile composite images. Numerous examples demonstrate the applicability of the method as well as its drawbacks.



Keywords: *M_1, NDWI, MNDWI, unsupervised classification, Otsu thresholding, Canny edge filter, snow, clouds, percentile composites.*

This chapter is based on G. Donchyts, J. Schellekens, H. Winsemius, E. Eisemann, and N. van de Giesen. A 30 m resolution surface water mask including estimation of positional and thematic differences using landsat 8, srtm and openstreetmap: A case study in the murray-darling basin, australia. *Remote Sensing*, 8(5):386, 2016b.

3.1. INTRODUCTION

AUTOMATED methods for surface water dynamics detection from multispectral satellite imagery were widely studied in the last two decades. Many studies were performed to improve the accuracy of water classification and to avoid the need for manual adjustments of thresholds. Early studies include the use of spectral indices and fixed thresholds, such as NDWI McFeeters [1996] and MNDWI Xu [2006]. Recently, many new spectral indices were introduced, such as AWEI Feyisa et al. [2014], WI_{2006} Homer et al. [2004], and WI_{2016} Fisher et al. [2016]. The main goal of these studies was to develop a single multi-band spectral index providing better separability of water from other land-use classes with the most stable threshold. While many of the studies have demonstrated improved accuracy of water classification for both clear sky and mixed pixels, their aim was mainly focused on the development of a new universal spectral index that can provide better separability between water and land. Furthermore, most of the method focused on atmospherically corrected imagery, and imagery with little to no noise.

Multiple studies provide a comprehensive literature overview of surface water detection from satellite imagery. Some of the works propose to categorize the methods of automated surface water detection, for example, Ji et al. [2009] has proposed the next categories: (*a*) thematic classification Lira [2006] (*b*) linear unmixing model Sethre et al. [2005] (*c*) single-band thresholding method Jain et al. [2005] (*d*) spectral water index method McFeeters [1996], Xu [2006], Feyisa et al. [2014], Hoberg et al. [2015], Fisher et al. [2016]. Later, Yang et al. [2015] grouped surface water detection methods using the following five categories: (*a*) digitizing through visual interpretation, a time and labor–consuming strategy which is unrealistic to repeat detection despite its high accuracy (*b*) density-slicing of a single band Frazier et al. [2000], Ryu et al. [2002], White and El Asmar [1999], which applies a fixed threshold to a given spectral band for water extraction (*c*) supervised or unsupervised classification (*d*) spectral water indexes McFeeters [1996], Xu [2006], Hoberg et al. [2015], Fisher et al. [2016], where a combination of two or more bands has proven to be effective as well as convenient (*e*) image processing methods, such as mathematical morphology and object-based analysis Blaschke [2010], Lira [2006], Yang et al. [2015].

Most of the classical methods of water detection mentioned above are based on the use of spectral water indices (*d*), or perhaps, a supervised binary classification, where images are classified into water and non-water classes. Even though recently developed methods provide better classification accuracy, they mostly require manual threshold adjustment to provide the best results. Therefore, their applicability to global studies, preserving high accuracy, is limited. In addition, many of the studies focus on methods that can be applied to process cloud-free satellite images or images, where the number of noisy pixels is very low.

In this and the next chapter, the focus will be on a new unsupervised method of water detection. The method was also tested with SAR backscatter imagery and provides good results for water detection. The method is fully automated and accurately predicts visible surface water, even in the presence of significant noise.

According to the classification of Yang et al. [2015], the method best fits the spectral water indexes and object-based analysis categories, (*d*) and (*e*). In the next chapter, the method will also be extended with a generative probabilistic framework to recon-

struct missing water pixels, for example, due to cloud, snow, or other noise. In the above classification, this extension falls under the category Machine Learning (ML) (*c*); and more specifically an unsupervised version of it, even though non-classical ML methods are used to implement the algorithms. Methods of ML (*c*) can be further separated into several smaller sub-categories, such as supervised ML, unsupervised ML, and reinforcement ML. A few more advanced methods were introduced recently, such as deep learning, generative adversarial networks Goodfellow et al. [2014]. Supervised ML is a classic example of ML and has been evolving for a long time. It solves a problem of classification or regression given a set of training examples. On the other hand, the definition of unsupervised ML is much vaguer. One of the definitions describes it as a task of inferring a function to describe hidden structure from unlabeled data. In general, the unsupervised ML solves tasks similar to those that are solved by supervised ML, but in this case, without any prior manual training required. Unsupervised ML does this by performing one of the following tasks: (*1*) clustering (*2*) dimensionality reduction (*3*) density estimation. The essential part of unsupervised ML is that all steps can be performed in an automated manner. In practice, some tuning of the algorithm parameters is still required from time to time.

Here, Otsu thresholding and Canny edge filters are used as building blocks for an automated water detection algorithm. The Canny edge filter is applied mainly to perform dimensionality reduction (*2*). This step is followed by the use of the Otsu thresholding step, resulting in an estimation of local threshold versus a global one. This results in a much more accurate water detection, mainly because the focus is shifted only to the most likely waterbody boundaries. In practice, a few additional steps may be required when working with different types of multispectral imagery, for example, removal of spurious edges based on the gradient of spectral index values, lengths of the edges, or topological properties. By combining image processing methods with the understanding of multispectral signals available in the satellite images, a fully automated step-wise algorithm for surface water detection is developed.

3.1.1. OTSU THRESHOLDING

In image processing and computer vision, Otsu's method Otsu [1975] is used to automatically cluster images using thresholding, or the reduction of a grayscale image into a binary image (in a more general form — into several classes). Otsu thresholding makes use of Fisher's linear discriminant analysis Fisher [1936] to compute the optimal threshold which discriminates the image into multiple classes in an optimal way. In a simple form, the method allows finding a threshold to classify the image into two classes, C_0 and C_1. This is done by analyzing the histogram of all values of the image, assuming that they represent a bivariate distribution.

In a simple binary form, the idea is to select a threshold that minimizes the intra-class (or within class) variance — the combined spread:

$$t = \underset{t}{\operatorname{argmin}} \left(\sigma_w^2(t)\right) \tag{3.1}$$

where $\sigma_w^2(t)$ is the within-class variance, defined as:

$$\sigma_w^2(t) = w_0^2(t)\sigma_0^2(t) + w_1^2(t)\sigma_1^2(t) \tag{3.2}$$

and $w_{0,1}$ and $\sigma_{0,1}^2$ are the probabilities and variances of two classes, separated by the threshold value at index t, where t is one of the indices of the histogram, defined by L levels $i \in \mathbb{Z}_L$.

In the case of a histogram with L levels, the probabilities $w_{0,1}$ are defined as:

$$w_0(t) = \sum_{i=0}^{t} p(i) \tag{3.3}$$

$$w_1(t) = \sum_{i=t+1}^{L-1} p(i) \tag{3.4}$$

And the means:

$$\mu_0(t) = \sum_{i=0}^{t} \frac{ip(i)}{w_0(t)} \tag{3.5}$$

$$\mu_1(t) = \sum_{i=t+1}^{L-1} \frac{ip(i)}{w_1(t)} \tag{3.6}$$

$$\mu = \sum_{i=0}^{L-1} ip(i) \tag{3.7}$$

where μ is the total mean.

Minimizing the above sum 3.1 for each possible threshold is expensive. Otsu shows that the same can be achieved by minimizing the intra-class (or within class) variance, rather than maximizing the inter-class (or between class) variance σ_b^2:

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = w_0\left(\mu_0(t) - \mu\right)^2 + w_1\left(\mu_1(t) - \mu\right)^2 \tag{3.8}$$

$$\sigma_b^2(t) = w_0(t)\,w_1(t)\left(\mu_0(t) - \mu_1(t)\right) \tag{3.9}$$

The optimal threshold is then obtained by maximizing the:

$$t = \underset{t}{\operatorname{argmax}}\left(\sigma_b^2(t)\right) \tag{3.10}$$

The resulting formula 3.9 is much easier to apply and can be computed by a single-pass algorithm applied to the image histogram (spectral index in our case). In this study, the algorithm is implemented using Google Earth Engine and JavaScript API so that it can be applied to a large number of images. Moreover, this method is frequently used in image processing libraries as one of the classical methods.

In our case, the red vertical line in the figures 3.2 and 3.9 indicates the optimal threshold found using this method.

3.1.2. CANNY EDGE DETECTOR

While global Otsu thresholding has been used successfully in many image processing applications, applying it to process satellite images with large spatial extent can be challenging. One of the limitations of the threshold being computed from all image pixels is that it includes all of the values present in the image, while we may be interested only in the pixels representing specific features, such as boundaries between water and land. To improve performance of the method, it may be easier to first reduce the image to include only pixels potentially representing a water and a land boundary before applying the Otsu thresholding. A popular algorithm used in image processing to detect edges is the Canny edge filter, which allows detection of a wide range of edges in images. The algorithm was developed in 1986 by Canny [1986]. The algorithm allows for the detection of edges with a low error rate, assuming that most of the edges are detected and that image noise is not marked as an edge. To satisfy these requirements, the algorithm is constructed from a number of steps: 1) images are smoothed using Gaussian kernel to remove the high-frequency noise 2) intensity gradients are computed 3) a non-maximum suppression step is applied to remove spurious edges 4) double thresholding is applied to determine potential edges 5) edges are tracked by hysteresis: edges that are weak and not connected to strong edges are suppressed.

During the first step, the image is smoothed, usually by convolving with the Gaussian kernel of some fixed size. After that, the gradient and aspect are computed as:

$$G = \sqrt{G_x^2 + G_y^2} \tag{3.11}$$

$$\theta = atan2\left(G_y, G_x\right) \tag{3.12}$$

The aspect θ is then rounded to one of the discrete angles representing vertical, horizontal and diagonal lines (0°, 45°, 90°, 135°).

During the non-maximum suppression step, the edges get thinned followed by thresholding and edge tracking, ensuring that only the strongest edges remain.

In our study, the Canny edge filter is combined with the Otsu thresholding method to compute local threshold for spectral indices in a dynamic way. The algorithm was coded using its JavaScript API. Some of the figures were generated using Python and the implementation of algorithms from the scikit-image library Van der Walt et al. [2014]. The main parameters used to tune the algorithm were σ and size of the Gaussian kernel and the threshold used to remove edges. Both parameters were found empirically to ensure detection of most of the water features present in the satellite images.

3.2. DYNAMIC LOCAL THRESHOLDING OF SPECTRAL WATER INDICES

Varying thresholds of spectral water indices may result in significant errors in resulting surface water masks Yang et al. [2014]. Manual adjustment of the threshold is usually required to overcome these errors, which become even more evident for multispectral images where significant atmospheric noise is present. Also, spectral properties of open water, as well as the land type next to the water may vary significantly across the globe.

In many cases, surface water constitutes only a small fraction of the overall land cover, making it harder to detect with threshold based methods. Large local errors may

be introduced when a constant threshold of 0.0 is used to distinguish between water and land. The challenge is to establish a varying threshold that can be derived automatically.

The fact that water is almost never clear in the real world may result in significant distortions in the observed spectral properties, limiting the applicability of spectral indices, especially when a single fixed threshold is used to separate water pixels from non-water pixels. Typical variations of threshold values for different spectral water indices can be found in Ji et al. [2009]. One of the approaches to overcome this problem is to use methods that allow detecting threshold values based on a histogram of all NDWI values in a given area. One such method is Otsu thresholding Li et al. [2013], Yang et al. [2014]. In fact, the Otsu method is very similar to the k-means method applied to the histogram of spectral index values Liu and Yu [2009].

The use of dynamic threshold detection methods, such as histogram-based Otsu thresholding, helps to overcome some of the problems. However, this approach does not work when the fraction of water pixels is small. The main reason is that a number of land-use types can be significant, resulting in that water pixels becoming practically invisible, making them impossible to detect. A more detailed example can be found in Chapter 7, where surface water of the Murray-Darling River Basin, Australia is estimated from a multitemporal Landsat 8 dataset.

To handle this problem, two methods are applied sequentially (Canny edge filter and Otsu thresholding) to spectral water index images. This allows detection of a local threshold versus the global one. The application of the Canny edge filter with a very high threshold applied to the spectral index image reveals only edges located near sharp value changes. In the case of water spectral indices, this usually takes place when the near-infrared band abruptly changes.

Potential water and land pixels located near water are then computed using morphological dilation applied to the detected edges. It is important to note, that this approach can result in a skewed distribution in the case of thin, single pixel wide water bodies (canals). To overcome this problem, a buffer with the size (dilation) equal to half of the pixel is used in step 3. In an ideal situation, the resulting distribution should look bimodal 3.1 so that a clear distinction between land and water can be made from this distribution.

The proposed method consists of the following steps: 1) compute spectral water index, 2) compute edges using the Canny edge detector 3) dilate edges to capture most probable water and land pixels located around waterbody edges 4) compute a threshold value using the Otsu method, using only the pixels in the buffer 5) compute water mask by applying the threshold.

As can be seen in Figure 3.2, the proposed method can capture smaller local features better when compared to the use of the default thresholds.

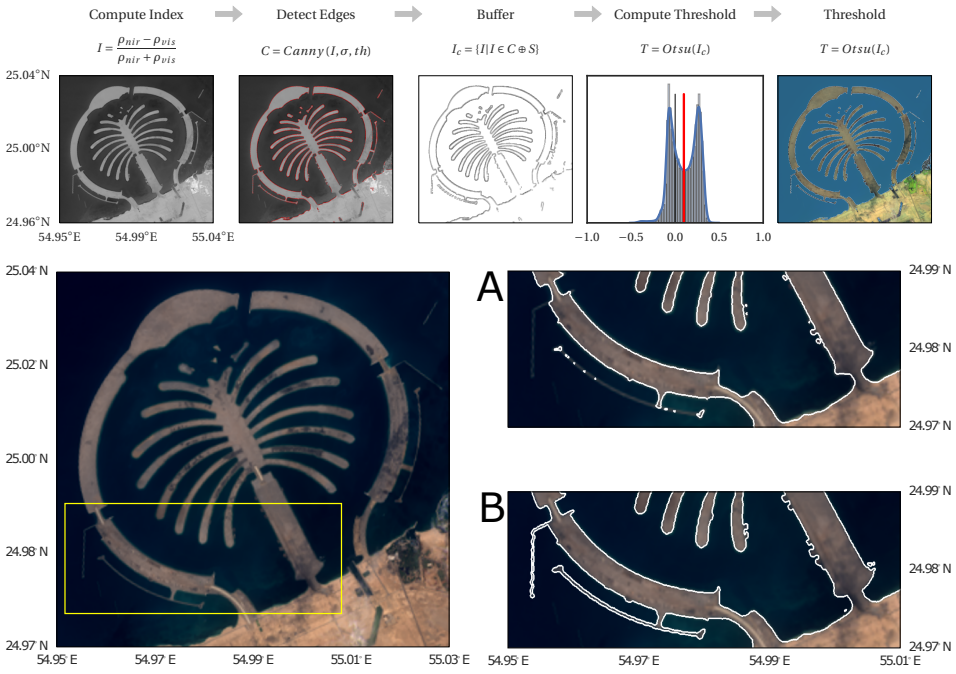


Figure 3.1: The method of dynamic local thresholding for water detection, Palm Jebel Ali, Dubai, UAE. Surface water detected using new method (B) reveals more details, when compared to a naive NDWI=0 method (A).

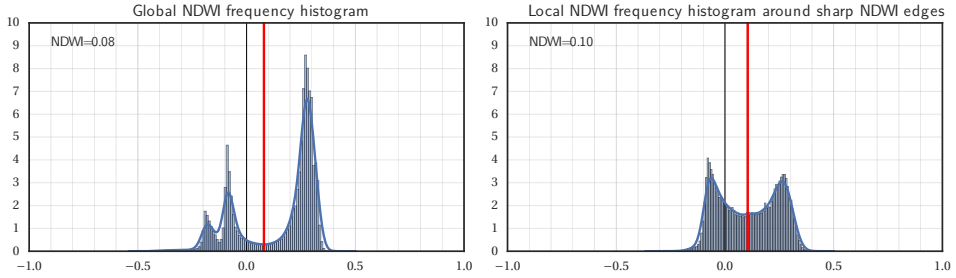


Figure 3.2: Frequency histogram of NDWI values without (left) and with (right) the use of Canny edge detection filter, $\sigma = 1$, $th = 0.5$. The red line shows the threshold estimate using the Otsu method.

$$I = \frac{\rho_{green} - \rho_{swir1}}{\rho_{green} + \rho_{swir1}} \tag{3.13}$$

$$C = Canny(I, \sigma, th) \tag{3.14}$$

$$I_c = \{I | I \in C \oplus S\} \tag{3.15}$$

$$T = Otsu(I_c) \tag{3.16}$$

$$Water = \{I | I < T\} \tag{3.17}$$

where $Canny$ - is a Canny edge detection filter, applied to the spectral water index I and resulting in a masked image of edges C (3.1 red edges), in this example - MNDWI; σ and th are the standard deviation and the threshold used by the Canny edge filter; S is a fixed size dilation kernel,

3.3. STEPWISE METHOD OF SURFACE WATER DETECTION FOR RESERVOIRS

To validate the applicability of the above method and to test its performance, the method was applied to reconstruct surface area changes for a relatively small reservoir with a maximum surface water area of about 300ha and capacity of about 36 800 000 m3.

3.3.1. PROSSER CREEK RESERVOIR

The Prosser Creek Reservoir is located in Nevada County, California, USA and was constructed between 1959 and 1962 by the Unites States Bureau of Reclamation at an altitude of 1750m above the sea level. The dam crest height is 50m, and the reservoir width is about 0.5km. It impounds Prosser Creek and is used for both irrigation and flood control during winter and spring.

The reservoir was selected for validation because high-frequency surface water level data are available from USGS (`http://nwis.waterdata.usgs.gov/nwi`).

Figure 3.3: Prosser Creek Reservoir, Nevada County, California, USA. Image: Bureau of Reclamation.

Daily water level data are used, measured at the dam during 1996-2016 (Figure 3.4). As can be seen, the reservoir water level follows a typical seasonal pattern, with almost constant water levels during the winter season. However, during some years, the reservoir was only partially filled.

To test the applicability of the surface water detection method outlined above, all freely available medium resolution (<60m) satellite images for this area were used, measured by multiple multispectral (Landsat 4-8 and Sentinel-2) and SAR (Sentinel-1) sensors. In total, 1297 images were collected using different sensors.

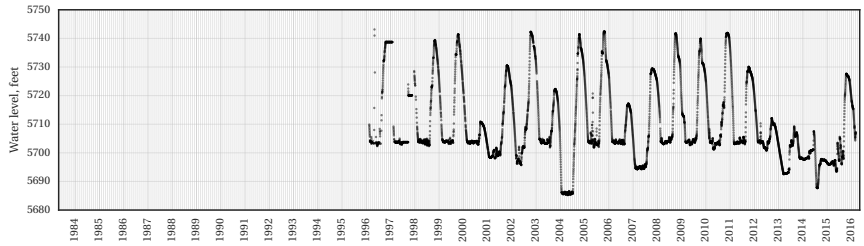


Figure 3.4: Measured water levels, source: USGS

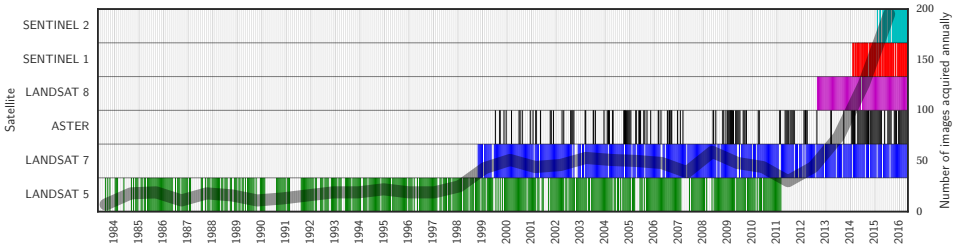


Figure 3.5: Number of images acquired by different satellite missions analyzed

The reservoir surface area is on average 41% covered by clouds Wilson and Jetz [2016], freezes during the winter season, and varies significantly throughout the years and seasons, resulting in many images being fully or partially covered by clouds, cloud cover or snow/ice.

The above method is applied to process TOA images, without the use of cloud or snow masking, to see how good or bad different surface water detection methods will perform. Subsequently, other processing steps were added, such as local Otsu thresholding, mentioned above, and topographic masking, using HAND.

In the figures 3.6 and 3.7 a few typical Landsat images for our study area are presented, showing the variability of reflectance values in visible and infrared parts of the spectrum. Twelve randomly selected images are used. Some of these images are especially difficult to process using most of the existing methods, while others are cloud-free. Cloud cover forms the main type of noise present in most of these images. Some of the images show significant snow cover presence, as well as ice.

As expected, surface water is slightly more visible in false-color composite images where short-wave infrared and near-infrared bands are used. This happens because infrared radiation penetrates thin clouds better than visible light.

On several images, the effects of hill shadows, which can cause false-positive surface water detection, are also present. The image acquired on 2015-10-08 also has a slight amount of algae present near the dam, and the image acquired on 2014-02-23 is partially covered by ice. We will see later in this chapter that these pixels will result in false-negative surface water detection, and in general, can't be corrected using the simple discriminative methods discussed in this chapter. More advanced methods to correct these

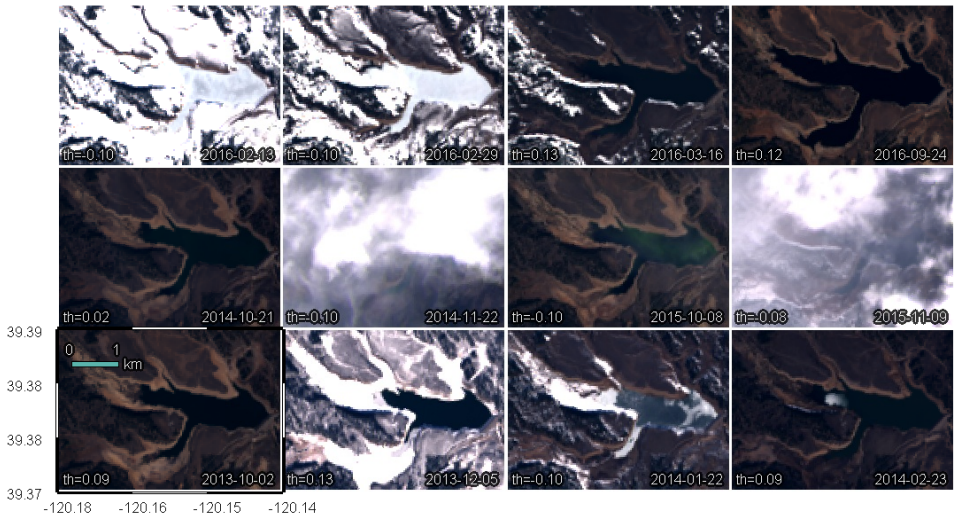


Figure 3.6: True-color image for Prosser Creek Reservoir, CA, USA. At-sensor reflectance ($\rho_{red}, \rho_{green}, \rho_{blue}$)

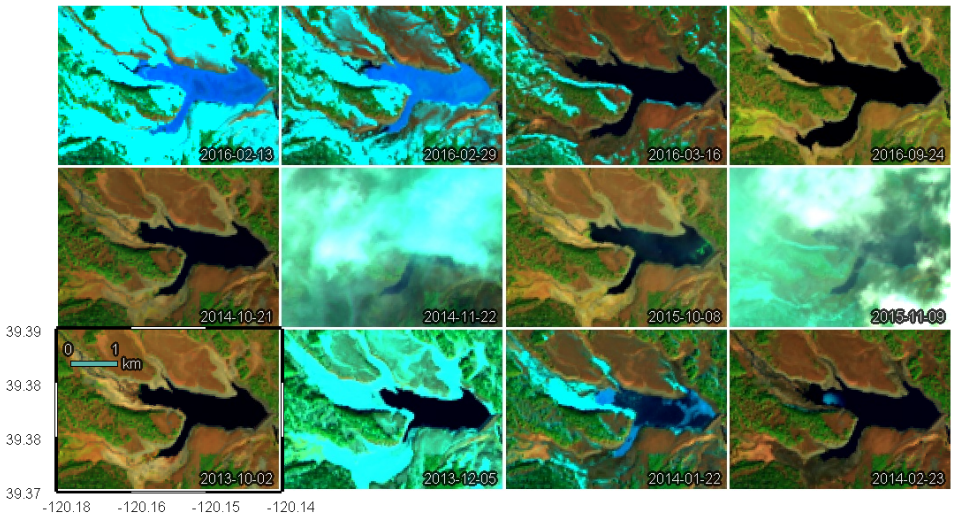


Figure 3.7: False-color image for Prosser Creek Reservoir, CA, USA. At-sensor reflectance ($\rho_{swir1}, \rho_{nir}, \rho_{green}$)

errors will be discussed in the next chapter, where probabilistic methods will be applied in a combination with local Otsu thresholding to estimate surface water.

3.3.2. APPLICATION OF THE DYNAMIC THRESHOLDING METHOD FOR WATER DETECTION

Applying the local dynamic thresholding method, introduced above, to the cloud-free images where little to no effects from hill shadows were present allowed for the removal of falsely detected water, and resulted in a slightly better water mask, Figure 3.8. Similar to the previous example for Palm Jebel Ali (3.1), we can see from the histogram in 3.9 that the objective selection of the threshold using global Otsu would be difficult. The reason is that the bimodal histogram on the left is skewed, and a very long flat valley makes it difficult to decide which threshold value would be the most optimal. At the same time, the histogram on the right represents NDWI values around strong edges in the NDWI image; the optimal threshold selection becomes easier to compute. However, we can also see that in this case, a much larger variability of NDWI values is present, due to different land use types. Even though these disturbances are small, in some images they may be significant, for example, when very distinctive land cover types are present around the waterbody (sand, vegetation, man-made constructions).

In some cases, when local NDWI values result in a multi-modal histogram, the use of the multi-class Otsu method would be more appropriate. As an alternative, smaller overlapping areas may be used to limit land cover variability. Later, in Chapter 7 this approach will be applied to estimate permanent surface water from Landsat 8 reflectance composite images for Murray-Darling River Basin in Australia.

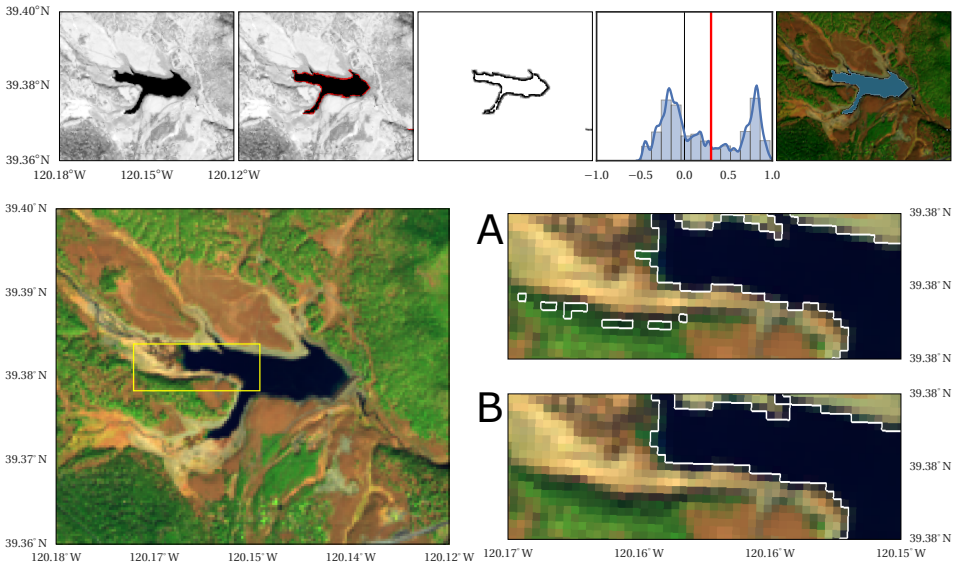


Figure 3.8: Method of dynamic local thresholding for water detection (Prosser Creek Reservoir). Dynamic method of surface water detection (B) provides lower commission error due to hill shadows, when compared to the naive NDWI=0 method (A).

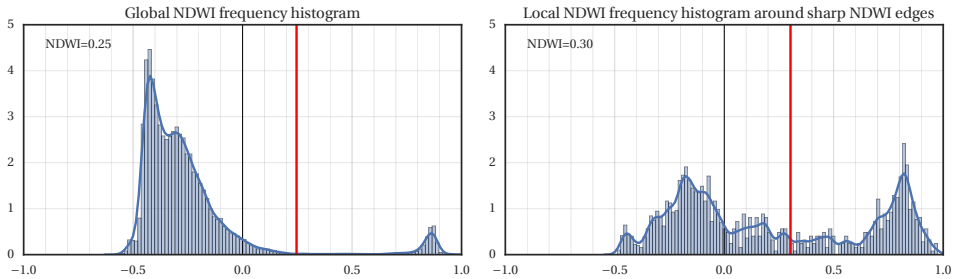


Figure 3.9: Frequency histogram of NDWI values without (left) and with (right) the use of local dynamic thresholding.

3.4. VARIABILITY OF NDWI AND MNDWI FOR NOISY IMAGES

A more detailed analysis of the values of NDWI (Figure 3.10) and MNDWI (3.11) spectral indices shows, that MNDWI may be a preferable index for some of the cloud and snow-free images, providing a more sensitive way to discriminate water and land. However, the index is, in general, much more sensitive to snow and cold clouds, making water detection very challenging under these conditions. On the other hand, the MNDWI index was reported to perform better in urban areas. In addition to snow, the MNDWI index is also more sensitive to hill and cloud shadows when compared to NDWI.

As can be seen from the figures, under cloud-free conditions, both indices can be used to discriminate surface water from the land. In addition, when working with atmospherically corrected images, a threshold value of zero is frequently a good starting point for water detection. However, when a very accurate surface water mask needs to be detected or when a high concentration of cloud (fog, haze) cover is present, the use of a dynamic threshold is the only option.

In the next figures, where the detected surface water mask will be shown, the NDWI will be used as the basis for water mask detection, combined with additional steps to detect surface water.

3.5. RECONSTRUCTION OF SURFACE WATER AREA FROM NOISY IMAGES

To test the method, it was sequentially applied to a set of randomly selected images with different atmospheric and land-use conditions as shown in figures 3.6 and 3.7. For most cloud-free images, as well as for images covered by clouds, the algorithm was able to detect surface water very accurately, without visible errors. However, for most images where a mix of snow/ice and cloud cover was present or where the waterbody was partially covered by snow (Figure 3.12), the results were unreliable, mainly due to the presence of snow pixels, which generated spurious edges. It is important to note that no snow or cloud masking was applied before the water mask detection algorithm was applied. In practical applications, most of the images where cloud or snow cover is present would be filtered as unreliable.

The performance of surface water detection for noisy images was improved dramatically after the additional topographic mask was applied to the Canny edges using HAND < 15m, which resulted in the removal of most of the spurious edges. The final water mask can be seen in Figure 3.13.

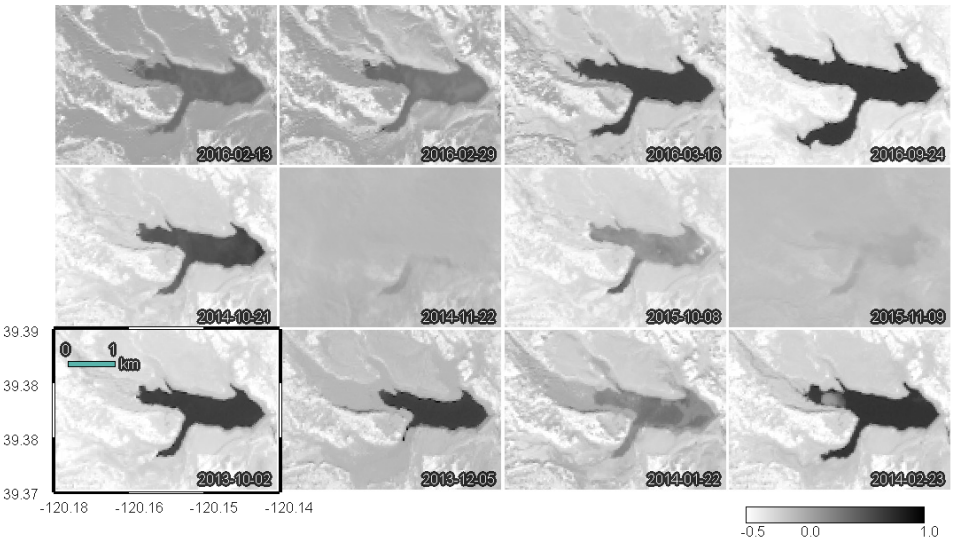


Figure 3.10: NDWI values for Prosser Creek Reservoir for selected dates, McFeeters [1996]. Source: `https://code.earthengine.google.com/59461a0aae48edf382670062fae1ea3a`

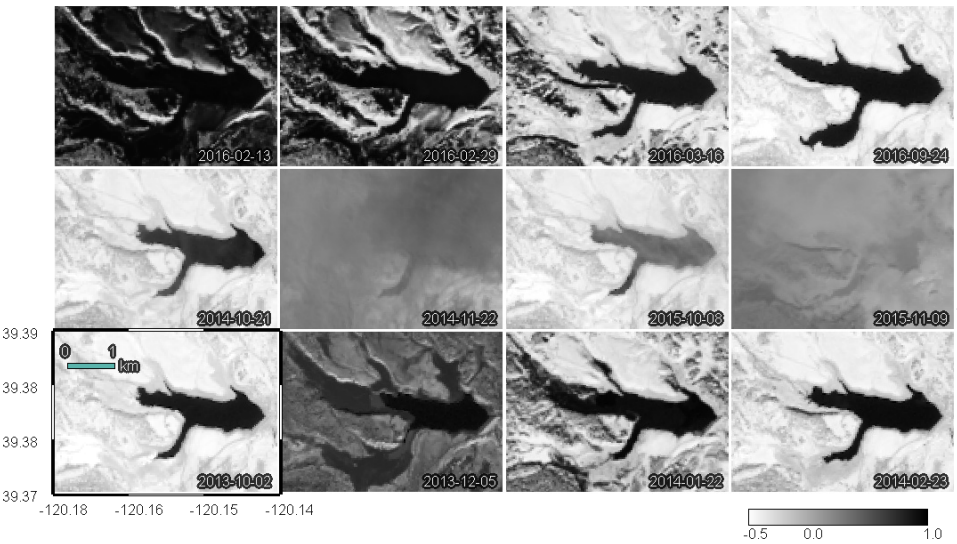


Figure 3.11: MNDWI values for Prosser Creek Reservoir for selected dates, Xu [2006]

Here we can see that the algorithm was able to detect water masks for all images, covered by clouds, snow/ice, or a combination of both. However, for practical use in water resource management applications, further additional processing is required to convert these partial water masks to the actual surface water area values.

Another approach to detect surface water reliably for images partially covered by snow and clouds, would be to take snow and cloud mask into consideration when filtering out edges during the edge detection step. However, a very reliable detection of surface water may be challenging when multiple types of noise are present at the same time.

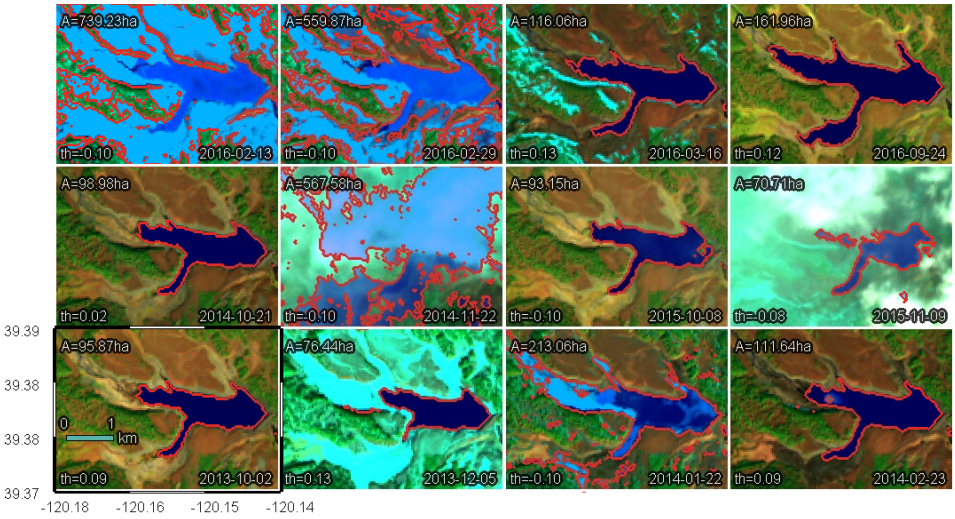


Figure 3.12: Canny/Otsu

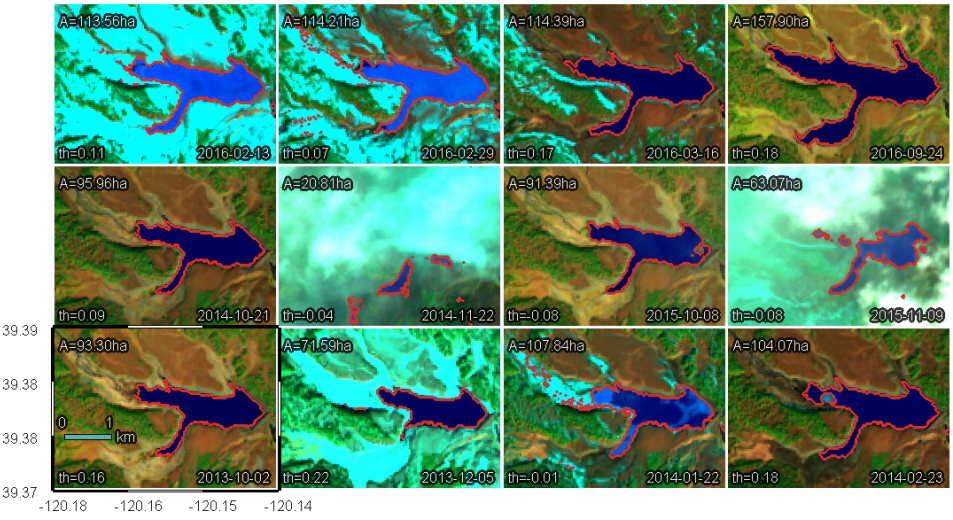


Figure 3.13: Canny/Otsu, HAND < 15m

3.6. VALIDATION OF THE SURFACE WATER DETECTION METHOD

To validate the surface water detection method outlined here, and, to develop a simple statistical model to be used for the validation in the next chapter, mostly cloud-free images were selected covering the reservoir area. The resulting surface water area values were compared to the in-situ observed water levels as shown in figures 3.14 and 3.15. Additionally, the outliers were identified, indicated in red in the figure. It appeared that

for some of the values, the water level station values (during 1997-1998) were unreliable, probably because of sensor problems. For two points, our criteria for the maximum allowed snow/ice and cloud mask (<10 pixels) appear to be incorrect, resulting in higher values for surface water.

It can be seen from the second chart that even when using only the cloud-free images most of the peaks of the reservoir water level time series can be detected. For 2011-2013, the only available images were measured by Landsat 7, with the SLC-OFF ETM+ sensor problems, resulting in gaps in the satellite images.

3.6.1. MODEL CONSTRUCTION USING CLOUD-FREE SATELLITE IMAGES

After outlier filtering, a second-order polynomial model can be derived from linear regression, resulting in correction of a 0.994 and $RMSE = 4.7ha$ or $0.047km^2$.

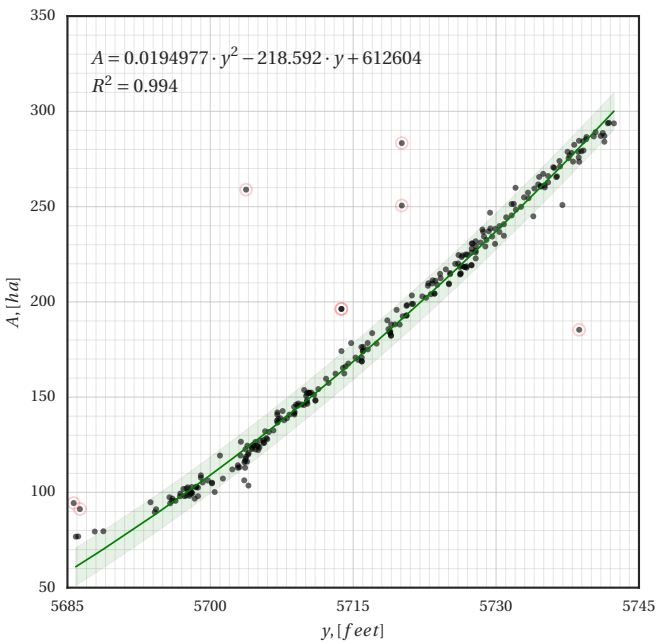


Figure 3.14: Prosser Creek Reservoir surface area and water levels for cloud-free images using multiple optical sensors (Landsat, ASTER, Sentinel-2)

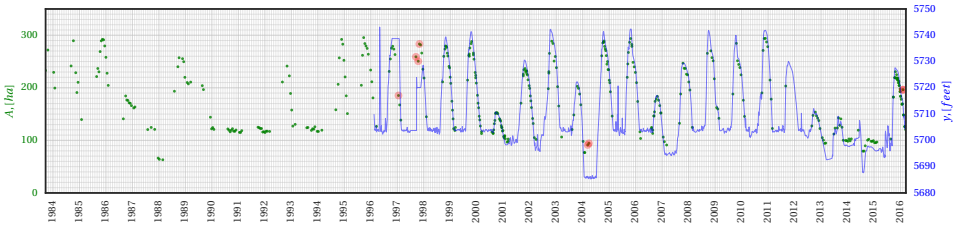


Figure 3.15: Prosser Creek Reservoir surface area and water levels for cloud-free images using multiple optical sensors (Landsat, ASTER, Sentinel-2)

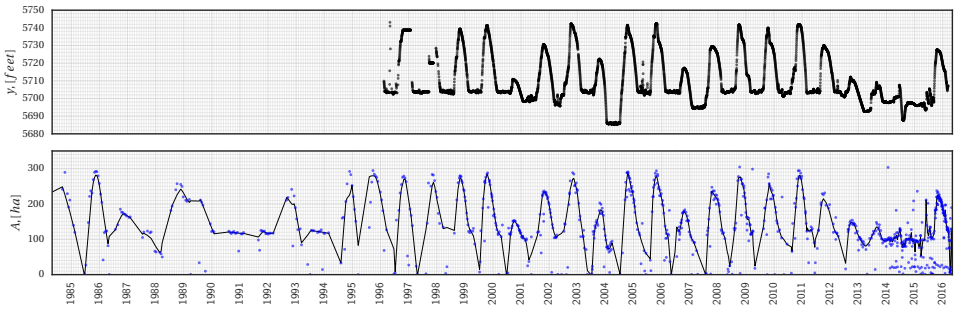


Figure 3.16: Measured surface water levels (top) and estimated surface water area form all images acquired during 1984-2016. The line in the lower chart is based on the Locally Weighted Scatterplot Smoothing (LOWESS) regression.

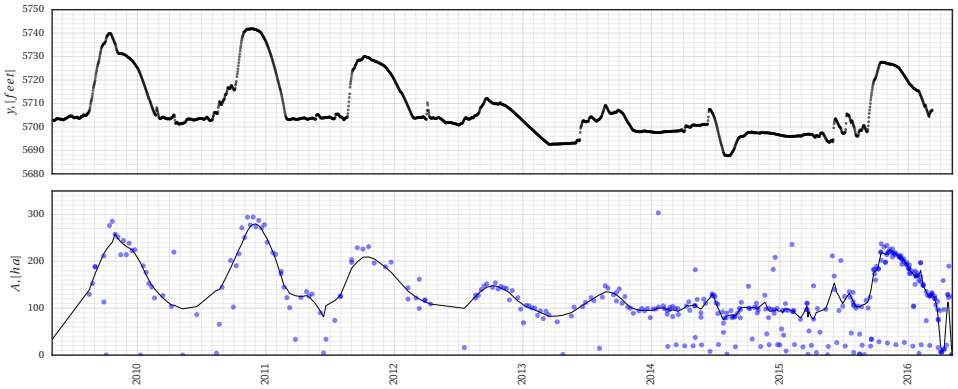


Figure 3.17: Measured surface water levels (top) and estimated surface water area form all images acquired during 2009-2016. The line in the lower chart is based on the LOWESS regression.

Finally, the method was applied to process the rest of the (noisy) images, where the reservoir is (partially) covered by clouds or cloud cover, resulting in significantly noised spectral signals measured at the sensor. However, the resulting surface water area values

were less reliable. The resulting surface water area values are shown in figures 3.16 and 3.17, combined with the non-parametric LOWESS regression Cleveland [1979].

The larger variability of surface water values after 2014 is caused by the estimates from ESA Sentinel-1A images, where only a very small portion of the reservoir can be seen in the image.

3.7. ON THE FUSION OF WATER MASKS ESTIMATED FROM MULTISPECTRAL AND SAR SENSORS

One of the advantages of the use of Otsu-based methods is that it is much less sensitive to radiometric differences between satellite sensors and may be used across images coming from different satellites. As can be seen in 3.5, the number of overlapping satellite missions has been increasing in recent years, resulting in daily, or more frequent, image availability over our study area. The main challenge of fully automating water detection, in this case, is still to correct evaluation of the confidence of the image pixels, to exclude, or to detect, the presence of clouds, cloud shadows, and snow. Developing fully automated software that may work across different sensors and also be applied to millions of locations becomes feasible. The Figure 3.18 demonstrates the results of water detection applied across all of the medium resolution sensors (10m-30m) acquired in 2016 during a period or high water level.

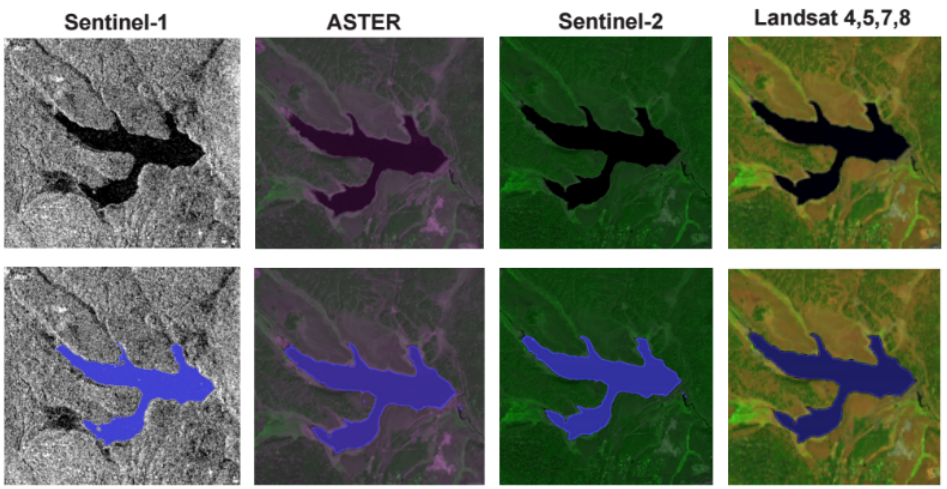


Figure 3.18: Surface water mask reconstructed from multiple medium-resolution sensors (Sentinel-1, ASTER, Sentinel-2, Landsat)

In fact, for the ASTER sensor, it was possible to successfully detect surface water even using raw uncalibrated, but normalized radiance values, represented as DN numbers, without converting them to radiance or reflectance values. However, reliable detection of clouds and snow was harder to perform due to the variability of sun parameters. Therefore, for the final version of the algorithm, the full version of the DN > radiance > reflectance algorithm was implemented within Google Earth Engine using equations 2.1

and 2.2.

3.7.1. SURFACE WATER DETECTION SAR AND SPECKLE NOISE REMOVAL

Not surprisingly, the algorithm has also shown good results when applied to backscatter amplitude SAR images acquired by the Sentinel-1 ESA satellite (3.18, left). However, in some images, it was very difficult to distinguish between spurious edges caused by Bragg scattering (Liebe et al. [2009]). Surface water masks estimated from these images were excluded for the construction of the regression model shown in 3.15.

The best results using SAR backscatter amplitude images were achieved after removal of speckle noise. Many methods exist to remove speckle noise in SAR images, with some of the most popular being Gamma Maximum a Posteriori (MAP) Lopes et al. [1990] or a refined Lee filter Lee [1981]. At the time when this study was performed, no implementation of a speckle filter existed within Google Earth Engine.

To address this, a new implementation was developed based on the method introduced by Perona and Malik Perona and Malik [1990]. The filter is in fact based on a simple anisotropic diffusion equation, where a diffusion coefficient is parametrized using values from the neighboring pixels.

$$\frac{\partial I}{\partial t} = \nabla \cdot (c \nabla I) \tag{3.18}$$

where $I = I(x, y, t)$ denotes in this case intensity values of the backscatter signal, ∇ is a gradient operator, and $c = c(x, y, t)$ is an anisotropic diffusion coefficient, parametrized using the nearest pixel values as:

$$c = \frac{1}{1 + (G/K)^2} \tag{3.19}$$

where $G = \|I\|$ is the absolute value of the gradient computed for the current image pixel, and K is the user-defined coefficient.

In fact, the original paper defines two alternative version of the parametrization for c, resulting in similar results.

The implementation (see Listing A.2) is based on the simple explicit central-differences numerical scheme, implemented using the convolution operator based on a set of 3x3 kernels.

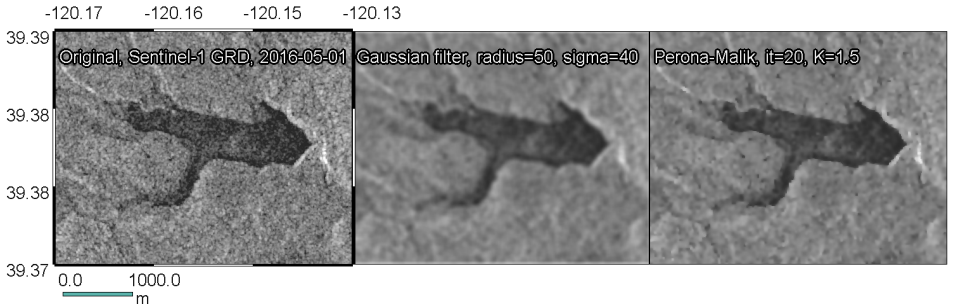


Figure 3.19: An example of results of Perona-Malik implementation within Google Earth Engine, compared to original and Gaussian-smoothed images. Source: https://code.earthengine.google.com/008a8e627389123fd61550929973463a

As can be seen in Figure 3.19, the image's high-frequency noise is significantly decreased after applying the filter, while strong edges remain. In contrast, applying simple Gaussian smoothing results in the removal of strong edges, reducing the accuracy of surface water detection.

3.8. CONCLUSIONS AND DISCUSSION

A new method of dynamic local thresholding was introduced for surface water detection from multispectral optical and SAR satellite imagery. The method is easy to implement and is based on two popular image processing techniques: Canny edge filter and Otsu thresholding.

One of the advantages of the algorithm is that it can be easily applied to process any multispectral satellite imagery measured by any satellite mission, regardless of its spatial and spectral resolution. This fact makes it attractive for global applications, where a large quantity of satellite data needs to be processed without user supervision.

All freely available medium resolution satellite images measured by NASA and ESA within the Landsat, ASTER, and Sentinel missions were analyzed, and the resulting surface water masks were compared to daily water level measurements for the period 1996-2017.

The method was applied to reconstruct surface area changes at Prosser Creek Reservoir in California, USA. The reconstruction strongly matched the in-situ observation data for cloud-free images. With cloud-free images, it was possible to achieve a perfect fit with the in-situ observation data. However, much more variability was observed when using images with a significant amount of clouds or snow cover present.

In addition to passive optical satellite imagery, the method was also adjusted to process Sentinel-1 SAR imagery. In contrast to multispectral imagery, where NDWI was used to detect surface water, for SAR imagery, the backscatter amplitude was used to detect surface water. To diminish the effects of speckle noise in SAR imagery, a new speckle filter was implemented within the Google Earth Engine, based on Perona-Malik, to improve the quality of images and to decrease the number of spurious sharp edges not belonging to surface water.

While the method of dynamic local thresholding presented in this chapter provides

a way to very accurately estimate surface water masks, it is still insufficient to reliably estimate surface water masks when surface water is only partially visible. In fact, the accuracy of the resulting water masks varies a lot depending on the type of noise present in the image. The method shows very good performance in the case of homogeneously distributed noise, such as fog or haze. However, for water bodies that are partially covered by thick clouds or snow/ice, the results are less accurate, causing significant variability of the final surface water area estimates.

In the next chapter, we will see how the resulting water mask can be further improved using Bayesian methods. We will see how cloud-free historical images can be processed using generative methods, by deriving a density function from cloud-free satellite images, followed by the reconstruction of the surface water area from partially-visible satellite images.

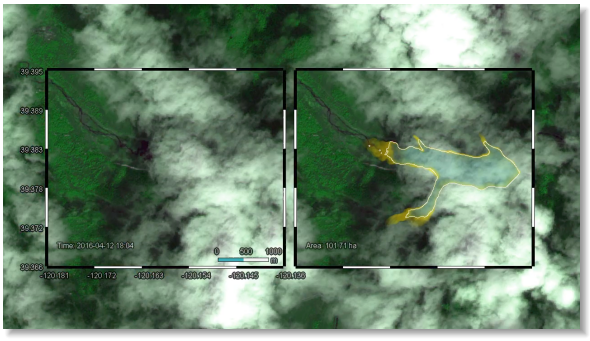
Eventhough the method introduced in this chapter and further extended in the next chapter has been validated for a single reservoir only (for demonstration purposes), it was successfully tested to reconstruct permanent surface water area for the Murray-Darling River Basin in Australia, where the resulting water mask was compared to other water datasets for a very large area. Also, early results show, that the method can be easily transferred globally, given that the cloud and snow mask is estimated correctly (low commission error). The next chapter discusses how these masks can be detected automatically using the current methods and datasets.

4

PROBABILISTIC METHODS OF SURFACE WATER DETECTION

A method has been developed to estimate surface water mask for noisy TOA reflectance images, where water bodies can only be partially visible under thin clouds, haze, fog or snow/ice cover. The presence of these effects increases the need for high accuracy and more frequent observations of surface water bodies, making it essential for rapidly varying surface water bodies such as small reservoirs, wetlands, and floodplains. Monitoring these waterbodies at high frequency is essential to understanding changes in the availability of water for food, conditions of fisheries, ecosystems, and the occurrence of floods.

Here, a Bayesian Network is used as a framework to introduce the method and, finally, to infer the actual surface water area from partially-visible surface water masks. The method makes use of a high-resolution probability density function, which is estimated using a set of cloud-free satellite images from measured by multiple freely available satellite missions from NASA and ESA. Variability of TOA reflectance values for different land use and atmospheric conditions to estimate cloud and snow detection parameters in an automated manner is explored. The method is validated using in-situ water level measurements.



Keywords: *M_2, probabilistic graphical models, filling, Bayesian Networks, KDE, inference.*

4.1. INTRODUCTION

In the previous chapters, we saw various methods to detect surface water at high accuracy. However, when a waterbody is partially invisible, for example, due to clouds or missing pixels, these methods are insufficient to estimate the actual surface water. In this case, more advanced methods are required to fill missing pixels. In this chapter, this problem is addressed by making use of probabilistic graphical models, in particular, by formulating the task of inferring the most probable water mask in the form of a simple Bayesian Network.

The general idea is to combine methods discussed in the previous chapter with the probabilistic estimate of surface water mask using a priori knowledge about surface water. This will be done by generating a joint probability density function using surface water estimates based on images where an accurate surface water mask can be generated.

This turns our task of detecting surface water from satellite data into a generative class of models, where the most probable surface water mask is generated using density functions, in contrast to the discriminative way of processing of satellite data.

Additionally, a Probabilistic Graphical Model (PGM) notation is used to discuss the steps required to estimate (or infer) the actual surface water mask from available raw satellite data.

4.2. PROBABILISTIC GRAPHICAL MODELS

PGMs (Koller and Friedman [2009]) is a rich framework for encoding probability distributions over complex domains, in the form of univariate or multivariate (joint) distributions. Random variables and their conditional dependencies are then expressed in the form of a graph, where child nodes are usually represented by Conditional Probability Distributions (CPDs), either defined analytically or in a tabular form, and root nodes are represented by marginal distributions. Two commonly used types of PGMs are Bayesian Networks (BNs) and Markov Random Fields (MRFs). However, many other types of PGMs exists today, such as Conditional Random Fields (CRFs), factor graphs, restricted Boltzmann machines and many others.

Here, a BN notation is used, which is usually expressed in the form of a Directed Acyclic Graph (DAG). Consider the example of a simple network as shown in 4.1. Here, $X_{1...3}$ nodes represent random variables and the shaded variable X_2 indicates a variable that is measured.

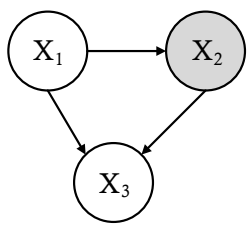


Figure 4.1: A simple Bayesian network.

The root nodes of the graph of a BN are defined as marginal distributions, which are in general expressed as single-variable Probability Distribution Functionss (PDFs), and all child nodes can then be represented as CPDs, which can be expressed as multivariate probability distribution functions.

This DAG can then be used to conduct statistical inference, for example, to infer the unknown variables X_3 and X_1 by making assumptions about their distributions, while at the same

time, satisfying the available (measured) distribution of X_2. The joint distribution of this BN can then be expressed (factorized) as:

$$P(X_1, X_2, X_3) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1, X_2) \tag{4.1}$$

or in a general form:

$$P(X_i, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid S_i) \tag{4.2}$$

where S_i represents a set of parents of every node X_i.

After defining the model, many types of questions can be answered, by means of direct inference methods, such as conditioning parent variables and computing resulting distributions of dependent variables, marginalization — deriving marginal distribution by integrating over other variables, or by applying methods of learning to simulate the graph as a whole and to infer hidden variables.

Many methods exist to perform inference in BNs, such as Markov Chain Monte Carlo (MCMC), or Expectation Minimization (EM). The main goal of these methods is to estimate unknown distributions from known variables and hypothetical relations between them. Many examples of PGMs as well as methods used to conduct inference in BNs are given in Koller and Friedman [2009].

4.3. PLATE NOTATION

One of the useful notations used for BNs is a *plate notation*, which allows grouping of multiple random variables of a similar type. In this way, more complex models can be expressed by the network, as shown in Figure 4.2.

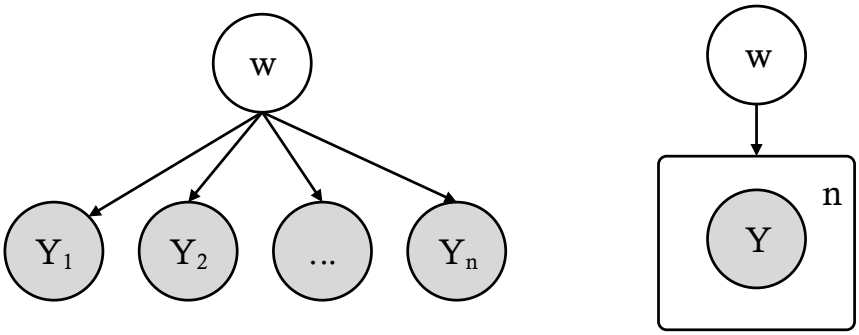


Figure 4.2: An example of the use of plate notation (right) for Naive Bayes classifier graphical model (left).

The above graph, in fact, represents a model of a Naive Bayes classifier, where the random variable w represents a distribution of some feature belonging to some class, given a set of independent variables X_i with known conditional distributions $P(X_i|w)$.

During training, the model variables X_i (conditional distributions) are defined. During inference, the most probable class is selected, for example, by estimating the maximum a posteriori probability.

4.4. DISCRIMINATIVE VS. GENERATIVE METHODS

In machine learning, a generative model is defined as a model where some variable values are generated from a joint probability density function. Generative models differ from discriminative models in that they require the use of joint probability density function to be constructed, while discriminative models provide a way to model only the target variables conditioned on observed variables Ng and Jordan [2002]. Some examples of generative models include Naive Bayes classifier, Gaussian Mixture Model, and the Hidden Markov Model while examples of discriminative models include Logistic Regression, Support Vector Machines, and Conditional Random Fields.

The method presented in this chapter can be classified as a generative one, as the joint probability density function is estimated for pixels belonging to water $P(x, y)$. However, it will be estimated in a discriminative way, by combining surface water masks detected from cloud-free images, as was discussed in the previous chapter.

4.5. PROBABILISTIC NATURE OF THE SATELLITE OBSERVATIONS

Here, the multispectral satellite observations are expressed in the form of a Bayesian network, to allow better reasoning about the methods used to detect surface water changes. Eventhough, standard algorithms used to reconstruct marginal distributions were not applied, the framework is a very rich way to express and explain probabilistic surface water inference methods introduced later in this chapter.

In multispectral remote sensing using satellites, we can think of observed radiance or reflectance values as random variables. In fact, the observed values represent a joint distribution in a multidimensional space (time, space, wavelength). Even though most of the sensors provide 2D surface radiance values, the actual values are reflected from the 3D surface of the Earth. The actual values measured at-the-sensor represent a signal, backscattered from the aerosol particles, and from the ground (frequently represented by non-Lambertian surfaces). The actual path of the light is extremely complex and may be impossible to model in a deterministic way. Interpreting this signal using probabilistic models makes more sense.

For passive multispectral satellite sensors, the most important factors disturbing the resulting radiance values are clouds and topographic effects. While many algorithms exist to detect clouds, modeling complex interaction of cloud and hill shadows, thin clouds, fog or haze is extremely difficult. Additionally, physical parameters of the sensors, such as spatial and radiometric resolution, result in observations of mixed signals, reflected from multiple surfaces, such as different types of soil, water, and vegetation. Furthermore, surface water pixels may be fully or partially covered by snow and ice.

In Figure 4.3 the most important variables influencing the observed signal are shown. Note, that the observed variables are shown as shaded circles and that wavelength, time and space ($x \in \mathbb{R}^2$) are defined using plate notation.

A more general task of remote sensing can be defined as a classification of the hidden, or latent, variable Y_{class} for every land-use type observed by the satellite. Clouds are

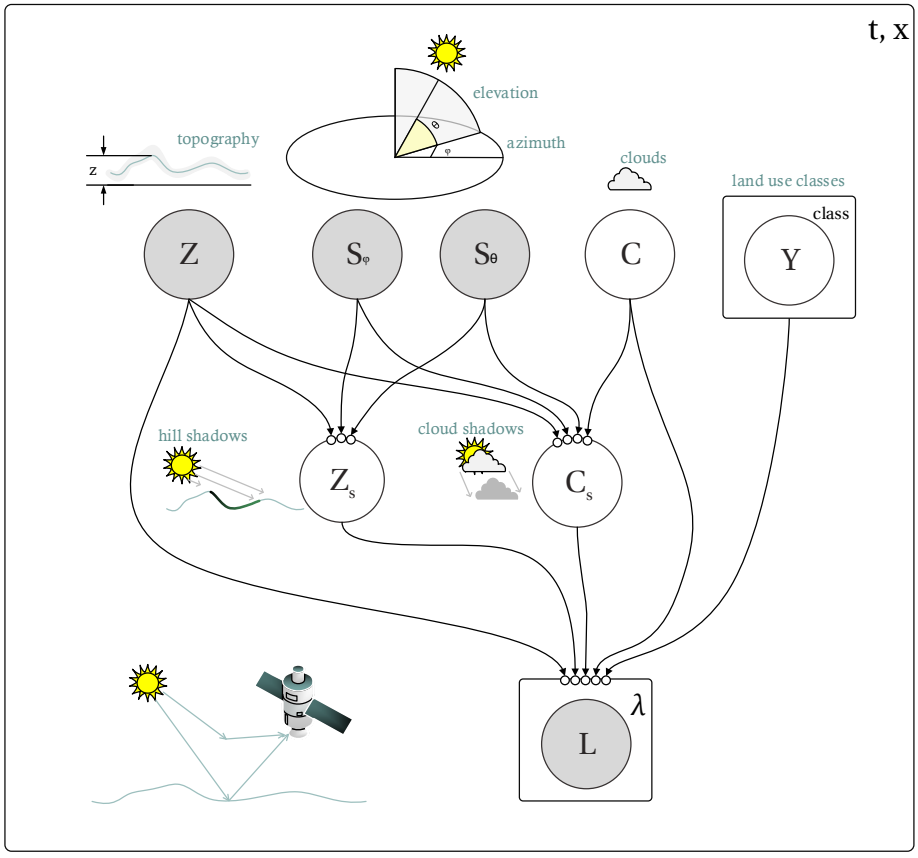


Figure 4.3: BN of a passive multispectral satellite sensor. Filled circles represent observed variables, with a known distributions. Empty circles indicate latent variables that need to be reconstructed. The variable L defines observed radiances, where values in the distribution are based on the solar parameters, cloud cover, topography, as well as on the actual land use classes occurring at a given location. Plate notation is used to indicate dependency on time, space and wavelength.

introduced as a separate variable C, to stress, that it represents a noise and is not mixed with the actual land use type hidden under the clouds or distorted by the cloud shadows.

To indicate a complex light interaction between clouds and Earth's surface topography, a separate random variable Z is added. It is a random variable, because it is never measured exactly, and may be very noisy for most existing global digital elevation datasets, such as 30m SRTM or 3m NED. These errors can be very large (sometimes 50m) and may result in an incorrect interpretation of the observed pixels when used, for example, to model illumination effects, as shown in Chapter 2.

The Sun parameters, S_ϕ and S_θ represent the Sun's azimuth and elevation at the time, when the image was measured, and are usually very accurate, perhaps making it more appropriate to introduce them as parameters instead of random variables.

The Sun's light interacts with clouds and topography, which results in a distortion of the signal, which is expressed by the variables C_s, Z_s, which may indicate the amount of light scattered and absorbed due to clouds and shadows.

The arrows between these variables represent conditional dependencies. For example, the conditional probability representing the amount of light influenced by cloud shadows can be defined as: $P(C_s|Z,S_\phi,S_\theta,C)$. Here, the cloud shadow variable depends on the topography because the actual illumination usually changes depending on the cloud height, which in turn depends on the actual elevation at that location.

To reconstruct (infer) unknown variables, one can try to simulate the joint probability for this PGM factored as:

$$\begin{aligned} P(L_{\lambda_1},...L_{\lambda_N},C,C_s,Z,Z_s,S_\phi,S_\theta,Y_{class_1}...Y_{class_M}) = P(Y_{class_1})...P(Y_{class_M}) \\ P(C)P(S)P(C_s)P(S_\theta)P(S_\phi)P(Z_s)P(Z) \\ P(C_s \mid C,S_\theta,S_\phi,Z)P(Z_s|Z,S_\theta,S_\phi) \end{aligned} \tag{4.3}$$

with N bands λ and M classes $class$ defined as:

$$\begin{aligned} \lambda &= \{blue, green, red, nir, swir1, swir2, temp\} \\ class &= \{water, snow, sand, vegetation, urban, ...\} \end{aligned} \tag{4.4}$$

Note that in addition to the class and wavelength dimensions, most of the variables also depend on space and time, which is indicated as indices of the corresponding plates.

While variables Y_{class} can be simulated as discrete variables, with a corresponding probability mass function representing the probability of a pixel belonging to a given class, we can also think of it as a continuous variable, defined for every class separately and indicating probability of a pixel fraction (varying from 0 to 1) belonging to a given land-use class. Many of the variables in the above equation are known or partially known as well as their respective errors, however, it may be difficult to simulate the whole joint distribution using standard methods (MCMC), mainly because of the amount of data to be processed.

For the water detection task, we are interested in the reconstruction of $P(C_{water})$ distribution, or more generally - as a multivariate distribution $P(C_{water},x,y)$, marginalized

in time and representing the probability of water in every pixel of a given area for some specific period. It is important to note that this can be done only under the assumption that morphology of the waterbody does not change during the selected period. For more complex configurations, this density function should be assumed as a time-dependent function, for example, to represent seasonal surface water variability.

One of the methods to infer this water distribution, as well as other latent variables in the above graphical model, is to estimate them in a discriminative way, directly from the observed reflectance values measured by satellite sensors. With sufficient measurements and under the assumption that morphology of the waterbody does not change significantly, we should be able to condition our joint distribution 4.3 into subsets where clouds are present and absent. Also, instead of integrating this joint distribution, marginalizing it to estimate $P(C_{water}, x, y)$, it can be directly computed by making use of the methods described in Chapter 3 and estimating the frequency of pixels to be covered by water or land.

An alternative could be to fully simulate the above graphical model by introducing some analytical models, like FMask or topographic correction formulas from Chapter 2 to simulate the CPDs representing cloud and hill shadows. After the reconstruction, standard methods used for BNs can be applied to infer all other unknown land-use class variables $P(C_{class})$.

4.5.1. SAMPLING SPECTRAL REFLECTANCE VALUES FROM MULTI-TEMPORAL LANDSAT IMAGES FOR WATER AND LAND PIXELS

In many cases, distributions of the marginal variables introduced in Figure 4.3 are unknown and may represent very dynamic processes occurring at the same time. To demonstrate this complexity, N=100 locations are sampled for the area outlined in Figure 4.4, and then, using only an area covered by surface water, in the middle of the reservoir. Because surface water of the reservoir changes in time, but also because image registration may be imperfect, it is hard to ensure that all sampled values correspond to water and clouds only. However, it should be sufficient to demonstrate the variability of the processes over a period of 30 years.

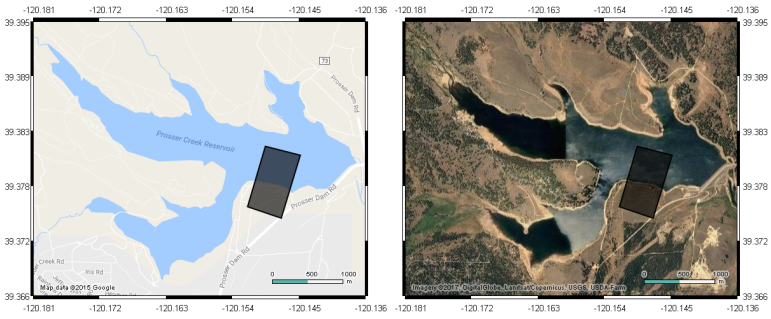


Figure 4.4: Sampling area used for scatter plots, Prosser Creek Reservoir, CA, USA

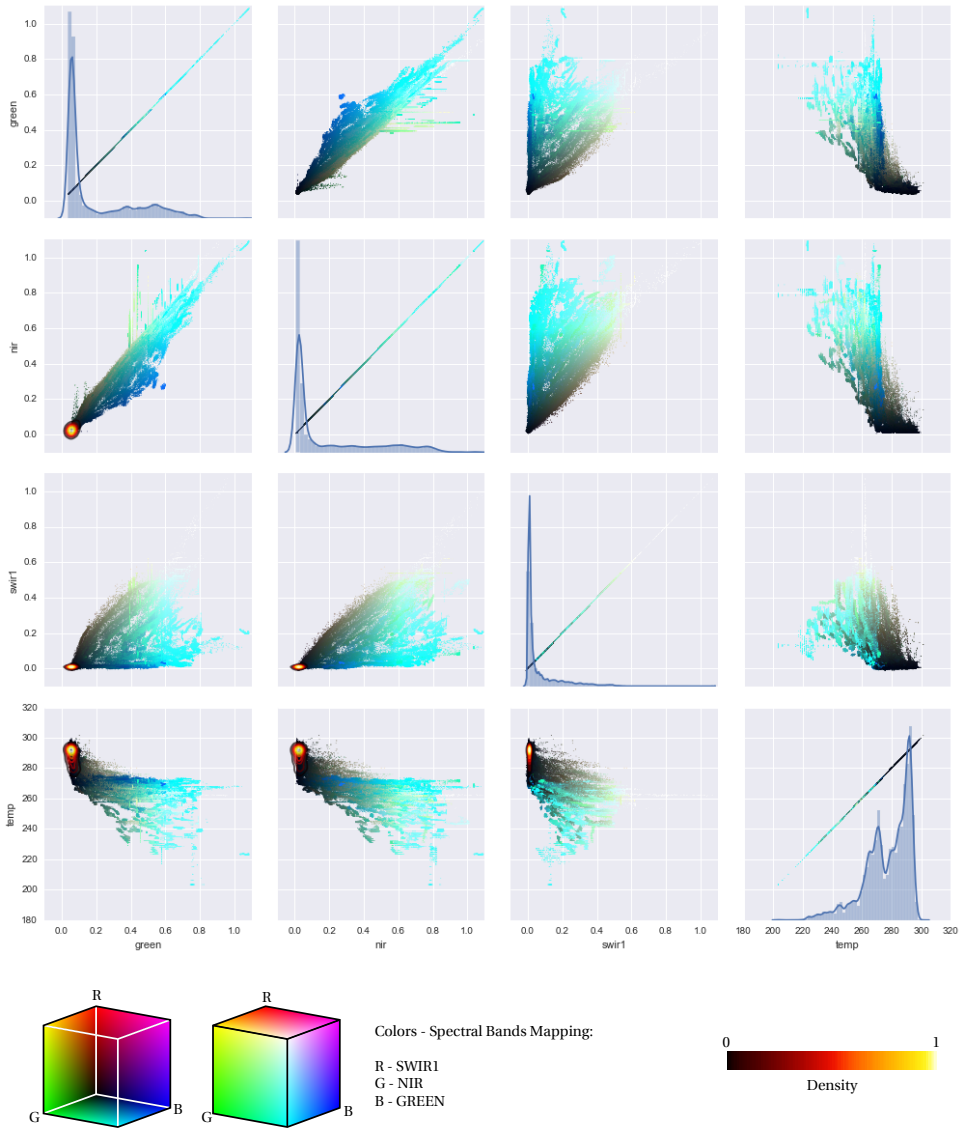


Figure 4.5: TOA reflectance values sampled over the study area but covering mainly water pixels. Scatter plot points are colored as a false-color combination: [swir1, nir, green]. The scalebar indicates probability after applying the Kernel Density Estimation (KDE) method.

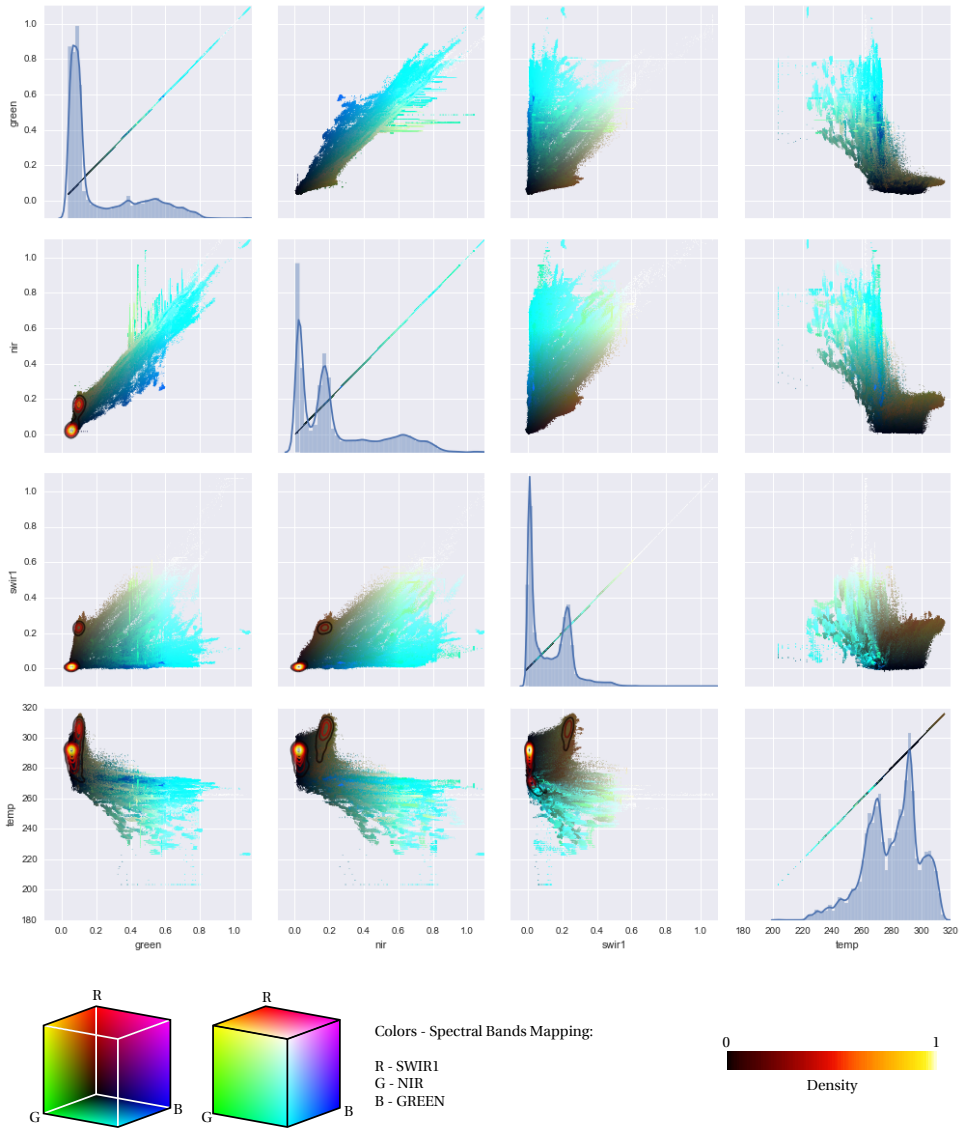


Figure 4.6: TOA reflectance values sampled over study area (water and soil pixels) from all Landsat images available for that area (N=930).

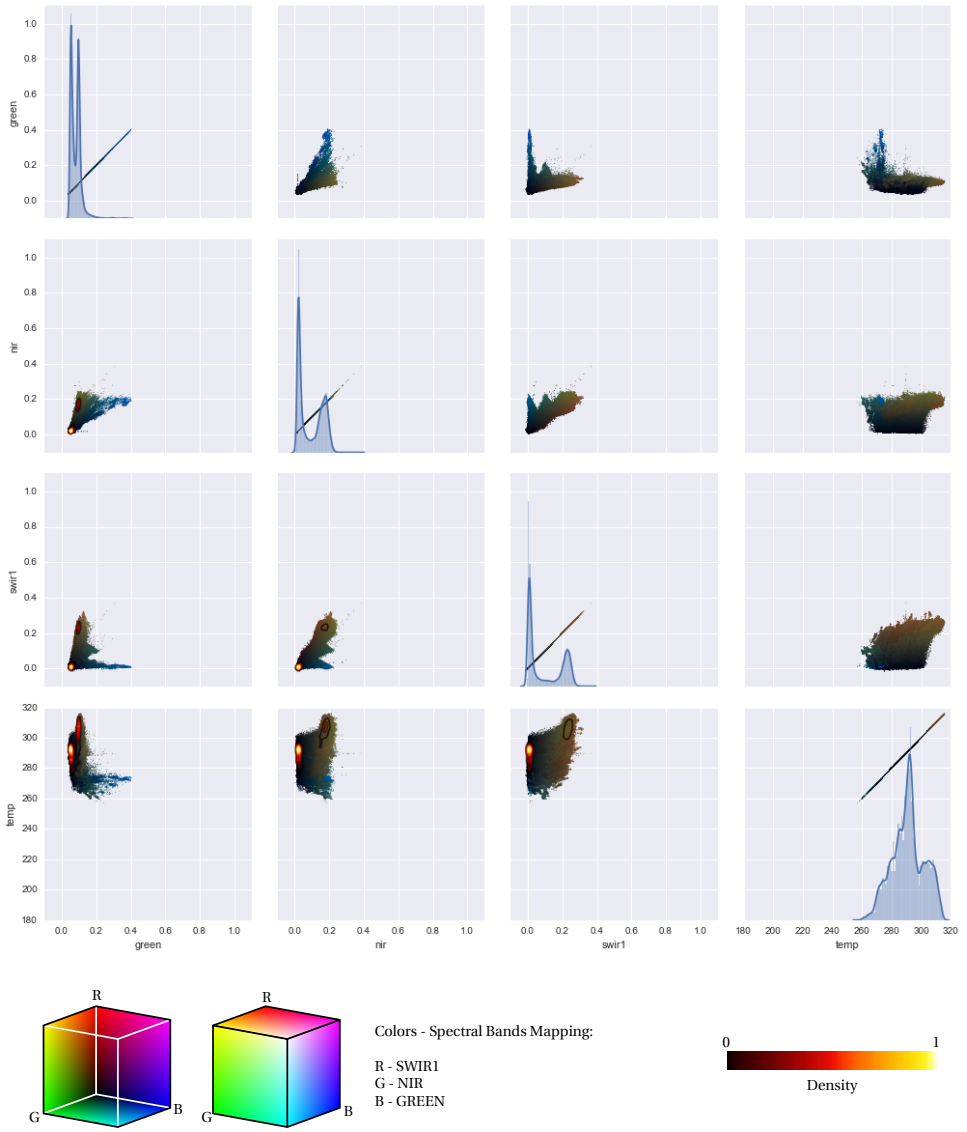


Figure 4.7: TOA reflectance values after conditioning on cloud and snow cover.

The resulting scatter matrix plots, colored as a false-color image using swir1, nir, and green bands, are shown in figures 4.5 and 4.6, for water-only pixels and water plus soil pixels. Additionally, the density of the pixels for every scatter plot in the lower triangle is analyzed. The reason for this exercise is to try to see if different classes can be determined from these distributions or at least to identify the most frequently occurring values in the case of water, soil, clouds or snow. It is interesting to see that for this area,

the distance between the most frequently occurring values, which can be used to separate water and land, correspond to nir and green bands. However, also the temperature band shows a lot of variability, where temperature values are formed by the presence of snow, soil, and water in pixels (three peaks in the histogram of the temperature band) as shown in the directed graph above. The Figure 4.5 represents a joint distribution, conditioned on some of the variables, in this case, on $P(C_{water} = 1)$. As can also be seen, reflectance values for clouds and snow vary significantly in contrast to lower reflectance values in the case of water or soil.

By sampling values in the center of the reservoir and the area covering water and the nearby land, some of the parameters can be automatically reconstructed from these distributions to be used in combination with classical remote sensing methods for cloud detection, such as ACCA or FMask. Adjusting the actual thresholds used in these methods with more objective parameters, it should be possible to automatically identify fully cloud-free images, and then use them to simulate the marginal distribution representing water occurrence.

Here, this approach is used to adjust threshold values used to estimate clouds and snow scores, to identify the most cloud-free images over the reservoir area (in fact, conditioning the joint distribution for L on a $C = 0$). These cloud-free images are used to reconstruct the marginal distribution $P(C_{water}, x, y)$) for a given time interval, assuming that the water dynamics under the clouds looks similar to when there are clouds.

4.6. PROBABILISTIC FILLING OF MISSING PIXELS IN SURFACE WATER DETECTION

One of the ways to solve the problem of filling missing pixels, or to remove the less probable false-positive water pixels is to make use of the a priori information about surface water dynamics. By adjusting cloud (or snow/ice) detection algorithms with local statistics, we can identify images where the number of bad or missing pixels is small, resulting in very accurate water masks. The figures 4.8 and 4.9 show edges for several of these water masks and a 2D probability density image estimated from these water masks.

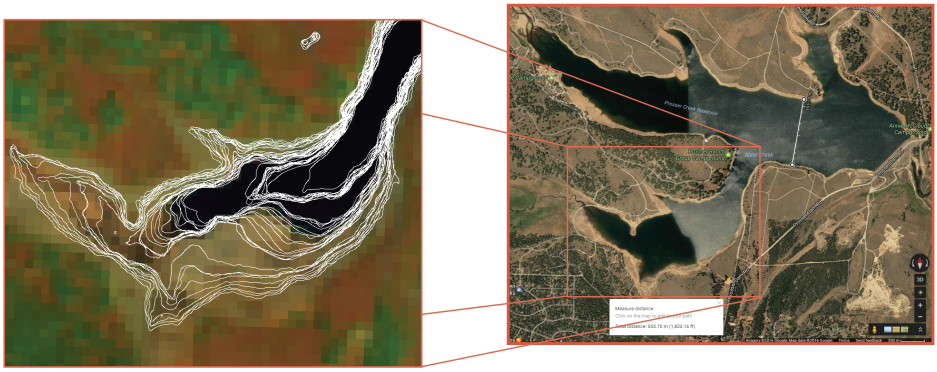


Figure 4.8: Surface water mask detected from cloud-free images (multiple sensors). The images were resampled using bicubic interpolation before the actual mask was detected. This has resulted in a better final water occurrence.

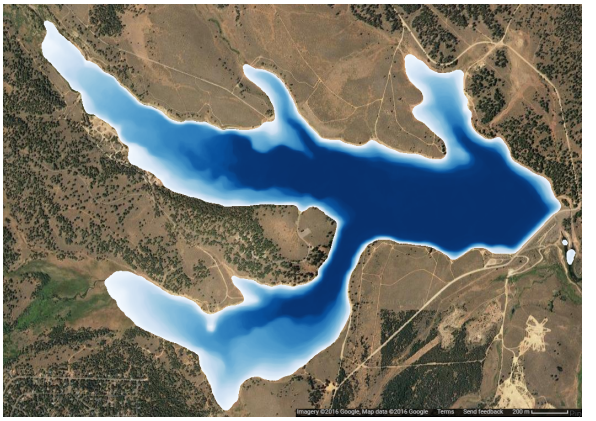


Figure 4.9: Water occurrence (prior) estimated from cloud-free images measured by multiple satellites (ASTER, Sentinel-2, and Landsat)

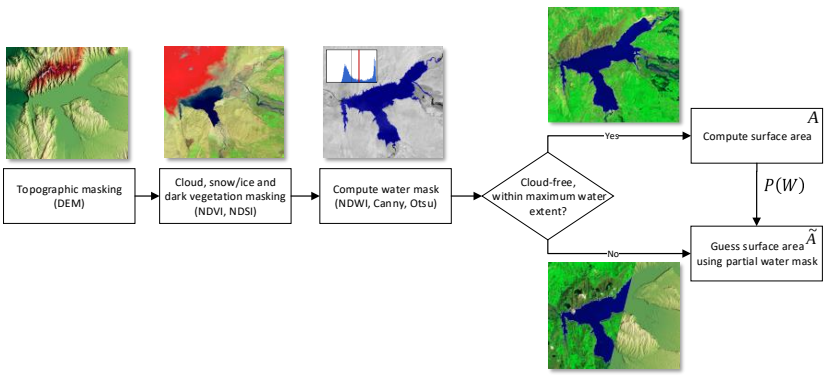


Figure 4.10: A high-level workflow of the surface water detection method

The method to fill surface area for missing pixels is based on the use of this probability density image to infer water mask, employing Bayes' rule. The actual algorithm

involves a few more steps, where the reconstructed water PDF is also used during edge detection (less probable water edges are removed). The resulting algorithm, outlined in Figure 4.10 allows for the automated estimation of reservoir surface area using both cloud-free and partially cloud-free images.

In fact, the inference step of the algorithm can be schematically visualized as in 4.11, where on the left side, the original graph is shown, with all irrelevant variables removed.

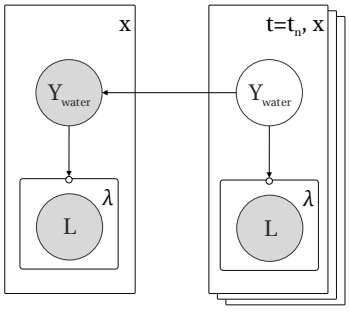


Figure 4.11: An example of a BN for a single scene, assuming marginal distribution on the left estimated during the previous step.

During the first step, where the water probability density function was estimated, the Landsat 7 and Sentinel-1 images were excluded, because of SLC-OFF errors and because the SAR water detection algorithm may be less reliable, and because of effects like Bragg scattering, resulting in false-negative pixels.

In many cases, slightly better results are achieved for noise-free pixels when compared to probabilistic estimate (our initial prior estimate is never perfect). An additional step was added to choose between a generative surface water estimate and the one performed using discriminative methods. The choice between different water masks detected using generative (density function) and discriminative (dynamic thresholding) methods is made based on the quality of the pixels, which depends on cloud and snow scores. For very hilly topographies, it may be necessary to make use of Z_s as well, to avoid misclassification of hill shadows as water pixels.

4.7. APPLICATION TO THE RESERVOIR SURFACE WATER AREA RECONSTRUCTION

A hybrid, stepwise method, combining image processing algorithms and remote sensing methods would be more appropriate for the accurate estimation of the water mask.

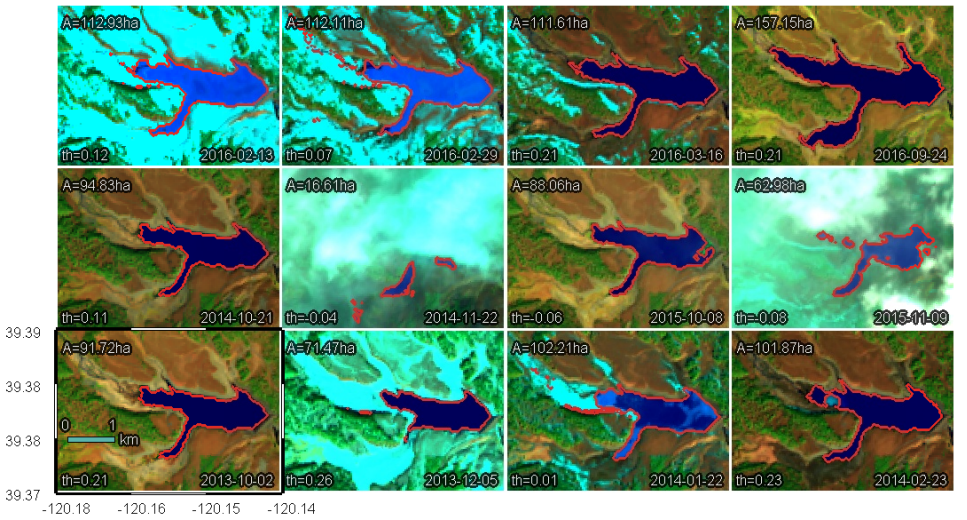


Figure 4.12: Canny/Otsu, probabilistic update

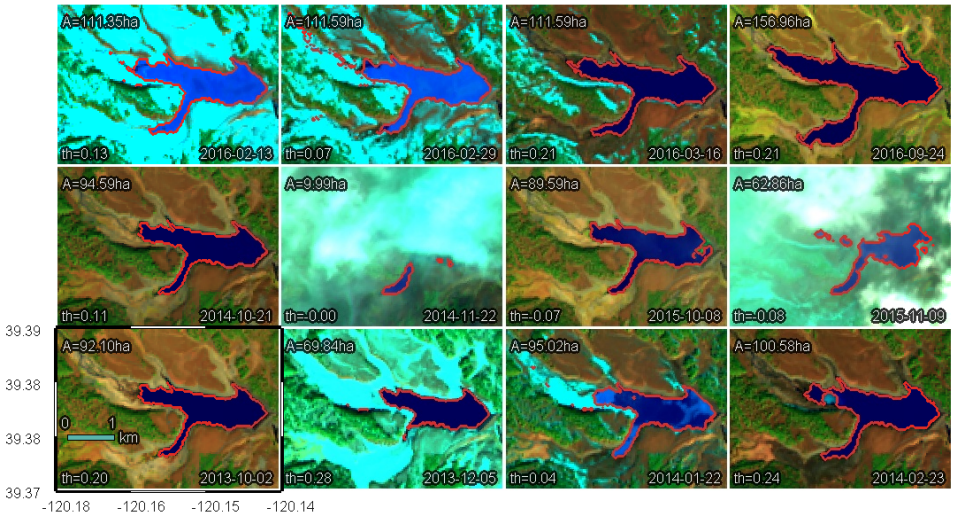


Figure 4.13: Canny/Otsu, HAND < 15m, probabilistic update

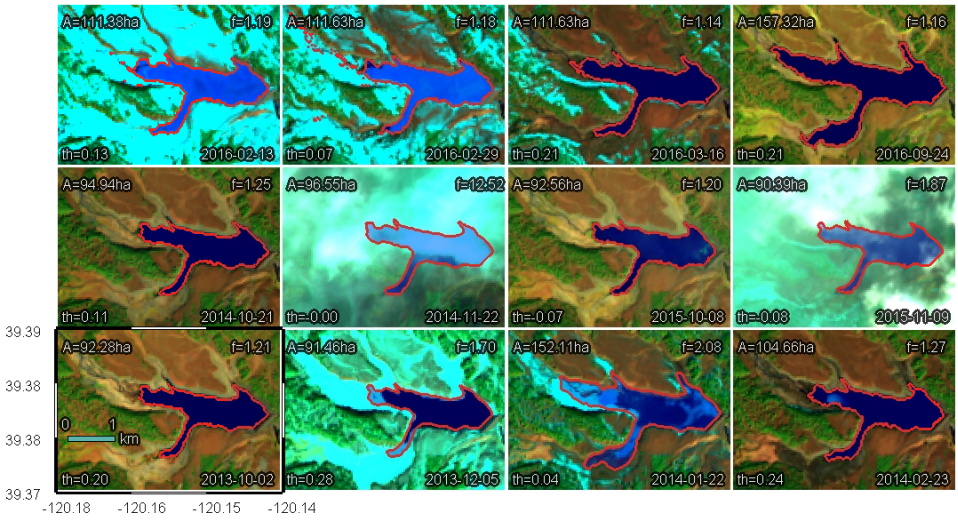


Figure 4.14: Canny/Otsu, probabilistic update, HAND < 15m, fill

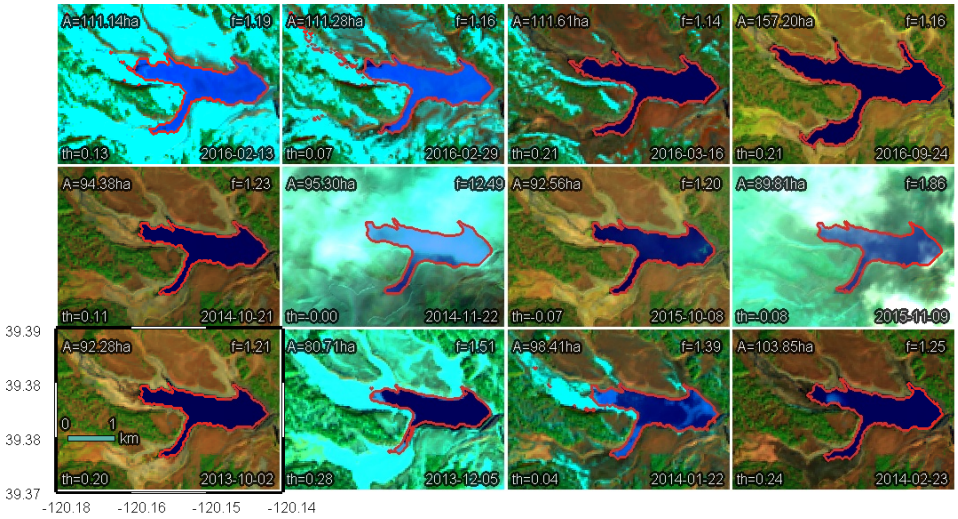


Figure 4.15: topographic, Canny/Otsu, probabilistic update, HAND < 15m, fill

4.8. MEASURING PERFORMANCE OF SURFACE WATER DETECTION WITH FILLING APPLIED

To evaluate the overall multi-step algorithm performance, where a combination of the dynamic thresholding and the generative estimation of surface water mask were involved, the final surface water area values were compared with the surface water area estimates computed using the statistical model introduced in Chapter 3.

Additionally, surface water masks not passing the quality criteria were excluded. The quality score was defined as a combination of multiple parameters, such as the fraction of false-positive, false-negative pixels, and the number of cloud and snow pixels.

The final surface water area time series, together with the original surface area, estimated from only optical satellite images and NDWI=0, are shown in Figure 4.17.

To further explore the quality of the estimates, a residuals analysis is performed for the final results, comparing them to the estimates based on the model constructed for the reservoir. We can see from the figures, that the errors are mostly normally distributed, $RMSE = 0.059 km^2$, indicating good performance of the algorithm.

In the case of cloud presence, the actual threshold variability of the spectral indices makes them less applicable, resulting in much larger variation of the surface water mask estimates. Use of the step-wise approach introduced in the last two chapters provides a way to automatically detect surface water with a very high degree of accuracy.

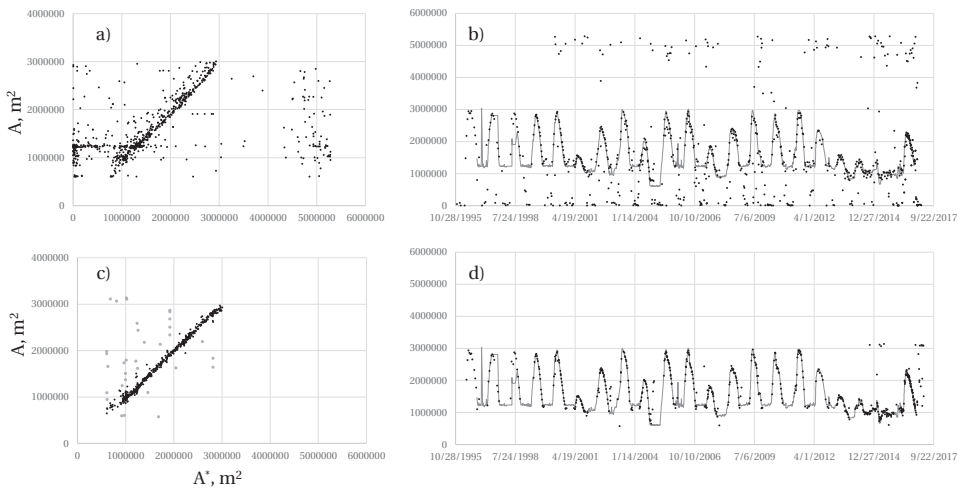


Figure 4.16: Comparison of estimated surface water area A vs. simulated surface water $A*$, based on the linear regression constructed previously. Scatter plots (a, b) show results of the surface water detection algorithm, without (a) and with (c) probabilistic filling. Time series charts (b, d) correspond to estimated (dots) and measured surface water area values.

4.9. ANALYSIS OF RESIDUALS AND OUTLIERS

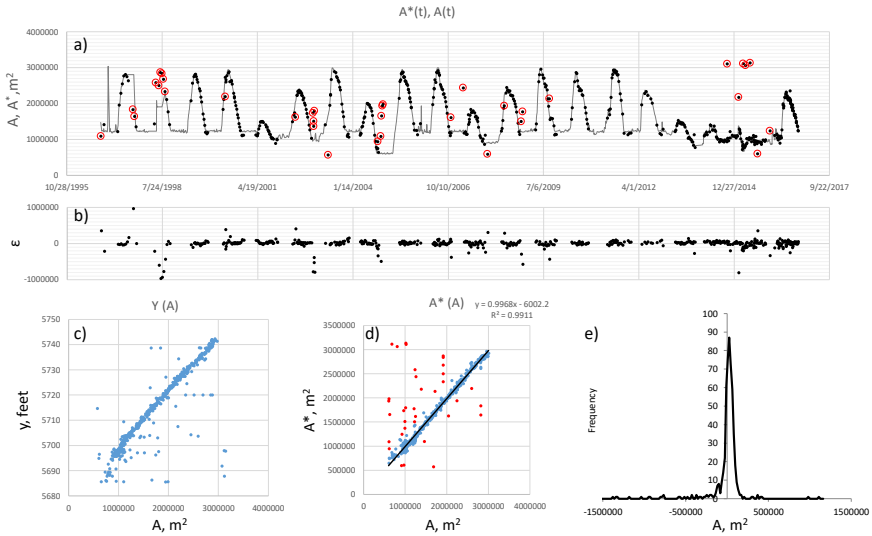


Figure 4.17: Analysis of residuals (surface area estimated from satellite images vs. surface area estimated from the linear model). The top chart (a) compares time series estimated from water level measurements and satellite photos (dots), as well as outliers ($\epsilon > \sigma$). Chart (b) shows residuals values, assuming surface water area A* as the ground truth. Scatter plots (c, d) show a comparison of measured water level and estimated surface water area (c) vs. water levels and A/A* plot (d). The chart (e) indicates histogram of residual values (<3%).

4.10. DIFFICULT SURFACE WATER DETECTION EXAMPLES

To demonstrate the method performance, figures 4.18, 4.19, 4.20 demonstrate a few examples where the surface water mask estimate is particularly difficult. Using the method presented here, an almost perfect water mask can be reconstructed, when compared to the observed water level values.

The first example in Figure 4.18 illustrates the algorithm performance for the case in which the reservoir is almost fully covered by clouds, and only a small area in the upper part of the reservoir can be seen. Because these pixels are located in a very flat area, where water surface area changes correlate with changes in water levels, accurate estimates of the actual surface area could be made by utilizing the density function constructed above. This case may result in larger errors in some reservoirs, where large morphological changes may take place in the river. In this case, our prior distribution, used to provide posterior estimates, is less informative. In this case, it would be more appropriate to utilize the temporal variability of the surface area. This is relatively easy to achieve by constructing the water occurrence distribution as three-dimensional, instead of two-dimensional - $P(C_{water}) = f(x, y, t)$, where the probability of water at a given location may change in time as well.

Figure 4.19 shows a case where the original surface water mask estimate is incorrect with regards to the very dark cloud shadow pixels. The image was acquired at a relatively low sun elevation, resulting in lower reflectance values, with a very evident cloud

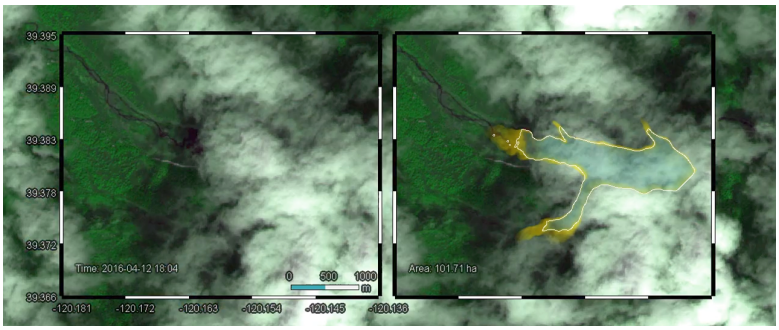


Figure 4.18: Surface water reconstructed from a small amount of visible water at the edge of the reservoir. Visible water pixels are located in a dynamic part of the reservoir.

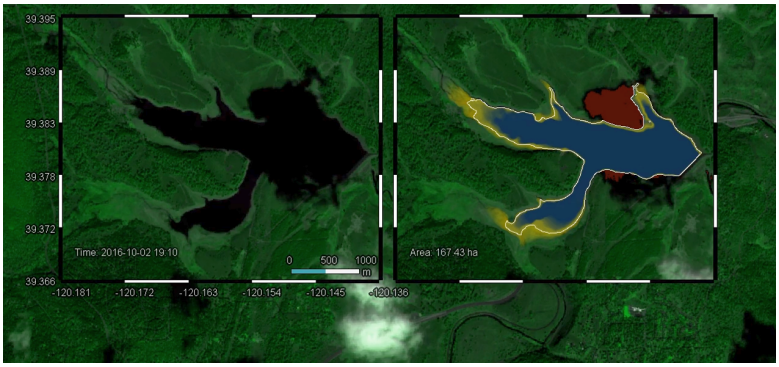


Figure 4.19: Surface water reconstructed from a Sentinel-2 image with a very dark cloud shadow present. False-positive surface water detected during first pass (red) is corrected after the use of Bayesian update step.

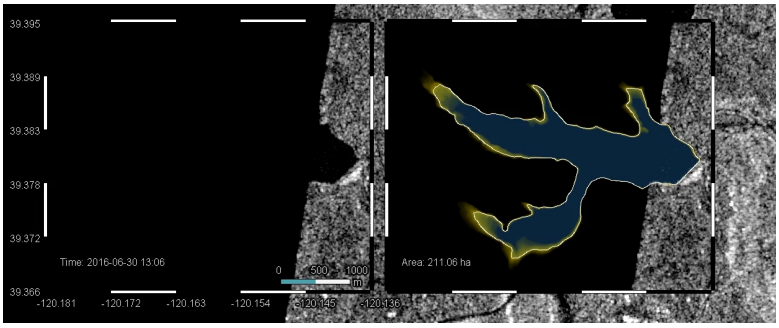


Figure 4.20: Surface water mask reconstructed from Sentinel-1 SAR image. A very limited number of pixels are visible in a steep part of the reservoir.

shadow. In fact, looking at the values NDWI of the pixels, corresponding to water and cloud shadow only, it was impossible to detect the surface water mask reliably. In this situation, as well as in the case of SLC-OFF images, the surface water mask can be accurately estimated. This claim holds true because a strong correspondence between water occurrence and detected surface water edges is expected to occur.

A slightly more difficult case can be seen in Figure 4.20, where a small part of the Sentinel-1 SAR backscatter intensity image is used to estimate surface water. The challenge here lies in the fact that this part of the reservoir is relatively steep, resulting in a much higher chance of an incorrect surface water area estimate. Another challenge here is related to the presence of a very dark edge present in the Sentinel-1 image, which was excluded by eliminating low entropy areas along the image edges.

Density-based methods may be extremely useful when the waterbody is partially covered by clouds, cloud shadows, or when only part of a waterbody is visible, due to limited swath or due to sensor artifacts (SLC-OFF for Landsat 7).

4.11. CONCLUSIONS AND DISCUSSION

In Chapter 3, it was demonstrated that automated surface water detection from multiple multispectral missions is feasible, however, the surface water detection from noisy satellite images remained difficult. Where such noisy images are frequently occurring (e.g. in predominantly clouded regions) or where high frequency observations are required, the applicability of the method presented in Chapter 3 is limited and more advanced methods are required.

In this chapter, it has been shown that even when a small part of a surface water body is observed and the remainder obscured, the rest of the surface water body can be inferred by analyzing historical imagery in which accurate estimates of the surface water body were established. The accuracy of the final surface water mask depends on how informative the observed surface water area is, which directly depends on how many changes (inter- or intra-annual) occur to the surface water at a given location.

The historical imagery is introduced in the water classification algorithm through Bayesian inference. However, this also assumes independence of the observations from each other, resulting in overestimation or underestimation of surface water area in several cases, such as very noisy images with multiple types of noise present: snow, clouds, shadows or very noisy backscatter data. To overcome these problems, many other methods can also be applied such as Hidden Markov Models (HMM), where transition probabilities can be estimated based on multiple historical observations.

Essentially, the method presented in this chapter allows enhancing the spatial resolution of the final estimated water mask. This is done by combining surface water masks from multiple sensors, with the spatial resolution varying between 10m (Sentinel-2) up to 30m (for Landsat). The final density function was estimated at a resolution of 5m combined with the bicubic resampling applied to the images to interpolate values. Even-though the actual spectral signature gets damaged at the sub-pixel level, it was possible to generate very accurate surface water mask estimates due to the use of local dynamic thresholding methods introduced in Chapter 3.

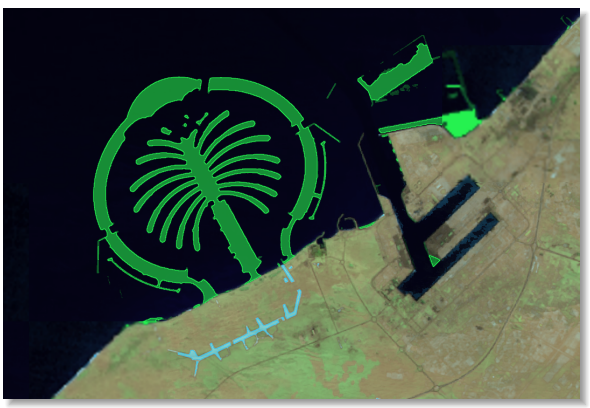
The framework presented in this chapter can also be used in combination with more general image segmentation methods instead of local thresholding. Alternatives may in-

clude image segmentation algorithms such as local k-means (SLIC, Achanta et al. [2012]), combined with methods such as Markov Random Fields (MRF) or Conditional Random Fields (CRF) to infer the actual land use types based on the spatial neighborhood and temporal variability of reflectance values.

5

LONG-TERM SURFACE WATER CHANGE DETECTION

Methods to detect long-term surface water changes are discussed. A simple method based on reflectance percentile composites and linear regression is presented. As an example, Siling Lake in the Tibetan Plateau is used as a study area. The advantages and drawbacks of the method are discussed.



Keywords: *M_3, reflectance percentiles, NDWI, linear regression, long-term surface water changes.*

This chapter is based on G. Donchyts, F. Baart, H. Winsemius, N. Gorelick, J. Kwadijk, and N. van de Giesen. Earth's surface water change over the past 30 years. *Nature Climate Change*, 6(9):810–813, 2016a.

5.1. INTRODUCTION

As we have seen in the previous three chapters, accurate surface water detection is a complex task and may require many steps. However, different land use changes may occur at different temporal scales. If we are interested only in long-term (inter-annual) land-use changes, more generic statistical methods can be applied to detect these changes. With sufficient observations available and under the assumption that statistics of reflectance values for different years remain the same, we can use relatively simple methods, when compared to those discussed in the Chapter 4.

The first step that should be done when analyzing optical passive sensor satellite imagery is the elimination of clouds and cloud shadow effects. Cleaning images from cloud, cloud shadows, and snow can be a challenging and time-consuming process, especially when images partially covered by clouds need to be considered. Many methods were developed to do this properly, as has been discussed in chapters 2, 3, and 4. These methods can be used to construct completely cloud-free images, or to extract information required to analyze temporal changes in different land cover classes.

As an alternative to this, we can generate inter-annual composites directly from the top-of-atmosphere or surface reflectance values. In recent years, this approach was used in Hansen et al. [2013] to study forest cover changes by employing a combination of metrics, such as: (*a*) selected percentile values (*b*) mean reflectance values for selected percentile ranges (*c*) slope of linear regression of band reflectance values versus image date. Furthermore, they make use of supervised classification methods based on decision trees to relate these metrics to homogeneously varying subsets of data.

Here, a similar approach to estimate long-term surface water changes is introduced. However, instead of developing multiple metrics like in Hansen et al. [2013], a single combined metric employing percentiles is used to estimate average cloud-free reflectance values and the slope of linear regression to identify pixels where long-term changes have been occurring.

The resulting method is also more resource efficient because applying linear regression after averaging is cheaper to compute. The cost of applying linear regression directly to reflectance values is $O(n^3)$, and the cost to calculate the average value for every pixel is $O(n)$, where n is the number of analyzed images. The O notation is frequently used in computer science to indicate scalability of the algorithm with the growth of the dataset. By employing a two-step approach, surface water changes can be estimated in almost $O(n)$ time, making it attractive to be applied across multiple sensors, assuming similarity in reflectance values among different satellite sensors for some spectral bands. For surface water change studies, the most used spectral index is NDWI. The index uses near-infrared and green bands. These bands have similar spatial and spectral resolution across multiple Landsat missions and seem to be in agreement with regards to spectral response and spectral resolution, except for Landsat 8. For Landsat 8, the OLI sensor covers the thinner spectral range for the near-infrared band when compared to TM and ETM+ sensors used by Landsat 4, 5, and 7 Roy et al. [2016], Angal et al. [2014]. A detailed comparison of spectral responses between Landsat 5 and 7 can be found in Teillet et al. [2001] and between Landsat 7 and 8 in Flood [2014]. Additionally, the literature says that data produced from Landsat was cross-calibrated to ensure consistency and continuity of values between different missions Mishra et al. [2014]. This makes combining images

across different Landsat sensors, and the use of statistical methods to process their values, very promising.

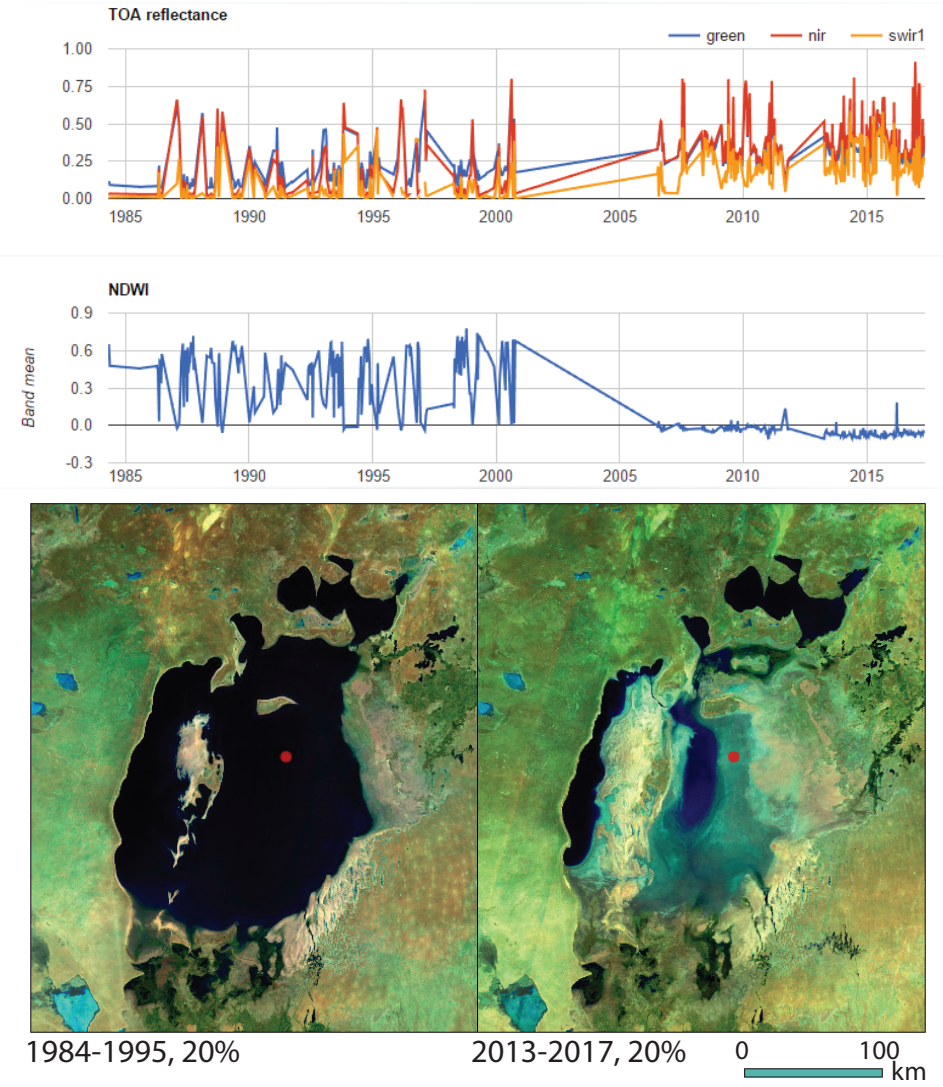


Figure 5.1: Aral Sea, Uzbekistan. False-color reflectance percentile composite (swir1, nir, green, 20%). The red circle indicates the location where reflectance and NDWI values shown on charts were plotted. The composite images and the time series were computed from 4325 Landsat scenes between 1984 and 2017. Source: `https://code.earthengine.google.com/c1ccc09c6b5e7d73bd3a3601f8edfaf3`

Surface water is very dynamic and may change on a daily, monthly, and annual basis. To estimate long-term, inter-annual surface water changes from raw satellite images, we have to exclude effects caused by both clouds and by intra-annual or seasonal varia-

tions, assuming these changes remain the same across years. Examples of intra-annual changes can be seasonal changes in surface water due to natural processes, a regular water cycle, changes of surface water levels caused by tidal effects in coastal areas, and variability in surface water levels due to man-made activities, such as the operation of reservoirs, which may also cause surface water changes in the downstream rivers. Long-term surface water changes may be due to natural changes of water boundaries, such as erosion or accretion, natural changes caused by climate changes, or manmade changes in rivers or coastal zones (embankments, land reclamation, construction of reservoirs).

In reality, a combination of these surface water changes may occur and interfere with each other, making the task of surface water change detection extremely difficult.

To demonstrate the variability of reflectance values in different bands for an area where both long and short-term changes took place, a long-term time series in the middle of the Aral Sea is obtained for pixels which have dried up during the past few decades, as shown in Figure 5.1. This area has faced dramatic changes regarding surface water and, as can be seen from the reflectance values, this has influenced both reflectance and corresponding NDWI value changes. While it would be possible to detect the trend of NDWI values from the raw reflectance value time series, doing this may be less efficient, mainly because NDWI values for many pixels corresponding to clouds may look very similar to the values corresponding to water. Additionally, it would require significantly more computing resources.

5.2. LONG-TERM SURFACE WATER CHANGE DETECTION

To address long-term surface water changes, we can first try to exclude short-term variability of reflectance values by computing average images for a long time intervals, excluding effects of clouds and cloud shadows in this way. By estimating this type of composite images for multiple time intervals, followed by the application of spectral water indices, such as NDWI, we should be able to reconstruct long-term variability of the surface water changes, which will be reflected in the variability of the NDWI values. One of the important questions arising using this method is how to select the best percentile, allowing to minimize the effect of clouds and cloud shadows, and at the same time, avoiding the appearance of cloud shadows in the images.

The method will work only if a sufficient number of observations is available. A large number of images is required to ensure that the distribution of values, used to compute reflectance percentile composites, converges to the actual distribution representing different land or atmosphere values for a given location is based on similar distributions. Ideally, the resulting percentile should look the same for two areas when no long-term surface water changes take place.

Because surface water has a very distinctive spectral signature, with very low reflectance values in almost all bands, the presence or absence of water in a given pixel will be reflected in the final distribution represented as a probability distribution function (PDF) or a cumulative distribution function (CDF) for a given location.

To demonstrate the main steps used in the methods, surface water changes which have occurred around the Siling Lake are analyzed - one of the lakes of Tibetan Plateau, China. In the last few decades, an enormous area of new surface water has been created due to climate changes. Figure 5.2 shows the final estimate for the submerged land,

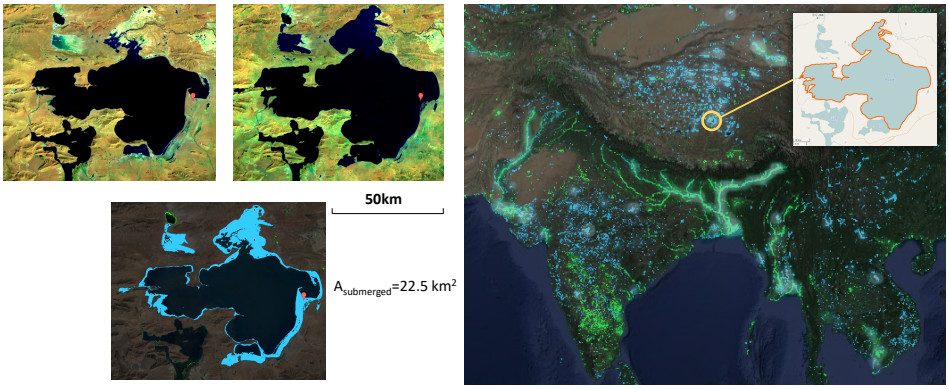


Figure 5.2: Siling Lake, Tibetan Plateau, China. Two reflectance composite images at the top-left show how the lake looked like in 80's (left) and how it looks like today (right). The blue color indicates submerged land due to climate changes. Source: `https://code.earthengine.google.com/87d0057d8b3d28817b1da7cfc1bd6b6e`

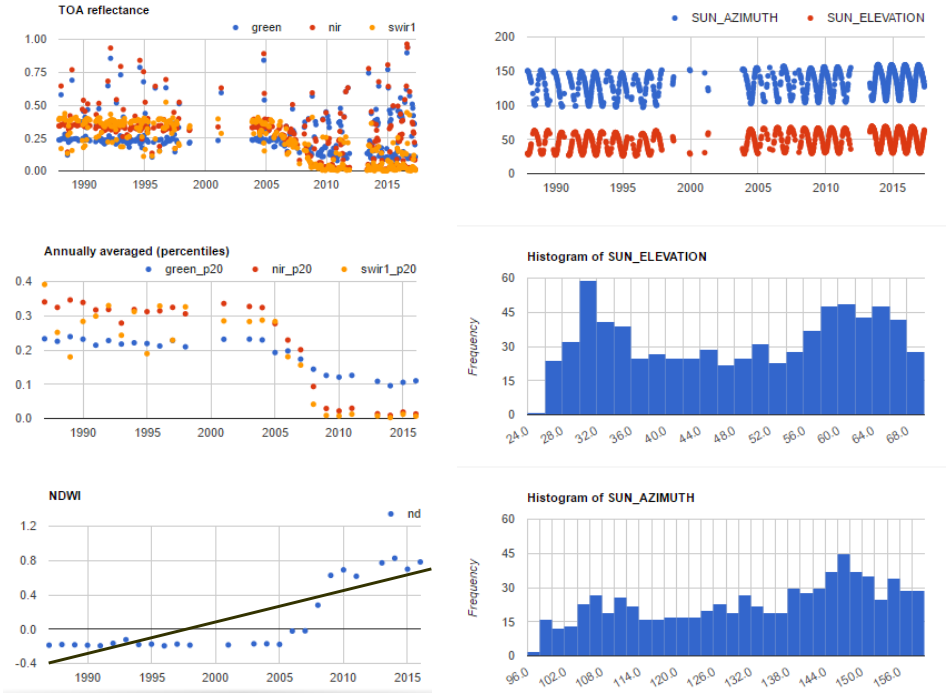


Figure 5.3: Estimating surface Siling Lake, Tibetan Plateau, China. The actual TOA reflectance values (left-top), annually-averaged (20%) reflectance values (left-middle) and NDWI values, estimated from annually-averaged reflectance values (left-bottom). The three charts on the left indicate values and distributions of the sun parameters for the images used to compute these values.

equal to $22.5 km^2$.

A straightforward and efficient way to estimate long-term surface water changes may be to use a two-step approach. First, the intensity percentile composites are computed for all temporal intervals T used for the change detection analysis:

$$F(\rho) = \int_{-\infty}^{\rho^i} P(\rho) d\rho \tag{5.1}$$

where F - cumulative distribution function, the superscript index i denotes the i^{th} percentile, ρ - reflectance values for a given band.

Then, for every percentile reflectance value ρ^i, the spectral index is computed as an image:

$$I = \frac{\rho^i_{green} - \rho^i_{nir}}{\rho^i_{green} + \rho^i_{nir}} \tag{5.2}$$

After that, the linear regression is performed on a set of all index values for all intervals T:

$$I = \beta_0 + \beta_1 T + \epsilon \tag{5.3}$$

The resulting slope of the linear regression β_1 is then analyzed to identify pixels, where significant surface water changes took place.

Usually, the percentile is determined empirically and corresponds to low reflectance values, but not too low, to avoid confusing water with cloud shadows, which can also be very dark for green and nir (or swir) bands used to estimate spectral indices. For those pixels, where the slope β_1 of the linear regression is significant. The spectral index values I are also tested to exclude false positive changes, detected for images where most images correspond to water or land:

$$\begin{aligned} I &= min\left(I, I^{water}_{min}\right) \\ I &= max\left(I, I^{land}_{max}\right) \end{aligned} \tag{5.4}$$

where I^{water}_{min} and I^{land}_{max} indicate minimum and maximum values of the spectral index to be considered as water and land correspondingly. The equation 5.4 is applied for every interval used during temporal averaging.

The first equation of the 5.4 allows to filter pixels, where the slope of the regression is significant, with all values still belonging to water. The second equation of the 5.4 does the same to eliminate locations where all of the temporal intensity values belong to the land.

Additionally, Normalized Difference Vegetation Index (NDVI) vegetation index is used similarly to remove locations, where, for example, deforestation took place. In this case, NDWI values may also change significantly in time, but the actual changes do not correspond to surface water changes.

The actual number of images available need to be adjusted based on the cloud fre-

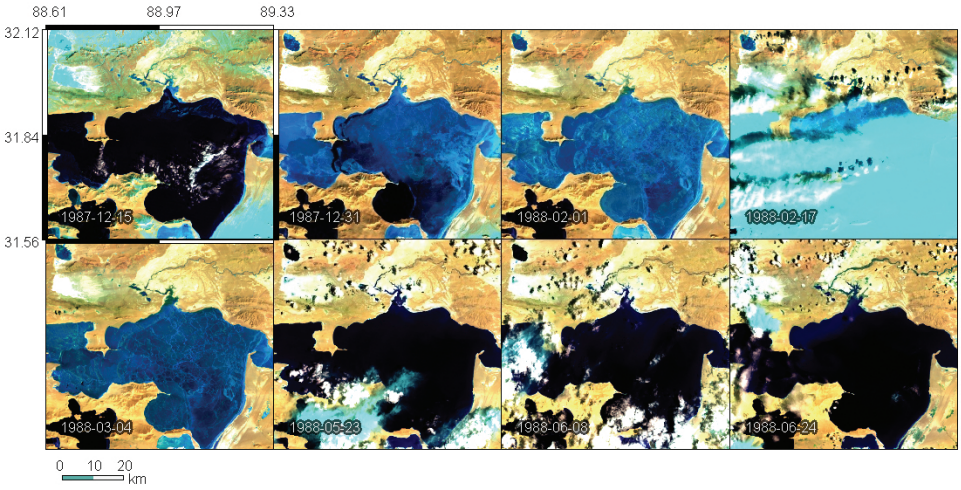


Figure 5.4: Examples of false-color (swir1, nir, green) images from 1998 acquired by Landsat TM sensor for Siling Lake, Tibetan Plateau, China. Source: `https://code.earthengine.google.com/882ab2929f536676d56bc79cbc895ded`

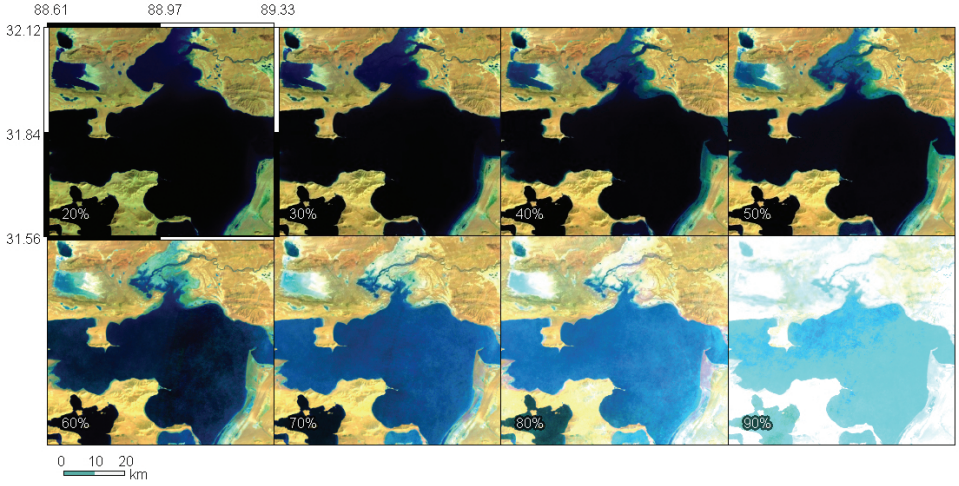


Figure 5.5: False-color (swir1, nir, green) reflectance percentile images estimated from 751 images acquired by multiple NASA Landsat missions during 1984-2017, Siling Lake, Tibetan Plateau, China

quency present for a given area. For the area around this lake, the average cloud frequency is 13%, with relatively low intra-annual variation, as estimated by Wilson and Jetz [2016] from MODIS images. However, for some areas on Earth, mainly near the equator, 95% of the images may be covered by clouds. This means, that to apply this method, many more images need to be analyzed. In this case, only long-term surface water changes may be estimated using passive optical sensors.

5.3. TOPOGRAPHIC NOISE FOR INCONSISTENT IMAGE COLLECTIONS

Even though the method described in Chapter 5 allows detection of surface water changes at high accuracy, a substantial commission error may occur in mountain regions. These errors are caused mainly by combined errors of elevation and azimuth, causing differences in hill shadows in composite images, used to detect surface water changes. For global analysis, most of these errors were eliminated by masking-out the final surface water changes images with the topographic index derived from HAND. The topographic mask, constructed by HAND < 40m, was used to filter out these errors. The HAND dataset used for the processing was generated from 30m SRTM and some auxiliary 90m datasets to cover areas where it was not acquired (>60 degree north latitude) Donchyts et al. [2016c]. However, for very accurate surface water change estimates this may also result in omission errors in the areas when DEM values are incorrect. For the time being, no free high-resolution DEM exists to be used as a topographic index (via HAND). Therefore, the Aqua Monitor website does not perform this correction during on-the-fly estimation of surface water changes. This may change in future versions, when a more appropriate method is introduced.

Examples of these errors are demonstrated in Figure 5.6.

The same errors were reported in the Hansen et al. [2013] dataset, which uses similar methods to estimate surface water changes (interval mean percentile composites versus percentile composites in Aqua Monitor).

A time consuming approach to eliminate these errors could be to perform topographic correction on every image before passing them to the change detection algorithm. However, this may significantly increase the amount of resources needed, making this type of analysis hardly achievable today.

5.4. CONCLUSIONS AND DISCUSSION

The method presented in this chapter showed surprisingly good performance for most of the places on Earth and was used to analyze global surface water changes over the last thirty years. Despite its simplicity, the method can very accurately detect long-term surface changes. Additionally, it is easy to implement using the Google Earth Engine parallel processing platform.

Another advantage of the method is that it is less resource intensive when compared to a more streamlined approach to detecting surface water changes. Hence it can be used for large planetary-scale studies and for a wide range of users, which is demonstrated in Chapter 6. For example, all images are first classified e.g. using supervised machine learning methods, and then the changes are computed from the resulting thematic maps.

However, it may be difficult to reliably detect surface water changes when:

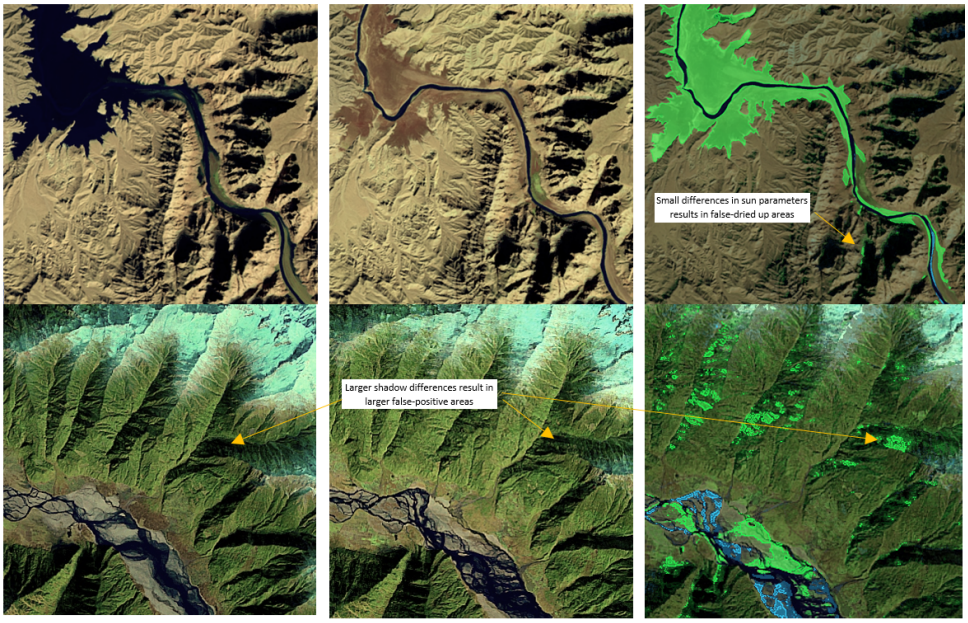


Figure 5.6: False-positive composites (swir1, nir, green) and surface water changes commission errors due to inconsistent sun parameters in the samples used to compute changes in the hilly areas (2000-2015). Top: upstream of the Lake Mead (36.10, -113.90); bottom: river near Yogongxiang, China (30.27, 94.90).

(a) the number of images used to compute percentiles is low

(b) surface water changes follow a more complex pattern than a one directional trend

(c) multiple land-use changes are present, represented by low reflectance values but irrelevant to surface water changes

(d) complex topographic conditions combined with unequally distributed sun parameters in different time intervals

Some of these issues can be easily addressed, for example, by changing empirically chosen percentiles to ensure one captures pixels free from cloud and snow effects.

Additionally, image samples used to compute percentiles can also be selected to only represent certain seasons, where some of the effects are less present, filtering image collections by day-of-year or by sun parameters, to avoid the most evident topographic effects.

The methods presented here work best when reflectance values, representing complex land-use and atmospheric changes, are similarly distributed. While for surface water this is frequently true, it may be less trivial under very complex land-use change (c). To detect these kind of changes automatically, additional inference steps may be required, for example, where different patterns can be recognized in the distributions such as those presented in Chapter 4, Figure 4.5.

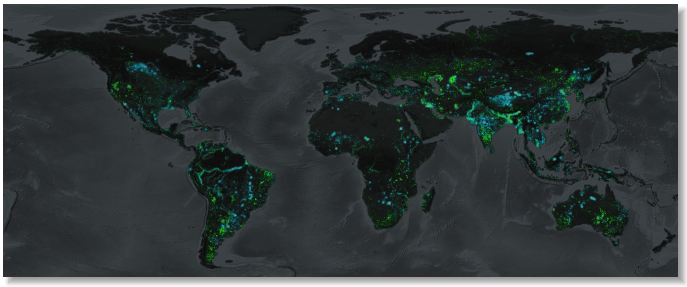
While linear regression may be a good choice to detect monotonic long-term changes, surface water may follow much more complex trends, requiring the use of more complex methods to capture them. For example, logistic regression can be used instead of linear regression to detect abrupt changes such as construction or decommissioning of reservoirs, or to analyze autocorrelation to detect recurring surface water changes.

With the increasing number of satellite images available, the use of statistical methods may be a powerful instrument, providing an easy way for change or anomaly detection before the use of more complex methods.

6

EARTH'S SURFACE WATER CHANGE FOR THE LAST 30 YEARS - AQUA MONITOR

Has the world become wetter or drier? Can we see global trends in the changes of coastlines, and are these trends also apparent where we live? Is the total surface water storage on land growing or shrinking at global and local scales? This chapter presents results of the global surface water change study using data from all Landsat missions (2PB). All calculations were performed using Google Earth Engine infrastructure and the algorithm discussed in the previous chapter. The new tool, Aqua Monitor, illustrates how emerging cloud platforms for large satellite data analysis, are rapidly removing the thresholds to the use of planetary-scale data. The main finding of the Aqua Monitor show that the largest contributors to the transition between water and land include Aral Sea (water to land) and Tibetan Plateau (land to water).



Keywords: *global surface water changes, climate changes, land reclamation, erosion, accretion.*

This chapter is based on G. Donchyts, F. Baart, H. Winsemius, N. Gorelick, J. Kwadijk, and N. van de Giesen. Earth's surface water change over the past 30 years. *Nature Climate Change*, 6(9):810–813, 2016a.

6.1. INTRODUCTION

CHANGES from land to water and vice versa are extremely relevant as witnessed by many recent news items. Rapid retreat of Tibetan Plateau glaciers observed during the last decades Chen et al. [2016]. The impoundment of the Three Gorges Dam in China is causing massive inundations, forcing about 1.3 million people to resettle Jackson and Sleigh [2000]. New islands along the coast of Dubai are created to provide new secluded areas for leisure and residence for the wealthy. The Mississippi Delta is losing thousands of hectares of land per year due to soil subsidence and lack of sediments Giosan et al. [2014], further aggravated by sea-level rise. The president of Kiribati declared that his people would need to move to new grounds to prevent them from dying from the effects of sea-level rise on the atoll Weiss [2015];

The causality of appearing or disappearing water surfaces may strongly depend on the case-specific context. Although atolls, such as Kiribati, are under severe threat, the exact effects of sea-level rise on coastal erosion, globally, may strongly depend on biophysical interactions as well, particularly in coastal marshes Storlazzi et al. [2015], as atolls may increase accretion rates as sea-level rise progresses Kirwan et al. [2016]. The impoundment of the Three Gorges Dam has resulted in a reduction in sediment concentrations in the downstream Yangtze River of about 70%. Unexpectedly, this reduction has not led to a retreat of the downstream submerged Yangtze River Delta so far Dai et al. [2014], contrasting to what is happening in the Mississippi Delta.

These examples demonstrate that conversions - and the stories and reasons behind them - can vary widely and are often the result of compounding causes. Therefore, general conclusions cannot be drawn from a limited sample of case studies. Instead, planetary-scale monitoring is needed to understand (and disentangle) the causes of detected changes and their attribution to natural variability, climate change or man-made change.

6.2. METHOD TO ESTIMATE EARTH'S LONG-TERM SURFACE WATER CHANGES

To obtain changes in water and land occurrence, the method described in the Chapter 5 is applied. The resulting algorithm was also implemented as an open-source website, called Deltares Aqua Monitor. Aqua Monitor prepares cloud-free average reflectance composite images for multiple years using all satellite images from all NASA Landsat missions and estimates surface water changes on-the-fly. The Deltares Aqua Monitor typically uses a 2-year period to obtain a sufficiently large number of images. However, to quantify long-term surface changes which have occurred during the last 30 years, we have extended this period to a longer baseline. The comparison was performed using two intervals: 1985-2005 and 2013-2016. The much longer first interval was required to ensure we have sufficiently large sample to compute statistics, as the number of images before 1999 is significantly smaller than those of recent years.

The Aqua Monitor establishes water-land and land-water occurrence on-the-fly by estimating the MNDWI spectral index values and performing trend analysis for these MNDWI values over both user-selected periods. For the analysis in this chapter, these selected periods were 1985-2005 and 2013-2016 as mentioned above. To decrease noise for high latitudes, all Landsat images acquired during night time were excluded. Additionally, we apply a topographic mask based on Height Above the Nearest Drainage (HAND)

to decrease the noise in hilly areas, occurring due to mismatches between sun elevation and azimuth parameters in both periods. All pixels representing mountain hills (HAND > 150m) are excluded, as described in Donchyts et al. [2016b]. The final estimates of the surface area (land to water and water to land) were obtained by including pixels where a large slope of the linear regression was observed.

6.3. EARTH'S SURFACE WATER CHANGES AT 30M SPATIAL RESOLUTION

Because the spatial scale of many surface water changes is small, we have upscaled the final results (30m to 300m) in a form of two global images: (*a*) surface water changes with transition water to land (green) (*b*) surface water changes with transition land to water (blue).

To produce the final global image and the Figure 6.1, both global images were mosaicked with an alpha channel, to ensure visibility of the pixels where both changes take place.

The Deltares Aqua Monitor (`http://aqua-monitor.deltares.nl`) is the first global scale tool that shows at 30-m resolution where water is converted to land and vice versa. With assistance from Google Earth Engine, it analyses satellite imagery from multiple Landsat missions, which observed Earth for more than three decades. The Aqua Monitor provides a much needed, fully planetary-scale view on changes in land and water occurrence (Garcia et al. [2016]). Documented and undocumented changes due to man-made interventions, natural variability and climate change are revealed. It is possible to look at any area of interest and use the outcomes for scientific advances at planetary-scale, review large-scale statistics on land and water conversion, or open a discussion with stakeholders in a given area on the basis of unbiased information on water and land occurrence and change.

6.4. SURFACE WATER CHANGE EXAMPLES

This chapter demonstrates the planetary-scale ability of the Aqua Monitor by showing some significant and contrasting water–land conversions. It provides a perspective of what these abilities - which are now available to any researcher or stakeholder - mean for climate research. First, the planetary-scale changes in the occurrence of water and land are shown. It can be seen that globally, between 1985 and 2015, an area of about 173,000 km^2 — about the size of the state of Washington — has been converted to land, and an area of 115,000 km^2 has been converted into water. An overview of the largest changes found globally, aggregated per drainage basin, are shown in Figure 6.2. The **Tibetan Plateau** and the **Amazon River** are the areas with the largest area conversion to water. The **Aral Sea** stands out for conversion to land. As changes in surface water only affect people at a regional and local scale, we show some contrasting cases for different areas in Figure 6.3 and describe these below.

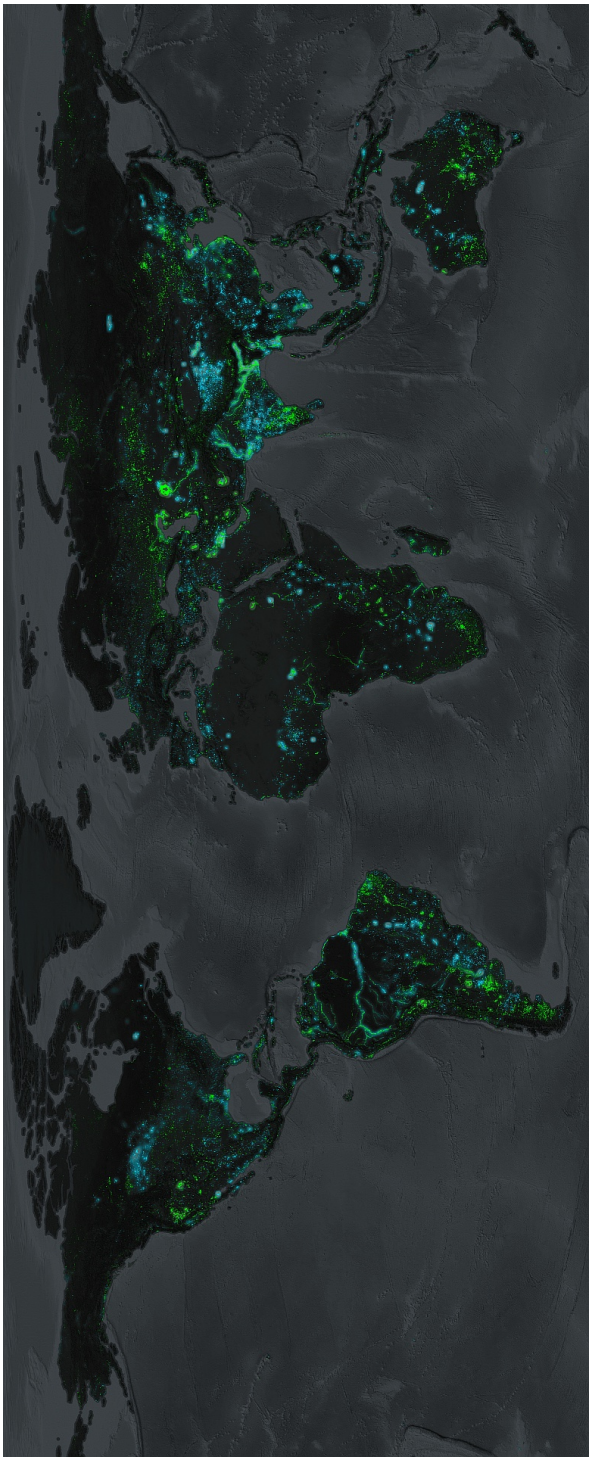
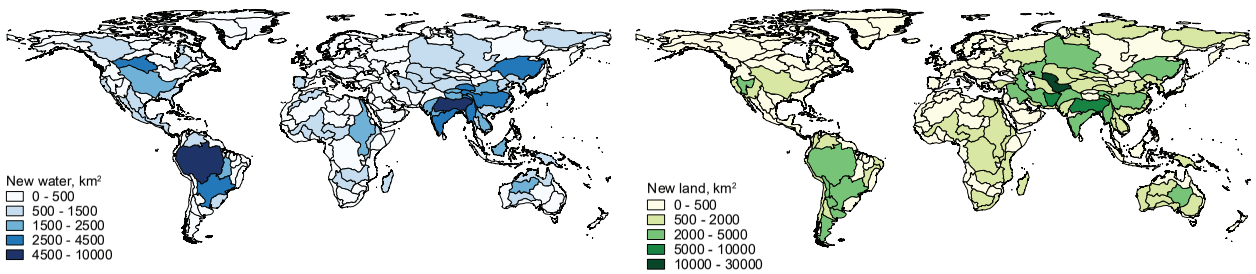


Figure 6.1: Heat map of global surface water and land changes. Blue lighting shows where land was converted into water over the period 1985-2015, Green lighting shows where water was converted into land over the same period. The intensity of the colors highlights the spatial magnitude of the change.



Name	Area, km2
Tibetan Plateau	7661
Amazon River	7058
Ganges-Brahmaputra Rivers	5495
Rio de la Plata	4410
River Nelson	4101
India	3677
River Amur	3494
Yangtze River	3238
Tigris & Euphrates Rivers	2636
Thanlyin;Sittang;Ayeyarwady	2592
Indus River	2312
Mississippi River	2231
Mekong River	2074

Name	Area, km2
Aral Sea	27841
Ganges-Brahmaputra Rivers	9580
Hamun Lake	6044
Tigris & Euphrates Rivers	4897
Amazon River	4888
Lake Eyre	4672
Rio de la Plata	4020
India	3851
Lake Urmia, Caspian Sea	3759
Great Salt Lake;Malheur Lake	3752
Ural River	3455
Amur River	3344
Ob River	3068

Figure 6.2: Largest surface water and land changes from 1985 until 2015 grouped by drainage basins. Left: changes from land to water in blue. Right: changes from water to land in green.

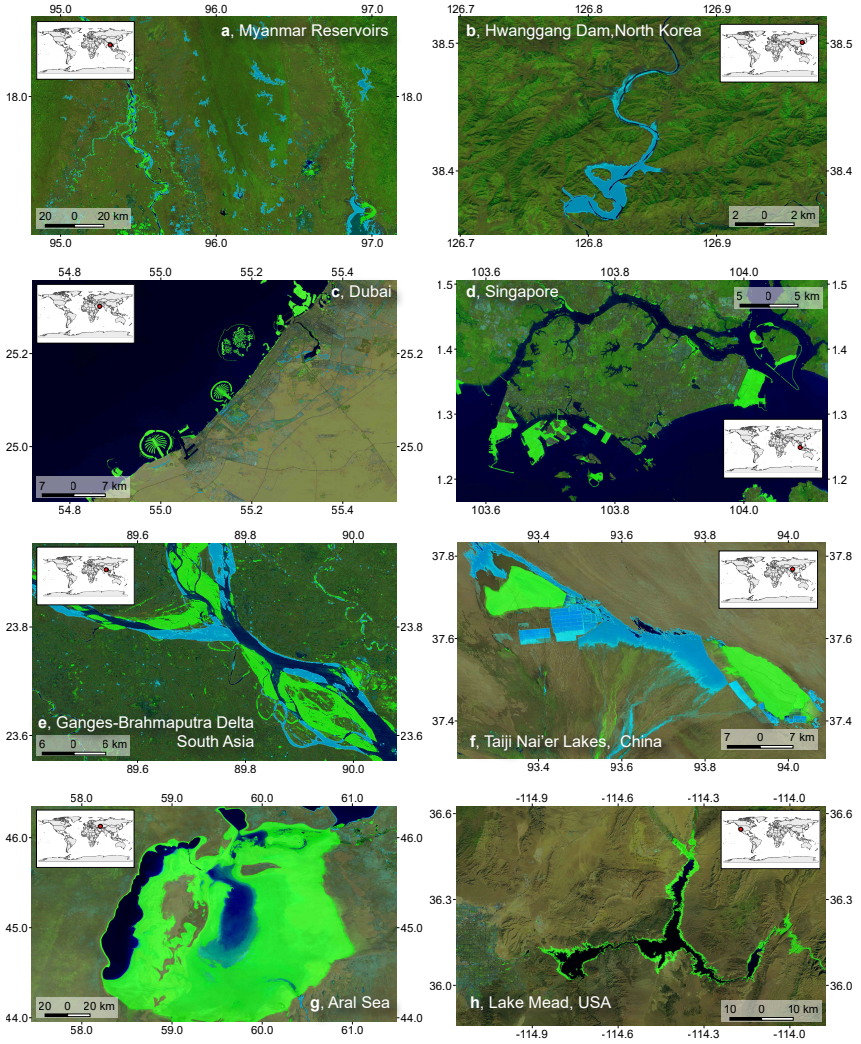


Figure 6.3: Examples of surface water changes between 1987 and 2015, detected using Aqua Monitor. Blue: conversion from land to water. Green: conversion from water to land

6.4.1. KNOWN AND UNKNOWN (MYANMAR VS NORTH KOREA)

Although many countries report on their dam construction, information in more remote or isolated areas is lacking. In Myanmar, the Global Reservoir and Dams database Lehner et al. [2011] shows an increase in water surface between 1985 and 2010 of about 400 km^2. Using the Aqua Monitor, we have counted the appearance of 1,180 km^2 of new water surface in this region over the same period (Figure 6.3a). The previously unmapped damming of the Rimjin River in North Korea, close to the border with South Korea, resulted in a storage surface of 12.4 km^2 (Figure 6.3b). This is, in fact, the Hwanggang Dam, mapped 35 km eastward. The dam was the topic of an international dispute between South and North Korea after the 2009 flash flood that killed six fisherman Sang-Hun [2009].

6.4.2. LUXURY VERSUS NEEDS (DUBAI VS SINAPORE)

The largest coastal water–land change is the construction of the Palm Island and adjacent islands along the coast of Dubai 80 km^2 (Figure 6.3c). Many countries have shaped and extended their coastlines by land reclamation. The motives to reclaim land are highly diverse. In Dubai, the main motivation was to increase the coast length, providing more room for recreation Davidson [2009]. In contrast, reclamations in Singapore (76 km^2; Figure 6.3d) are necessary to support its economic growth `http://www.mnd.gov.sg/landuseplan`).

6.4.3. NATURE VERSUS MAN-MADE (GANGES-BRAHMAPUTRA DELTA VS TAJI NAJER LAKE)

Results of the Aqua Monitor only show compound impacts of natural and human change or variability. It is often hard to tell what the causes are for a change without looking at the details of the local water and sediment budget. Although changes in meanders in the Brahmaputra River Delta are clearly natural (Figure 6.3e), the Mondrianlike shapes formed near Taiji Nai'er lakes in China, are clearly man-made (Figure 6.3f).

6.4.4. DISRUPTIVE VERSUS GRADUAL (ARAL LAKE VS LAKE MEAD)

An example of disruptive change can be found at the Aral Sea, once the fourth-largest lake in the world. Since the 1960s, Soviet engineers diverted the rivers away from this endorheic lake to irrigate cotton and wheat agriculture Glantz [1999]. The lake has almost entirely dried up, losing about 27,650 km^2 of surface water (Figure 6.3g). The positive impacts of a recent restoration program Micklin [2016] in the northern part can be observed as well. A slower drying lake can be found near Las Vegas at Lake Mead, the largest freshwater supply in the United States. It lost 222 km^2 over the same period (Figure 6.3h). The 10% probability scenario that the lake would have already dried out by 2013 Barnett and Pierce [2008] did not come true, but the lack of inflow from the Colorado River will cause the lake to gradually disappear.

6.5. NEAR-SHORE COASTAL SURFACE WATER CHANGES

One of the questions which can be answered by remote sensing is how much surface water changes took place along the global coastline. This can be important to detect effects such as sea level raise and other changes caused by natural processes or anthropogenic factors. Some examples of natural processes causing the largest coastline changes are

migration of mudbanks along the East-North part of South America, causing erosion and accretion of the coastline and mangrove areas up to 3km in width; changes along the coastline surrounding the mouth of the Mississippi River, though here many of these processes are also reported to be caused by man-made changes, such as construction of levees, causing reduction of deposits of fresh water and silt from the river; erosion and accretion of the Brahmaputra River Delta, causing appearance of 20m long islands and erosion of large areas - this is probably the area where the largest natural changes take place along the coastline.

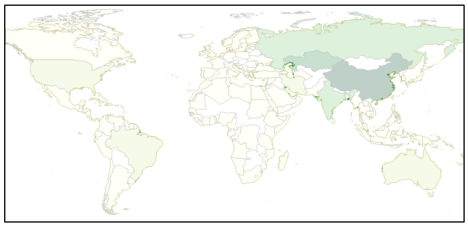
At the same time, we can conclude from the results of our study, that many man-made changes along the coast have resulted in a reclamation of large areas. When not considering the Aral Sea, the major contributor to the man-made changes is China, where about 6000 km^2 of new land has been claimed from the sea. Land reclamation occurs practically along the whole Chinese coastline, resulting in areas up to 5km wide claimed from the seas.

Even though this would be a very interesting study to perform on its own, we have tried to roughly estimate these near-shore changes. One of the main challenges we have faced when performing this analysis is to select a baseline coastline. In our experience, the accuracy of all existing and freely available vector and raster datasets defining coastlines is insufficient for this kind of study, mainly because the coastline is dynamic in many locations. Therefore, a multi-step approach is required to firstly estimate the baseline coastline for a given period and to compute long-term surface water changes from this baseline. To simplify the analysis, we have used a 40km buffer around OSM coastline, followed by aggregation of these changes for every country. The resulting surface areas for both accretion and erosion are reported in the Figure 6.4.

6.6. CONSLUSIONS AND DISCUSSION

Big satellite data analytics at anyone's fingertips may have strong implications on monitoring capacities and associated actions. At a very local scale, a civilian can now assess without any expert assistance, if coastal erosion threatens their house. At a regional scale, a downstream riparian state can monitor from year to year, if upstream neighbours are establishing new impoundments. Finally, at a global scale, agencies such as the United Nations International Strategy for Disaster Reduction can monitor the appearance of new, possibly flood hazard reducing, reservoir storage capacity.

Implications for climate research follow from the fact that the available time series are long enough to cover a climatologically relevant period. The period of 30 years allows distinction between noise of (multi) annual variations, such as the lake surface area of Lake Nasser, and long-term trends in land and water distribution, such as the vanishing of the Aral Sea. Feeding changes in land and water surfaces into regional climate models will lead to better representation of circulation patterns, as well as local climate, in particular in the vicinity of large wetlands Mohamed et al. [2005]. Another example is the attribution to sea-level rise or other drivers of coastal erosion in soft sediment coastal areas Barros et al. [2015]. Drivers such as sea-level rise, sediment delivery and subsidence, and the biophysical properties of the coastline, can cause highly nonlinear erosion and accretion. Quantifying the contribution of these drivers would benefit tremendously from information on multiscale patterns of erosion and accretion from



Country	Area, km2
India	2593
China	2402
USA	1904
Indonesia	1540
Mexico	1506
Russia	1119
Myanmar	1118
Vietnam	907
Bangladesh	876
Brazil	842
Pakistan	825
Australia	815
Iran	644
Canada	617
Iraq	576

Country	Area, km2
China	6019
Kazakhstan	3635
India	2757
Russia	2618
Australia	1803
Brazil	1664
USA	1571
Bangladesh	1356
Iran	1226
Indonesia	1109
Pakistan	1094
Myanmar	878
Turkmenistan	817
Vietnam	783
Egypt	620

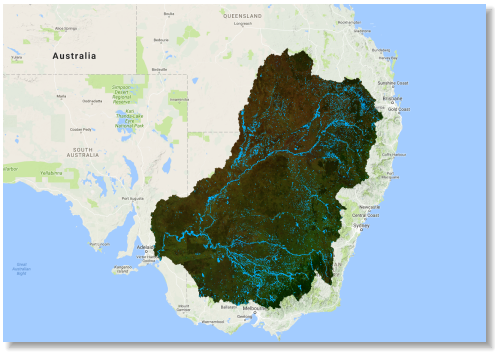
Figure 6.4: Largest land-water changes within 40km distance around the coastline from 1985 until 2015. Light colors indicate country-scale changes, while bright colours show the actual location of the changes. Left: changes from land to water. Right: changes from water to land.

low (global) to very high (local) resolution. The climate community is presented with the capacity to take into account these new planetary-scale observation abilities.

7

SURFACE WATER MAP FOR MURRAY-DARLING RIVER BASIN, AUSTRALIA

This chapter studies derivation of a surface water map of the Murry-Darling River Basin, Australia. Three surface maps are derived from three datasets: Landsat 8, 30m DEM (SRTM), and OSM. Methods developed in Chapters 3 and 5 are applied to detect permanent surface water. Positional differences between datasets are analyzed and demonstrated to be less than 60m for OSM and Landsat 8. The differences between the new water mask and SRTM-based linear features and hilly areas are slightly larger (110m). The overall agreement between OSM and Landsat 8 water masks is about 30%. It is demonstrated that all three datasets complement each other in terms of their quality and coverage.



Keywords: *Murray-Darling River, permanent water, Landsat 8, MNDWI, SRTM, HAND, OpenStreetMap, CART, positional accuracy*

This chapter is based on G. Donchyts, J. Schellekens, H. Winsemius, E. Eisemann, and N. van de Giesen. A 30 m resolution surface water mask including estimation of positional and thematic differences using landsat 8, srtm and openstreetmap: A case study in the murray-darling basin, australia. *Remote Sensing*, 8(5):386, 2016b.

ACCURATE maps of surface water are essential for many environmental applications. Surface water maps can be generated by combining measurements from multiple sources. Precise estimation of surface water using satellite imagery remains a challenging task for remote sensing due to sensor limitations, complex land cover, topography and atmospheric conditions. Alternatively, in case of hilly landscapes, a drainage network can be extracted from high-resolution digital elevation models.

Additionally, VGI initiatives such as OSM can also be used to produce high-resolution water body maps. They are frequently digitized and validated manually using the highest resolution available data sources. In this study, a high-resolution water mask is generated using Landsat 8 imagery and OSM as well as the (potential) drainage network using 30m SRTM. The approach presented here focuses on the surface water detection from Landsat 8 imagery and comprises the use of a 15% intensity percentile Landsat 8 imagery measured during 2013-2015.

To detect the surface water mask, a new non-parametric unsupervised method is used, based on Canny edge filter and Otsu thresholding, as discussed in Chapter 3. For hilly areas, the method is extended with an additional supervised classification step used to refine the water mask. Furthermore, it was applied across the Murray-Darling Basin, Australia. Differences were analyzed between the new Landsat 8 based water mask, OSM, and potential water mask derived from the digital elevation models. The results show that about 50% of the OSM linear water features can be confirmed using the water mask extracted from Landsat 8 imagery and the drainage network derived from SRTM.

7.1. REMOTE SENISNG AND VOLUNTEERED GEOGRAPHIC INFORMATION (VGI)

The main reason why OSM was chosen in the current study over local, authoritative datasets, is that it provides a global coverage, even though its local coverage and quality may vary. Additional research would be required to perform a detailed comparison of the datasets presented in this paper to the local Australian authoritative datasets, such as Surface Hydrology Crossman, S. & Li [2015], Water Observations from Space Mueller et al. [2015], or 5m Digital Elevation Model (DEM) of Australia Geoscience Australia [2016].

We derive a high-resolution water mask using Landsat 8 imagery and OSM as well as the (potential) drainage network using 30m SRTM. Extracting a water mask from OSM data is relatively straightforward (Section 2.2), but the other data sources require a specialized workflow. Our approach to derive a surface water mask from Landsat 8 imagery is described in Section 2.4, involving the steps to compute cloud-free average reflectance composites in Section 2.4.1. Additionally, we introduce a new non-parametric unsupervised method to detect water in flat areas (Section 2.4.2). We also propose a supervised classification step to refine the water mask in hilly areas (Section 2.4.3).

We make use of several open geospatial and remote sensing datasets to construct an open water map. Section 2 provides an overview of the main input datasets utilized in the study, as well as methods applied or developed to detect water, and to compare the resulting water masks. The main input datasets include (1) images acquired by the NASA Landsat 8 mission Roy et al. [2014]s; (2) a new revision of a nearly-global 30m DEM, measured by the SRTM mission NASA [2016]; and (3) OSM data for the Murray-Darling basin in Australia.

Additional research would be required to perform a detailed comparison of the datasets presented in this paper to the local Australian authoritative datasets, such as Surface Hydrology Crossman, S. & Li [2015], Water Observations from Space Mueller et al. [2015], or 5m Digital Elevation Model (DEM) of Australia Geoscience Australia [2016].

7.2. METHODS AND STUDY LOCATION

The Murray-Darling Basin[1], named after the two main rivers in the basin, is located in semi-arid and arid climate zones, covering 1 059 000 square kilometers, an equivalent of 14% of Australia's total surface area. OSM water features include 10 106 linear and 6 708 aerial hydrographic features representing both natural and manmade water features such as rivers, lakes, and canals (Figure 7.2). The features also include water bodies, which are only partially covered by water during the year. Figure 7.2 shows both hydrographic features extracted from OSM and the (potential) river network obtained from HydroSHEDS Lehner et al. [2008].

Figure 7.1: Murray-Darling junction, Australia. Photo by: Michael Storer

The Murray-Darling Basin receives most of its rainfall from a very small percentage of the Basin; mainly along the southern and eastern parts. The rest of the basin is flat and low-lying, contributing very little or no run-off to the rivers. Relevant for this study, many tributaries within the basin are intermittent streams, with highly variable flows dependent on the wetness of the year.

[1] http://www.mdba.gov.au/about-basin

7.2.1. STUDY SITE: MURRAY-DARLING RIVER BASIN

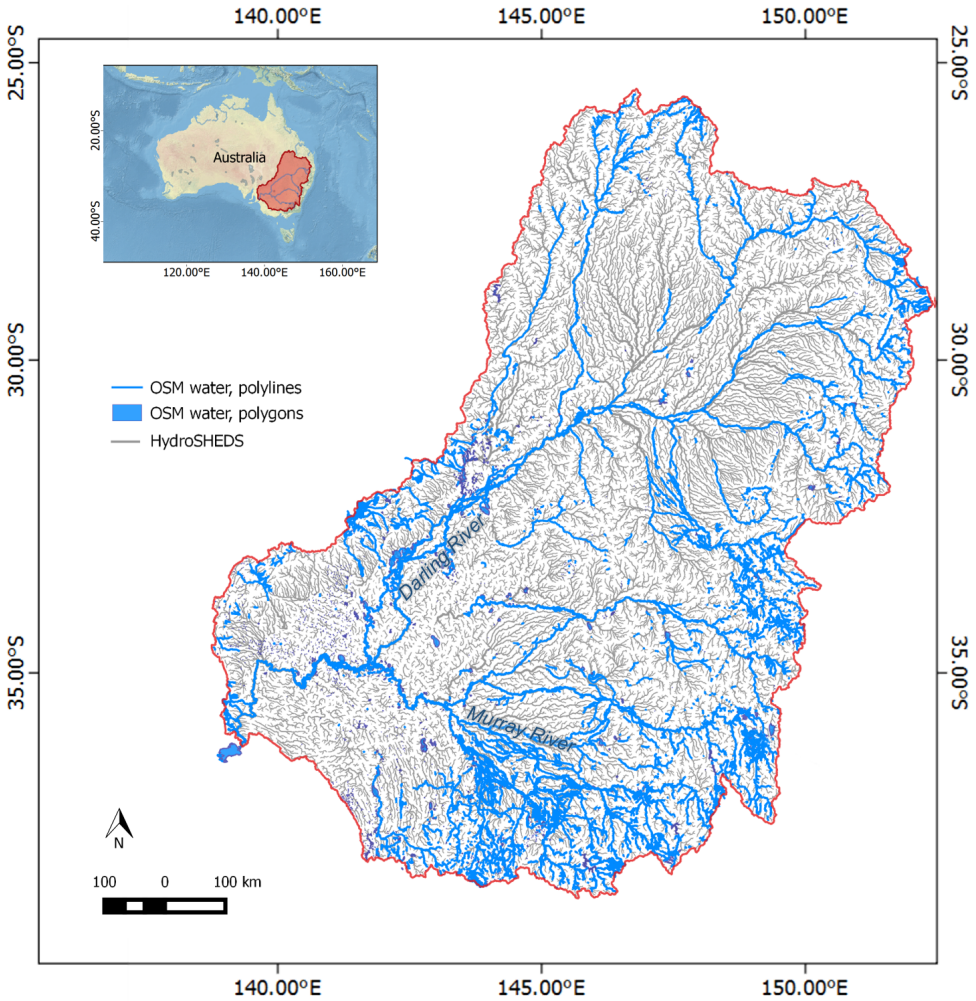


Figure 7.2: Overview of the study area, Murray-Darling River Basin, and vector water mask dataset extracted from OSM. The inset map is based on Natural Earth 1 raster dataset (http://www.naturalearthdata.com/)

7.2.2. INPUT DATASETS USED TO EXTRACT WATER MASK

Even though HydroSHEDS provides a much better coverage, its resolution is limited to 15 arc seconds (450m), which is insufficient to resolve all small water features required for detailed water-related applications. Furthermore, it was based on older, 90m revision of SRTM, which is less detailed compared to a recently released 30m version. Therefore, to develop a method to generate the water mask, we have used four input datasets (see Table 7.1 and Figure 7.3). The first dataset, Landsat 8, combines 2743 scenes of optical multi-spectral satellite imagery acquired during 2013-2015 over the study area. The Landsat 8 mission already operates for more than two years and, therefore, provides a

Table 7.1: Input datasets used for the analysis

Dataset	Type	Resolution	Notes
Landsat 8 TOA	Raster	15m, 30m,60m	2743 scenes were used acquired during 2013-2015, top of atmosphere (TOA) reflectance.
SRTM	Raster	30m	Effective resolution is lower due to the presenceof high-frequency noise
OpenStreetMap	Vector	1m-100m	Planet file from August 2015, the following tagsquery was used to indicate water features:natural=water or natural=spring or waterway=or landuse=basin or landuse=reservoir orbarrier=ditch or landuse=saltpond
HydroBASINS	Vector	450m	Level 8 basins were used to delineate HAND using 30m version of SRTM

reasonably long period of information to estimate water dynamics on a global scale. The second dataset is based on the new 30m revision of the SRTM digital elevation model. The last data source consists of all water features, extracted from a recent version of OSM Planet File OpenStreetMap [2016a]. Additionally, HydroBASINS Lehner and Grill [2013] was used as a supplementary dataset to simplify extraction of a drainage network from SRTM.

Extraction of water features from OSM is a relatively trivial task, mainly consisting of a careful selection of proper filters and data conversion tools. We have used data conversion tools provided by OSM and a set of programs based on GDAL (`http://www.gdal.org/`) to select only relevant features representing surface water. To derive a high-resolution drainage network from SRTM, we have used the D8 method O'Callaghan and Mark [1984], implemented using PCRaster Karssenberg and Schmitz [2010]. The stream delineation step was applied for every -catchment of the Murray-Darling basin. The HydroBASINS level 8 catchment geometries were used for this purpose. This step was required due to the large size of the catchment making it difficult to perform processing in a single step.

Extraction of the water mask using Landsat 8 imagery was the most challenging task and required the development of a new method to derive a water mask from multi-spectral imagery. The new method combines the use of the MNDWI spectral index (7.2) extended with a non-parametric detection of a local threshold to improve the accuracy of water detection. The MNDWI is very similar to the NDWI spectral index (7.1), but uses a short wave infrared band instead of near-infrared. Additional steps include the use of the NDVI Tucker [1979] (7.3) with a high threshold value (0.3) to exclude false water detection in very dark vegetated areas.

$$NDWI = \left(\rho_{green} - \rho_{nir}\right) / \left(\rho_{green} + \rho_{nir}\right) \tag{7.1}$$

$$MNDWI = \left(\rho_{green} - \rho_{swir1}\right) / \left(\rho_{green} + \rho_{swir1}\right) \tag{7.2}$$

$$NDVI = \left(\rho_{nir} - \rho_{red}\right) / \left(\rho_{nir} + \rho_{red}\right) \tag{7.3}$$

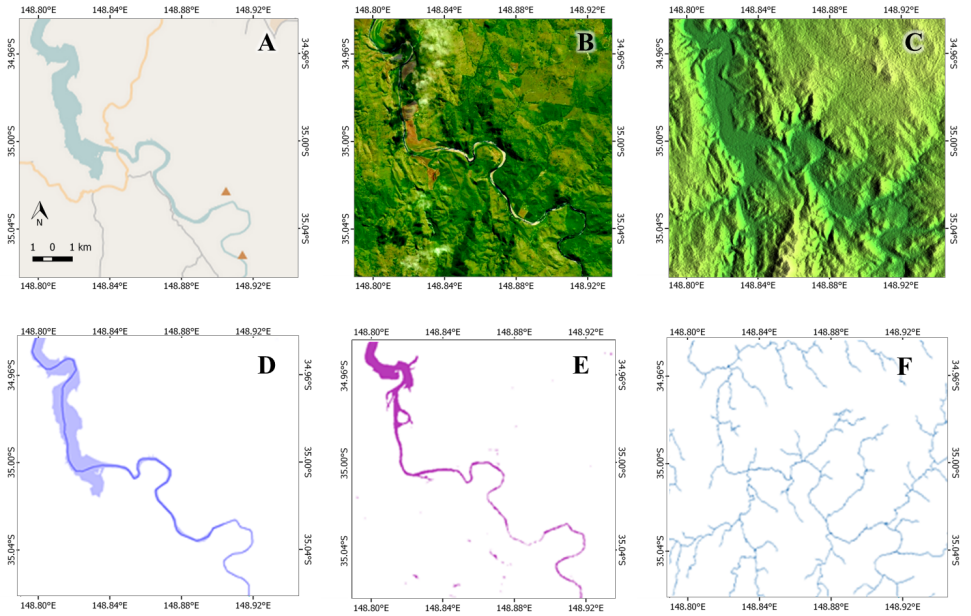


Figure 7.3: The datasets and their corresponding water masks: OSM (A and D), Landsat 8 (B and E) and 30m SRTM (C and F). Murrumbidgee River, north of Canberra.

where ρ_{green}, ρ_{swir1}, ρ_{nir}, and ρ_{red} represent TOA reflectance for corresponding Landsat 8 bands.

We have included an additional classification step to refine the water mask for hilly areas, where the results of automated classification led to high commission errors. We have used a supervised classification based on CART Breiman et al. [1984], which was trained using a manually digitized training set to distinguish between water and land pixels.

Our selection of the input datasets was based on the assumption that the accuracy of all three datasets is similar. However, estimation of the actual errors of OSM would be difficult, mainly because OSM features are usually based on different measurement methods (GPS traces, manual digitizing using medium or high-resolution imagery of bulk import from other databases of varying nature). For Landsat 8 and SRTM the main limitations are well known Roy et al. [2014], Rodriguez et al. [2006]. Horizontal accuracy of Landsat 8 is known to be better than 12m in 90% of the 30m resolution images of the Operational Land Imager (OLI). For SRTM, vertical relative and absolute errors can be explained by its radar nature and are in the order of 10 and 16 meters for 90% of the data, with an absolute geolocation error below 13m.

7.2.3. DERIVATION OF HYDROLOGICAL VARIABLES: DRAINAGE NETWORK AND HAND

In addition to the drainage network, a number of other hydrological parameters were derived during SRTM analysis, such as local drainage direction, flow accumulation, and HAND (see Figure 7.4).

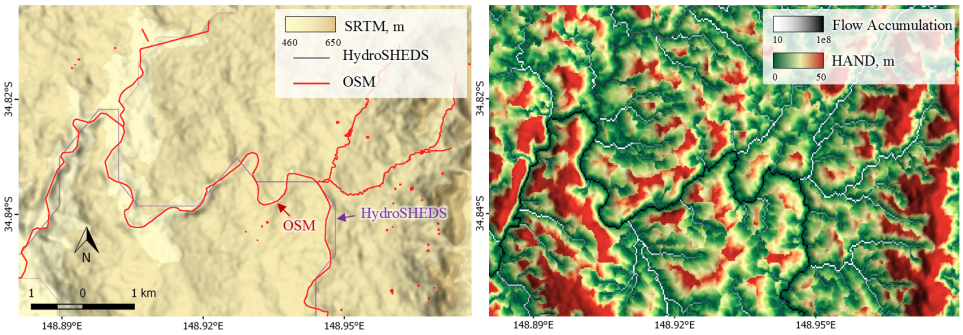


Figure 7.4: Digital elevation model, SRTM 30m (left) and HAND (right).

A (potential) water mask was generated by thresholding HAND values using 1m elevation above the nearest drain. We found this approach to work better in estimating water masks in flat areas compared to the methods based on flow accumulation.

All hydrological variables were derived using the following steps: 1) clip DEM using a Google Earth Engine script and one of 1494 HydroBASIN polygons and download it to Google Compute Engine (GCE); 2) Compute slopes, local drainage directions (LDD, includes pit-removal step), flow accumulations, and HAND. 3) Upload results to Google Earth Engine (`http://earthengine.google.com`) and Google Fusion Table (`http://tables.googlelabs.com/`) so they can be used in parallel processing scripts for further analysis. To compute HAND, a drainage network had to be estimated by thresholding the flow accumulation. We have used a threshold equal to 100 upstream cells, which was sufficient to detect most of the potential rivers.

The resulting HAND can be used for the estimation of potential flood areas, but also to detect pixels where potential errors occur due to hill shadows. These areas were estimated by generating a binary mask based on a variation of HAND values. The mask showing potential hilly areas was computed by marking pixels as hilly in a case where HAND values larger or equal to 30m were found in the 300m radius neighborhood.

7.2.4. METHOD OF WATER DETECTION USING LANDSAT 8

Our method of water detection from multi-spectral multi-temporal imagery is based on a step-wise approach combining unsupervised and supervised classification steps (Figure 7.9). The unsupervised step was applied first to detect the initial water mask using percentile images of reflectance, resulting in minor omission errors in flat areas. However, we obtain very high commission errors in hilly areas due to terrain shadow. Therefore, an additional step was required to refine the water mask using supervised classification. The unsupervised classification step is based on the local adaptive threshold detection method presented in Chapter 3.

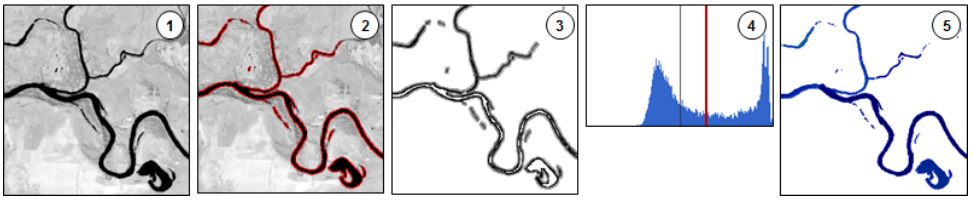


Figure 7.5: Optimal threshold computed around potential surface water edges

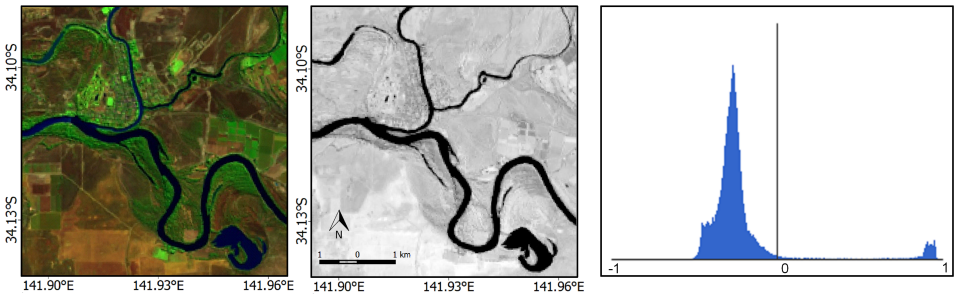


Figure 7.6: False-color intensity percentile composite image (swir1, nir and green) (left), MNDWI index values (middle) and its histogram (right)

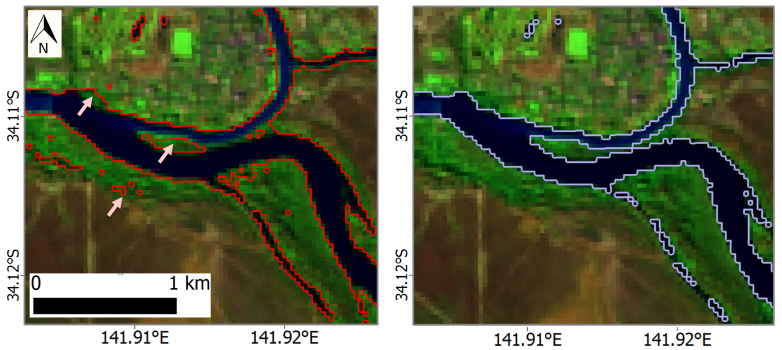


Figure 7.7: Water boundary classified using threshold = 0.0 (left) and threshold = 0.32 automatically detected for this location using new method (right).

The proposed method was applied to the cloud-free percentile images. In a case of two classes in the grid tile, we were able to get an almost perfect detection of water pixels using the following parameters: s=0.7,th=0.99 for the Canny edge filter, and a structuring element with the size 15m x 15m to dilate the edges in step 3 and create a surrounding buffer region. The *s* and *th* parameters are used to define the standard deviation of the Gaussian smoothing kernel and the threshold used to define sensitivity of the filter, respectively.

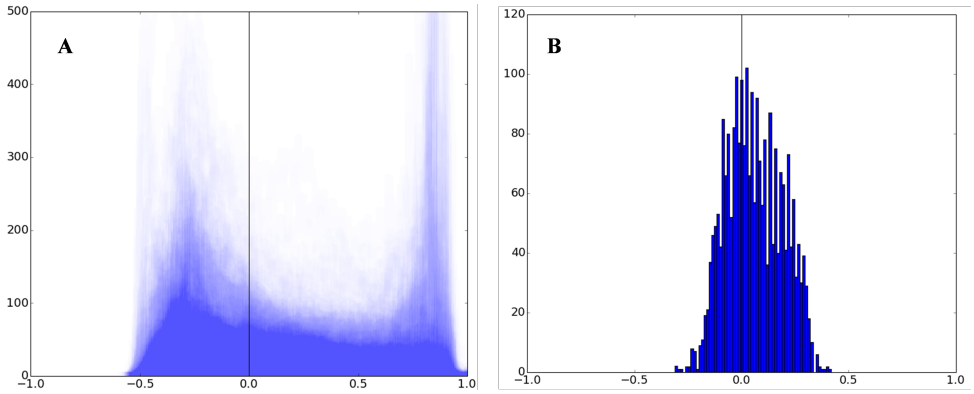


Figure 7.8: Histograms of MNDWI values within 15m of detected edges (A) and a histogram of detected threshold values (B)

7.2.5. REFINING WATER DETECTION USING SUPERVISED CLASSIFICATION BASED ON CART AND HAND

The method was applied over 1725 spatial boxes of 20km x 20km covering the Murray-Darling basin. The 20x20 km area was chosen arbitrarily, and it is assumed that the threshold values are the same within each spatial box area. The resulting MNDWI threshold values have a range of -0.25 to 0.4 (Figure 7.8), which clearly highlights the need for varying the MNDWI threshold spatially.

The supervised classification step, based on CART, was introduced to reduce commission errors found in the case of shadows and snow/ice pixels. It was performed only for hilly areas where misclassified pixels were observed. Hilly areas were detected using a threshold of Hmax, representing the minimum HAND value (30m in our case). In this way, we could keep omission errors low for flat areas and ensure low commission errors for hilly areas. The final error of water detection was very low (less than 1%), mainly due to a presence of mixed pixels or incomplete training data for hilly areas. An additional step was required to exclude very dark vegetated areas, resulting in high MNDWI values and, therefore, misclassified as water. These errors were removed by eliminating pixels with NDVI values greater than 0.3.

In order to compute the final water mask, the study area was divided into regular grid tiles of 0.2 x 0.2 degrees in size. This step was needed to make sure that the dynamic MNDWI threshold is estimated for every tile, but also to parallelize the processing. Finally, the above workflow was applied for every tile.

7.2.6. CLOUD-FREE LANDSAT 8 PERCENTILE IMAGES

Our method to exclude clouds and shadows in the satellite imagery is based on the use of percentile images extracted from an image collection, spanning a two year period (2013-2015), instead of the original images. The percentile images were computed on a per-pixel and per-band basis using 50-130 top of atmosphere (TOA) intensity values. The larger number of images comes from a higher revisit frequency due to an overlap in the satellite swath.

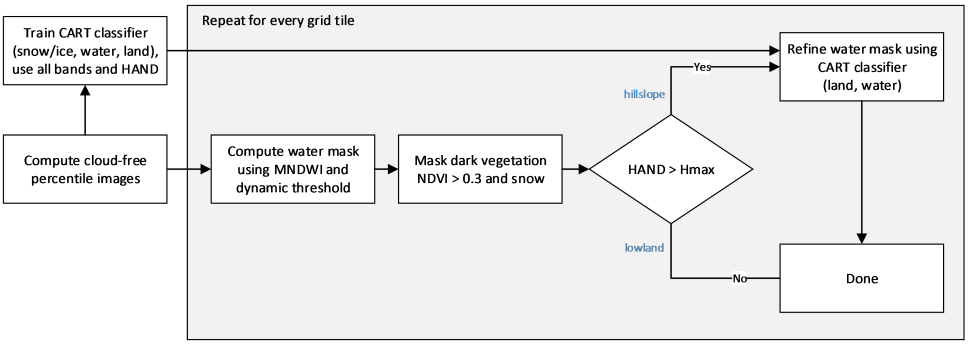


Figure 7.9: Water detection processing pipeline

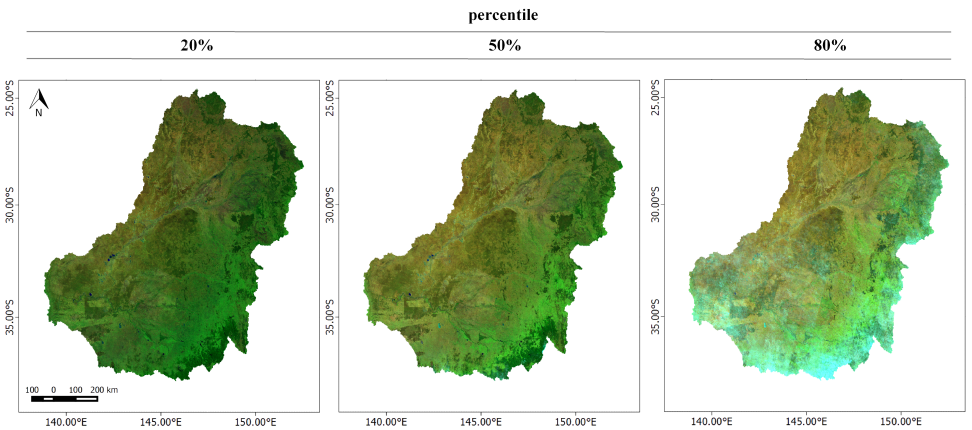


Figure 7.10: Intensity percentile false-color image (swir1, nir, green) based on 2743 Landsat 8 images for 2013-2015, 20% (left), 50% (middle), 80% (right)

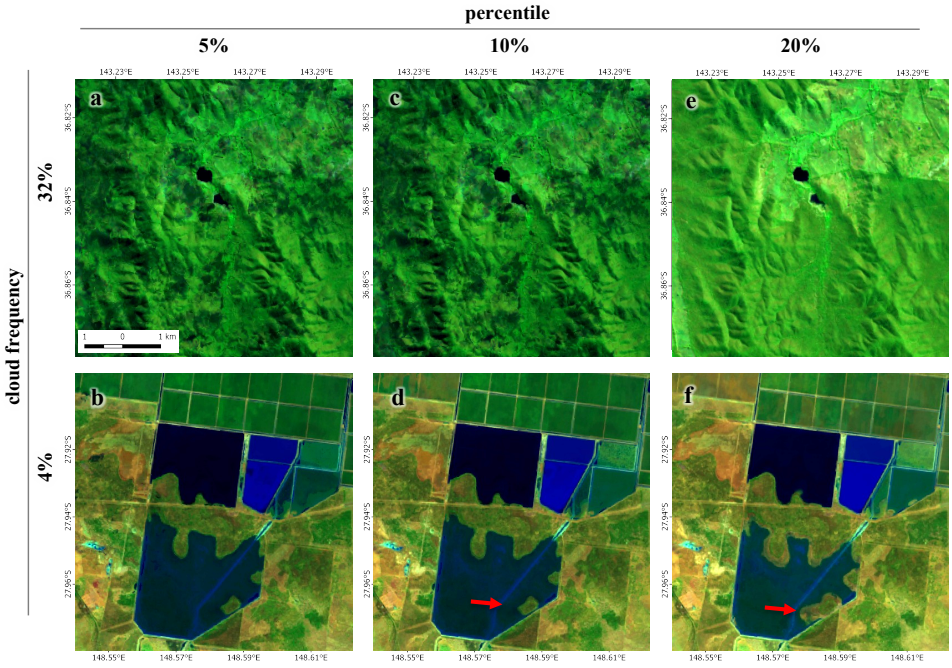


Figure 7.11: Intensity percentile false-color (swir1, nir, green) images. Left to right: 5%, 10%, 20% for areas with average cloud cover 32% (top) and 4% (bottom).

The percentile images (Figure 7.10) appeared to describe the water dynamics in a better way than the interval mean images used in other studies Potapov et al. [2012], Hansen et al. [2013]. Evensthough we have confirmed this result only by visual inspection, the reasoning comes from the fact that the water surface area may change sharply depending on local topographic conditions. This water area change results in sharp changes between water masks present in different percentiles (see Figure 7.11).

Over the Murray-Darling basin, the percentile range of 15%-55% of all TOA intensities was empirically found to be suitable for permanent water detection. However, for semi-arid and flat areas, where cloud frequency is very low, a larger range of percentiles (up to 90%) could be used as well. Average images generated for very low percentiles usually result in too many artifacts present in the images due to cloud and hill shadows, making them difficult to interpret automatically (Figure 7.11) top/left. At the same time, the use of lower percentiles has a higher chance to represent a larger amount of surface water, present during floods and wet seasons and a smaller amount of surface water observed during dry seasons (bottom row with cloud frequency = 4%). For a more detailed analysis the choice of the upper and lower percentiles can be estimated by taking cloud frequency and topographic conditions into account.

The use of simple water spectral indices in hilly areas is usually not sufficient, as they frequently result in very high commission errors due to false detection of water pixels. This misclassification is especially true for MNDWI, which is more sensitive to hill shad-

ows than NDWI. The reason is the spectral signature of hill shadows, which looks similar to the one corresponding to water, resulting in large MNDWI values. To remove these errors, we have trained a CART classifier using a manually created training set, all Landsat 8 bands, and HAND. The training set was created only to include those pixels that were misclassified during the unsupervised step, which appeared to be true only in hilly areas. The HAND values used to train the classifier were included only when they were greater than 10m. This constraint was introduced to reduce the influence of SRTM errors that get higher near water bodies.

The classification was performed using a Google Earth Engine implementation of CART with a tree depth increased from 10 (default value) to 20. The larger tree depth was required to avoid overfitting and because our study basin covers a relatively large area, resulting in a large variation of water and land use types. Overfitting was detected by observing the confusion matrix generated after training the classifier. In the final training set, the confusion error was very close to zero.

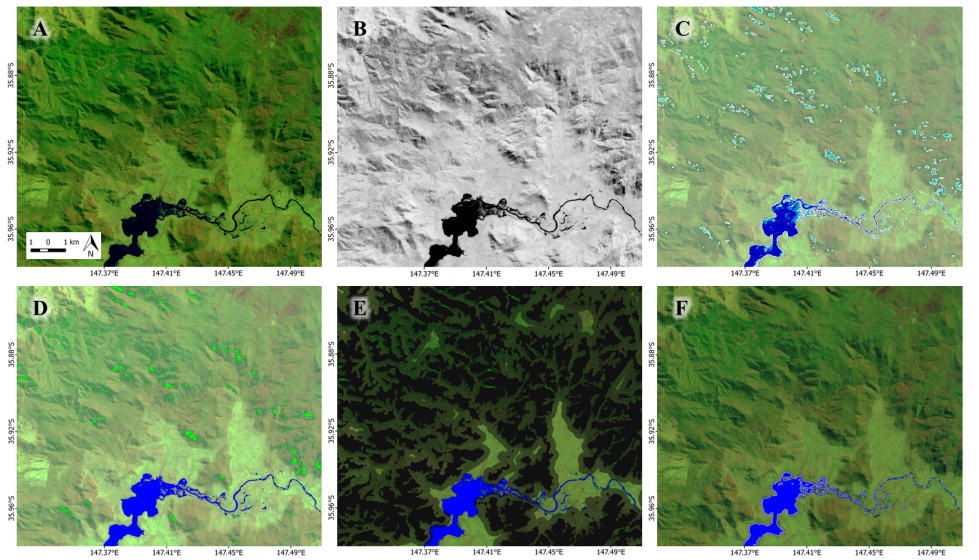


Figure 7.12: Steps performed during supervised classification post-processing step for hillslope areas. (tile 2847). False color 15% percentile composite image (A), MNDWI scaled from (B), water detected using adaptive threshold method (C), results of application of CART classifier (D), 300m buffer areas around HAND > 30m used to select final water mask (E) and the final water mask (F).

The final training set contains about 500 polygons created manually and iteratively by training the classifier for one set of tiles and then validating it for all other tiles where supervised classification was required. This step was required mainly for the grid tiles located in the southern part of the catchment (hilly landscapes).

Since Landsat can reveal water features better than SRTM, we have also decided to analyze the HAND values of all pixels used during the training stage (Figure 7.13). The results reveal that HAND values can take values up to 40m even though visual inspection shows that these pixels belong to water. A low SRTM accuracy can explain this mismatch

near water. Another confusing result comes from a comparison of NDWI and MNDWI. Theoretically, they should be strongly correlated. However, substantial differences were observed in multiple locations. These differences occur because many pixels requiring manual classification are mixed pixels, partially covered by land. Further research would be necessary to explain these results.

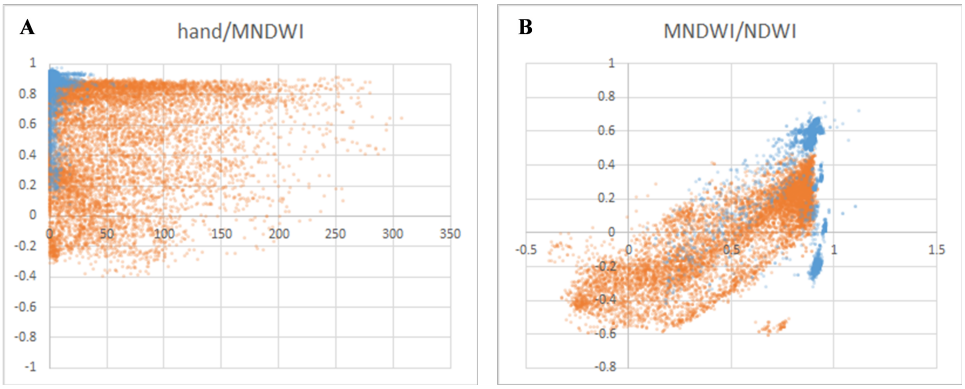


Figure 7.13: Values of HAND vs. MNDWI (A) and MNDWI vs. NDWI (B) used during training of CART classifier. Orange and blue dots represent land and water correspondingly

7.2.7. RIVER CENTERLINE ESTIMATION FROM LANDSAT 8 WATER MASK

An additional skeletonization step was required before computing positional differences between OSM linear water features and water masks detected from Landsat imagery (or HAND). We have used the method of mathematical morphology Serra [1982] by applying an iterative thinning operator applying a hit-or-miss transform to a binary water mask image. The actual steps include:

$$S(W) = \bigcap_{k=0}^{K} (W \otimes kB) \tag{7.4}$$

where $\otimes$ is a binary thinning operator defined as:

$$W \otimes B = W - (W \odot B) \tag{7.5}$$

where $\otimes$ is a hit-or-miss transform operator defined as:

$$W \odot B = (W \ominus B_1) \bigcap (W_c \ominus B_2) \tag{7.6}$$

The structuring elements B_1 and B_2 used for skeletonization (Figure 7.14) allow reconstruction of river skeletons even without the need to introduce pruning (the removal of small branches), which is usually required during such processing. The actual implementation performs a hit-or-miss transform using four rotated versions of the structuring elements applied sequentially during every thinning operation.

W_c denotes a set complement of W, referring to elements not in W.

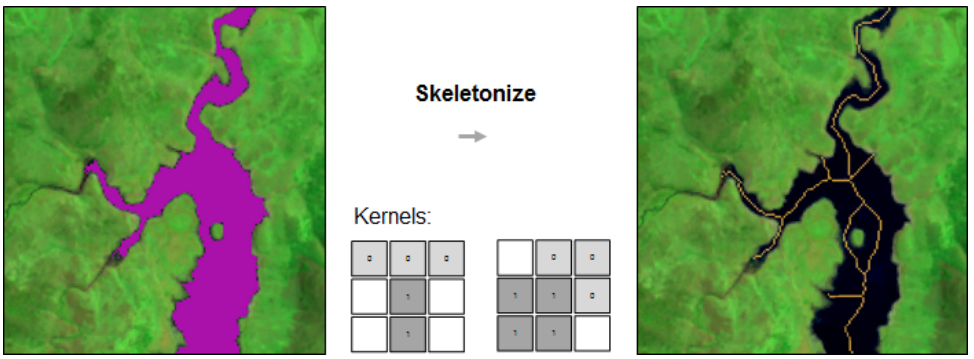


Figure 7.14: Input water mask (left) and results (right) of medial axis detection algorithm using mathematical morphology. Structuring elements (kernels) shown in the middle

Additionally, morphological smoothing was performed before the skeletonization step, to make sure fewer branches were present in the final centerline.

To implement most algorithms, we have used the Google Earth Engine (GEE) parallel computing platform Gorelick [2012]. Some of the computations, where GEE was less suitable, were performed using Google Compute Engine (GCE) running Ubuntu 15.04 (Vivid Vervet). The use of a dedicated machine was more suitable for the computation of hydrological parameters from the DEM, for which no algorithms are available yet in the GEE environment. JavaScript and Python were used as programming languages, as well as some open-source tools and libraries (GDAL GDAL [2017], Fiona Toblerity [2016a], Shapely Toblerity [2016b], ee-runner Donchyts [2016] and PCRaster Karssenberg and Schmitz [2010]). We aimed at automating most of the steps to make sure that they would easily scale to a planetary scale, and that the results of the research can be reproduced when updated versions of the underlying datasets will be released.

7.3. RESULTS

7.3.1. ESTIMATION OF POSITIONAL DIFFERENCES BETWEEN RIVERS

After the centerline is computed, positional differences can be easily computed using the Goodchild's method of increasing overlay polygons.

The above method works best when the lengths of the segments and the distance between segments are significantly larger than the maximum buffer size D_{bmax} used during dilation.

7.4. POSITIONAL DIFFERENCES BETWEEN OPENSTREETMAP, LANDSAT, AND SRTM

The Goodchild's method to estimate positional differences was applied for every line segment of the OSM dataset and two water mask raster datasets: 1) drainage network estimated using HAND and 2) water centerline estimated using Landsat 8 water mask. The final differences were estimated using Score = 0.85 as a threshold, which corresponds to a distance where 85% of the pixels from the second dataset (black line in Figure 7.15) are covered by a dilation via this distance. The frequency and cumulative histograms of the resulting differences are presented in Figure 7.16.

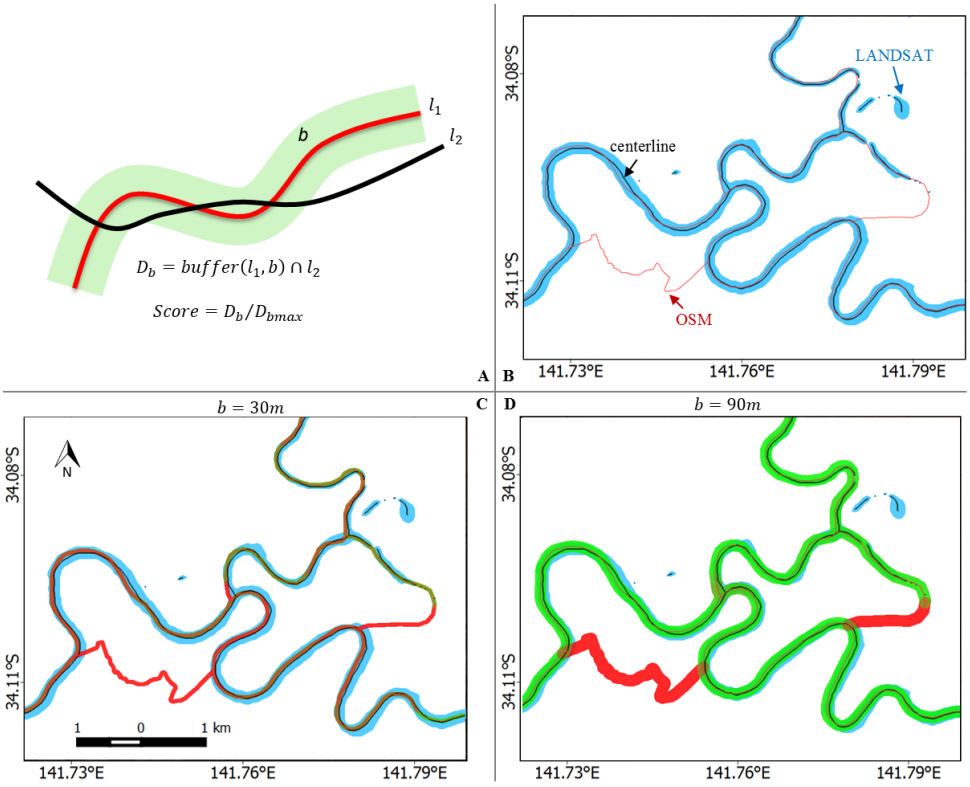


Figure 7.15: Method used to estimate positional differences between vector and raster datasets. Increasing overlay buffer method (A), example of OSM water polyline and Landsat mask (B), results for two different buffer sizes (C, D)

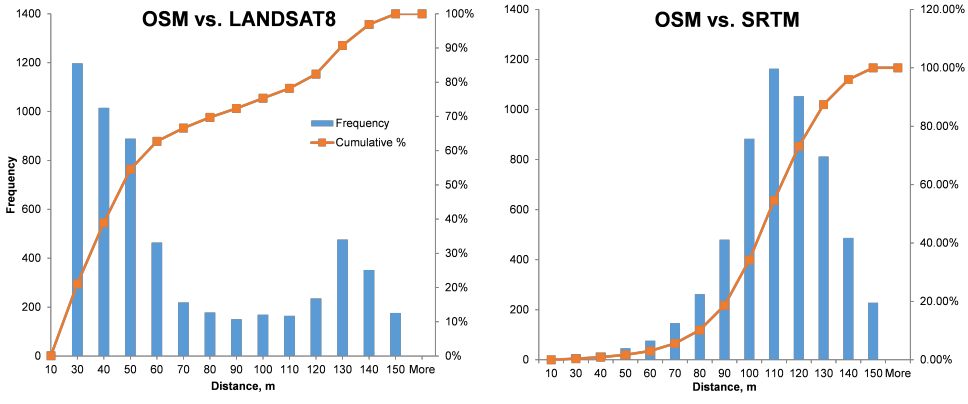


Figure 7.16: Frequency histogram of distances between OSM river segments, Landsat 8 centerline, and SRTM (30m) drainage network.

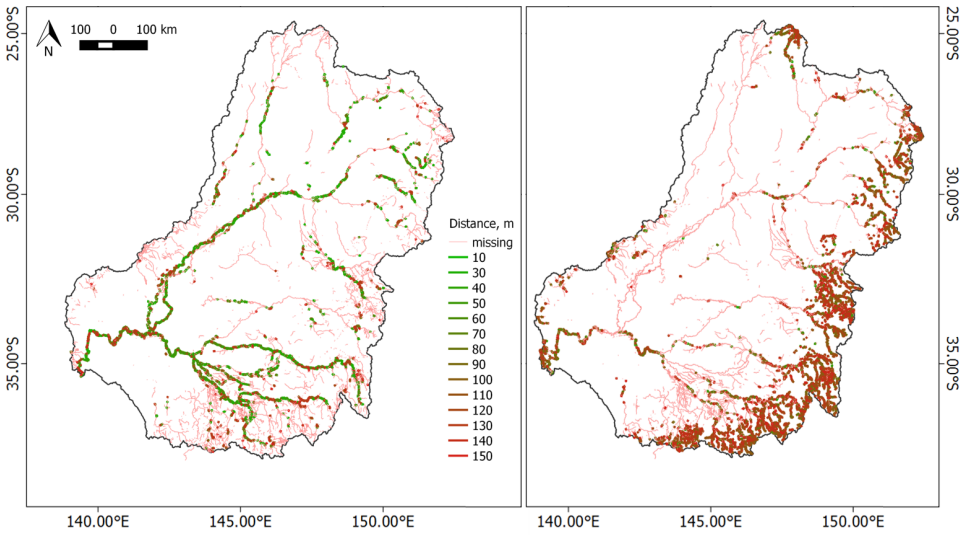


Figure 7.17: Positional differences between linear water features extracted from OSM, water mask centerline extracted from Landsat (left) and drainage network extracted from SRTM (right)

The second peak of the left histogram (distance > 100 m) occurs mainly due to wider water bodies, where the centerline cannot be reached within 20 thinning steps, but also for water bodies that start to overlap within the maximum size dilation, such as oxbows or meandering rivers located close to each other. Positional differences between Landsat centerlines and OSM become larger when a river width becomes larger. These large differences are observed because OSM polylines frequently represent the thalweg (the line of lowest elevation within a river) instead of a medial axis.

It can be seen that the distance between OSM water polylines and the centerlines computed using the Landsat 8 water mask falls in the range of 30-60 meters for 60% of the water features. We were able to confirm this for about 17% (N=5687) of the linear OSM water features. The length of the river segments was selected to be about 0.02 degree (2.2 km).

In the case of SRTM, about 30% (N=9887) of the OSM segments could be compared, mainly located in the southern, hilly part of the catchment (see Figure 7.17). The distances are slightly larger than those for LANDSAT, and also appear normally distributed with a mean value of 110 meters. This may be explained by a systematic shift between the OSM and SRTM datasets, combined with other factors, like the fact that smaller river meanders were not resolved well by SRTM dataset.

Spatial representation of the distances between OSM river segments and centerlines of the LANDSAT water mask, as well as the distances between OSM river segments and drainage network cells, are shown in Figure 7.17. It can be clearly seen that, in the case of Landsat, mainly large rivers could be compared due to the 30m resolution limitation of the OLI sensor. At the same time, a much larger number of the OSM segments could be confirmed using the drainage network derived from 30m SRTM, even though positional differences are not as good as in the case of LANDSAT. On the other hand, the drainage

network derived from SRTM seems to be very inaccurate for flatter landscapes. Therefore, we have excluded drainage network pixels for the cases where HAND values are smaller than 30m in the 300m radius neighborhood. These parameters were identified empirically after careful visual inspection.

7.4.1. GOODNESS OF FIT BETWEEN OPENSTREETMAP AND LANDSAT WATER MASKS

To estimate the overall overlap between water masks extracted from both OSM and Landsat, we have aggregated the overlap using a regular grid for a better understanding of the results. The analysis was performed using the actual water mask (polygonal and linear features of the OSM). For every grid cell, a total surface area of water mask was computed. For linear features, a dilation buffer of 15m was applied using a square kernel to make sure it matches, at least, one Landsat grid cell. Then, the resulting thematic differences were computed (Figure 7.18).

While the centerline analysis reveals a good fit between Landsat and OSM, the surface area analysis demonstrates quite a large mismatch. There can be several explanations for this mismatch. Firstly, OSM (as well as Google Maps) misses a large number of small agricultural reservoirs in this area. Secondly, many of the large reservoirs are intermittent and were mostly dry during 2013-2015. Thirdly, river bank information frequently does not match between two datasets; in some cases it is missing in OSM, in other instances, Landsat misses small rivers (W < 30m) or the water bodies are partially covered by vegetation.

The surface water area of both OSM and Landsat constitutes about 0.85% of the total catchment area. Only about one-third (32%) of the total surface water area can be observed using both datasets. The rest of the surface water can be seen using only Landsat or only OSM (Figure 7.19).

7.5. SUPPLEMENTARY MATERIALS

All scripts used in this chapter, including training set used for water mask refinement in hilly areas, can be found on the following GitHub repository: `http://github.com/gena/paper-osm-2015`. The repository contains all scripts which allow to: 1) extract water features from OSM planet file; 2) download SRTM files clipped by HydroBASIN catchments and perform calculation of HAND using Python version of PCRaster tools; 3) Google Earth Engine JavaScript scripts used to generate regular grid for processing, detect water mask for a given grid tile as well as scripts required to skeletonize that water mask and compute distance between OSM segments and another river raster dataset (drainage network derived from HAND or Landsat water mask centerline). Additional supplementary scripts are available in Jupyter Notebooks. These scripts were used to clean-up HydroBASINS catchments, generate tiled version of HAND from basin-based images, split OSM water geometries into smaller segments to perform local spatial analysis as well as scripts used to generate histograms from GeoJSON files produced by water detection scripts.

7.5.1. GRIDS USED DURING ANALYSIS

The following two grids were used during analysis: 1) HydroBASINS level 8 catchment polygons 2) regular 20x20km grid. The first one was used to parallelize generation of

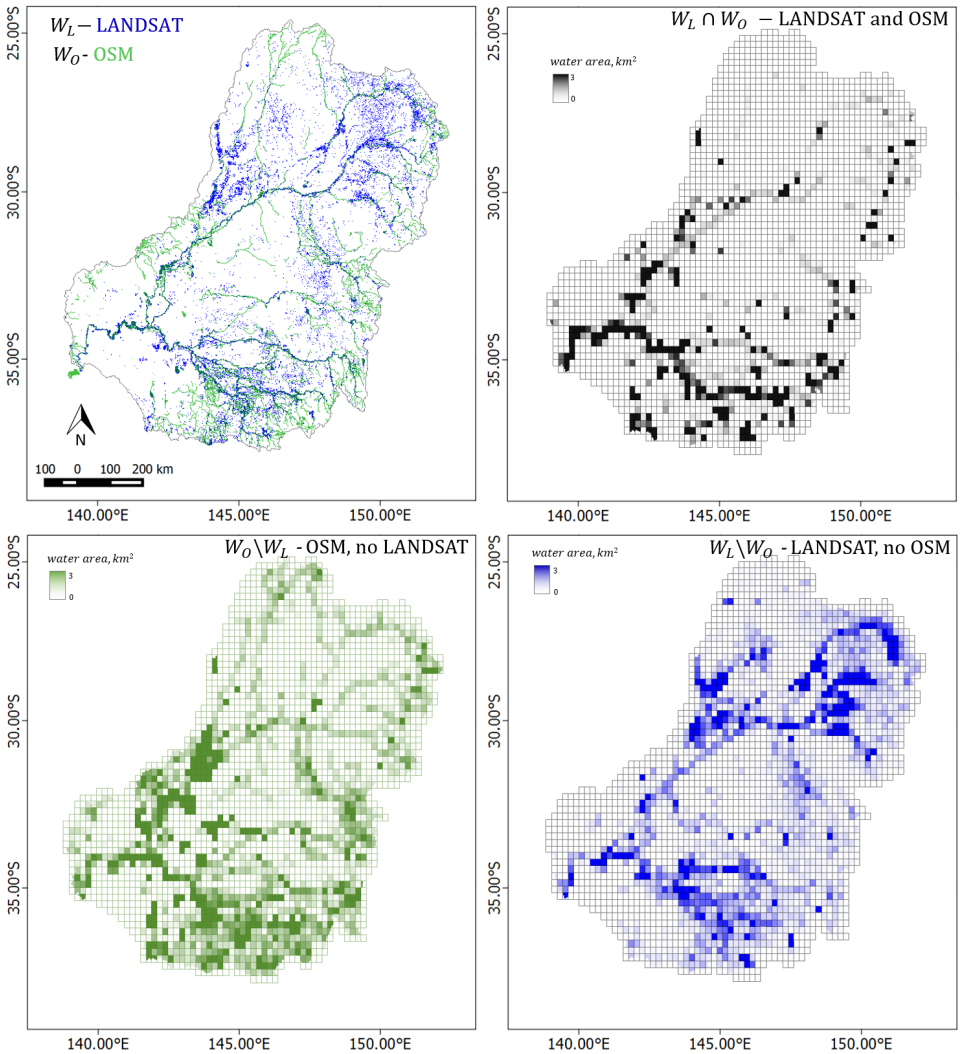


Figure 7.18: Thematic differences between water surface area provided by OSM and detected using Landsat 8

Variable	Area, km²	Ratio, %
$Total\ Water$	8981	100%
W_O	7073	79%
W_L	4799	53%
$W_L \cap W_O$	2891	32%
$W_O \setminus W_L$	4182	47%
$W_L \setminus W_O$	1908	21%

Figure 7.19: Overlap between surface water detected using OSM and Landsat

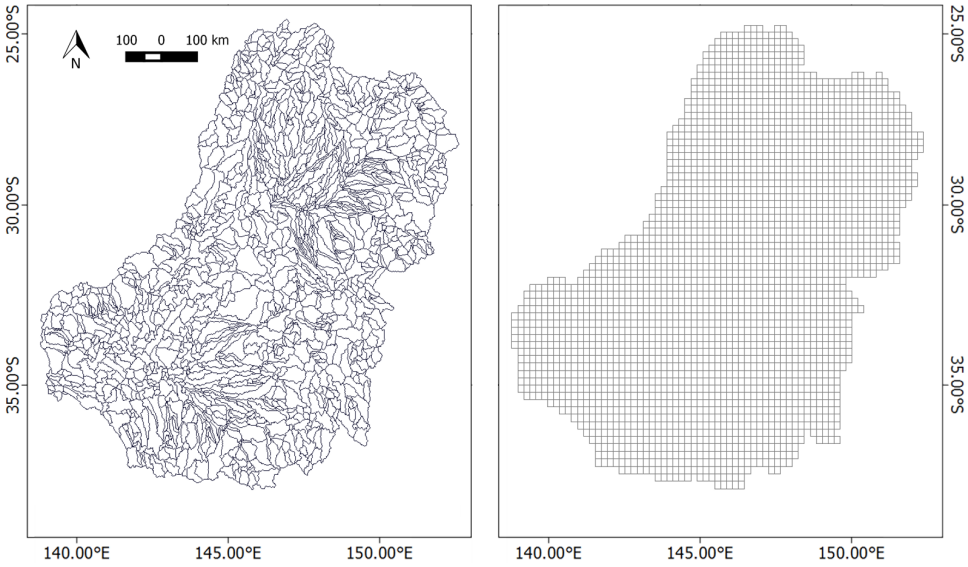


Figure 7.20: HydroBASINS Pfafstetter level 8 catchments (left) and regular 20x20km grid

Table 7.2: List of vector and raster datasets available online

Name	Type	Link
OpenStreetMap water features	Fusion Table	http://bit.ly/paper-osm-2016-fusion-tables
HydroBASIN catchments	Fusion Table	http://bit.ly/paper-osm-2016-fusion-tables
Landsat water mask	EE Asset	users/gena/AU_Murray_Darling/MNDWI_15_water_WGS
HAND	EE Asset	users/gena/AU_Murray_Darling/SRTM_30_Murray_Darling_hand
Local flow accumulation	EE Asset	users/gena/AU_Murray_Darling/SRTM_30_Murray_Darling_flow_accumulation
Distance to the nearest drainage	EE Assets	users/gena/AU_Murray_Darling/SRTM_30_Murray_Darling_dist
Google Earth Engine script	JavaScript	http://bit.ly/paper-osm-2016-gee-assets

hydrological parameters. The latter one was used to perform water detection and to visualize the results in an aggregated form, and to parallelize the water detection analysis using Landsat imagery.

7.5.2. RESULTS AS RASTER AND VECTOR DATASETS

Some of the datasets used during processing were uploaded and shared as Google Fusion Tables and Google Earth Engine Raster Assets, including:

7.6. WEBSITE

Most of the datasets produced by this research can be accessed using a website (`http://osmwater.appspot.com`) dedicated to this study. The website is hosted in a Google Cloud and uses Google App Engine infrastructure, providing integration with Google Earth Engine where most of the datasets produced in this study can be found.

7.7. CONCLUSIONS AND DISCUSSION

The research demonstrates clear benefits of the use of the new imagery acquired by Landsat 8 to detect water bodies when combined with water masks derived from other sources. We have also found the SRTM to be an excellent complimentary dataset enabling improvement of the water mask detection method for hilly areas, after its transformation into HAND.

However, none of the three water masks was found to be perfect regarding positional differences and completeness. The main issue of the water masks derived from Landsat 8 is its shortcomings in detecting small water features such as small rivers or man-made canals and detecting water bodies (partially) covered by riparian or surface water vegetation. The main limitation with the drainage network derived from SRTM is its inability to detect river features for flat terrain conditions. The latter constituted a major part of our study area. An additional challenge is related to the presence of high-frequency noise and a relatively poor quality of SRTM near water bodies. The noise can be explained by the radar origin of the dataset.

One of the next logical steps of the present research could be the development of a data fusion algorithm using the strengths of all three datasets. Such development would require the introduction of objective criteria regarding confidence of every water mask depending on topographic and other conditions. Another step might be to perform the same analysis globally. However, performing global analysis would require significantly larger computational efforts and includes both detection of the water mask and estimation of HAND at 30m resolution. Additional validation of the OSM and its fusion with the datasets produced by the local governmental agencies (Surface Hydrology, Water Observations from Space, local high-resolution elevation models) will help harmonizing existing vector and raster water mask datasets.

Possible improvements to the method of water detection might include utilization of the panchromatic band and entropy-based methods in addition to the spectral methods. Additional significant improvements can be achieved through the use of the other medium or high-resolution satellite missions such as Sentinel 2, PlanetLabs and SkyBox. The use of higher resolution imagery would allow detection of much smaller (width < 30m) river features, resulting in improved coverage. The method of water detection can be easily extended to use Landsat 7 or any other multi-spectral imagery, for example, to generate an inter-annual water mask or to study water dynamics.

The proposed method of water detection might face difficulties in the areas where an insufficient number of cloud-free observations is available, for example, in very wet or cold climates. In this case, it might be difficult to determine a correct range of the cloud-free percentiles to be used for water detection.

8

Conclusions, Discussion and Recommendations

This concluding chapter explains the scientific and technical implications for the academic community and society of the research findings.

Automated surface water detection from satellite imagery remains a challenging task, especially in situations, where only a limited number of satellite images is available. In this cases, the use of multiple data sources, combining imagery from multiple satellite sensors, is feasible. Eventhough this approach may require the use of algorithms that are tuned for every satellite sensor. I'm convinced, that the first step in every study involving multitemporal satellite data should be to generate and visually inspect time lapse videos based on all available satellite images for a given study area. This way, variability of reflectance values can be visually inspected in order to develop an optimal strategy for further algorithm selection and tuning. Mostly, these videos will be polluted with atmospheric noise and all kinds of artifacts, caused by sensor limitations or data processing inconsistencies. For many surface water studies, atmospheric effects can be easily eliminated by producing simple percentile composite images instead of the real ones, as discussed in chapters 5 and 7. Parameterizing algorithms to use typical cloud frequency for a given area to tune this kind of averaging sounds a feasible technique. This way, long-term trends can be easily explored, to help better understand how the algorithms, such as discussed throughout this thesis, can be further tuned to focus on specific natural or human-made processes. Understanding longer-term trends is useful for a more specific tuning of more expensive algorithms, focusing on processing of satellite images on a per-scene basis. Such algorithms are discussed in chapters 3 and 4. However, scaling this kind of algorithms to perform processing globally, at high spatial and temporal resolution, using both passive multispectral and active SAR sensors, can be a tremendous task, facing many technical and methodological limitations.

In this thesis, a number of algorithms to overcome these limitations were discussed, either by reviewing existing methods, as discussed in Chapter 2, or by introducing new algorithms, useful for many surface water detection applications. Some of these algorithms focus on the elimination of effects caused by clouds, cloud shadows, snow/ice or hills, while others provide a way for a more accurate or more resource-efficient detection

of surface water from noisy satellite imagery.

A new discriminative method was developed using two popular image processing algorithms: Canny Edge filter and Otsu thresholding. The method allows very accurate surface water detection. The method was further extended with a generative step, allowing to infer surface water for non-observed parts of the images, either due to limited swath or due to presence of noisy pixels, covered by clouds, snow, ice, very dark shadows, and vegetation. The method has been applied to reconstruct surface water dynamics of a reservoir, where all freely available multi-sensor imagery was used with a spatial resolution higher or equal to 30m. The validation shows excellent performance of the method.

We have discussed the potential of the use of Bayesian methods to infer a set of unobserved or indirectly observed variables from multi-temporal satellite data. This kind of technique is needed when applying the method at global scale. The Bayesian framework introduced in the thesis can help detecting surface water for partially-observed water bodies. The resulting reconstructed water masks perfectly match with the observed water level data. Even more, the satellite-derived water surface water area values allow detection of sensor artifacts, where measurements were not available.

For an automated long-term surface water change detection, we have developed a simple method, based on trend analysis of wetness indices computed for reflectance percentile composites. A very simple yet robust algorithm allows very accurate detection of surface water changes at global scale, provided that sufficient observations are available to ensure continuity in statistical properties of the analyzed image collections. The method was then applied to perform a global surface water change study and to develop a software tool called Aqua Monitor.

To further investigate the applicability of the methods and to compare satellite-derived surface water masks to existing global surface water datasets, surface water of the Murray-Darling Basin in Australia was studied. Here, the local dynamic thresholding method was applied to process Landsat 8 reflectance percentile composite images to detect permanent surface water mask at 30m resolution. The resulting water mask, together with the newly-generated 30m-resolution drainage network derived from SRTM, was compared to the water mask extracted from OSM.

The results reveal that the best surface water mask should be generated by combining all three data sets that were analyzed in the current study; water masks extracted from OSM, optical satellite imagery and the drainage network derived from high-resolution digital elevation models for hilly areas.

A good agreement was found, concerning positional accuracy, between river water features from OSM and water masks derived from Landsat 8. However, only 32% of the total OSM and Landsat 8 water mask matches when analyzing the actual intersecting surface area.

The newly generated Landsat 8 water mask reveals many new water bodies previously not present in OSM or any other vector dataset we have explored. A large part constitutes a large number of small agricultural reservoirs located in the northern and southern parts of the catchment.

8.1. OUTLOOK

The use of all available satellite imagery from multiple passive optical and active radar sensors increases our chances to detect the actual variability of surface water. However, in many cases the amount of information observed from satellites may be insufficient to represent all dynamics. Passive multispectral images may be useless in case of thick clouds, when no land surface can be observed. Therefore, combining information extracted from EO data with the actual local data, including alternative sources such as social media, may be feasible to reconstruct the actual surface water dynamics. This becomes even more relevant for real-time applications, for example, when information on flood extent is needed to better deal with emergency situations.

Furthermore, the number of free EO observations available increases exponentially, challenging remote sensing researchers to develop more effective methods to extract higher level variables from these datasets in an automated way. Effective infrastructures allowing to store, but also to analyze these data will be crucial to process these datasets.

While doing this research, I've been faced with the challenges that the amount of the new datasets and new data science methods to address remote sensing or more general classification questions develop very rapidly. Therefore, a lot of focus was shifted towards development of new methods and exploratory tools instead of derivation of new higher-level datasets. With the introduction of tools such as Google Earth Engine, processing of huge amounts of EO data have become a reality.

8.2. RECOMMENDATIONS

Further development of multi-sensor Bayesian methods is feasible, taking into account changes in the waterbody configuration, either due to natural processes or human activities (meandering rivers, sedimentation of reservoirs, land reclamation and so on). This kind of changes would be possible to incorporate into the Bayesian framework introduced in the Chapter 4 by making density functions (prior) time-dependent. Adding temporal variability may be useful to further increase accuracy and transferability of the method even when no such changes occur to the waterbody, but due to seasonal changes. Further extending of these methods to include temporal correlation between measurements may be needed to eliminate the effect of outliers. This can, for example, be done by incorporating sequence learning methods, such as for example Long short-term memory (LSTM) Hochreiter and Schmidhuber [1997].

One important aspect to be taken into account when detecting surface water form EO data at the planetary scale is that all of the steps need to be automated. This can be achieved by the use of fully automated methods. Prospective directions include the use of unsupervised machine learning methods, such as efficient image segmentation methods like SLIC Achanta et al. [2012] combined with the use of automated inference methods to perform further classification of these clustered images. The use of more suitable computational hardware and software platforms to speed-up data processing Donchyts [2017] is important to allow researchers to explore faster and measure performance of these approaches.

The parallel processing platform Google Earth Engine has opened a new era of EO data processing by allowing researchers around the world to freely process petabytes of satellite data. Built on simple algorithms as components, which are easy to reproduce

locally, the platform allows development of new EO algorithms that can be scaled easily to perform global analysis at high spatio-temporal resolution. The platform solves many technical problems , but also demands new way of working with the data and algorithms.

Surface water detection in hilly areas can be challenging, especially when no recent DEM is available or when its vertical accuracy is low. In this case, the use of topographic indices such as HAND Nobre et al. [2011] may help to eliminate false-positive surface water pixels or to identify areas where additional processing may be required. For single-image processing in topographically complex areas the use of topographic correction methods may be feasible, as discussed in the Chapter 2. However, applicability of these methods to study long-term changes at large spatial extents may be too resource-demanding.

For better applicability of statistical models, the incorporation of cloud frequency as one of the variables used for classification methods can help in implementing automated methods of surface water detection. Knowing how cloud frequency varies will allow to better adjust methods such as introduced in the Chapter 5. Temporal variability of the cloud frequency can help to better parameterize the statistical algorithms to generate the range of reflectance percentile composites to study variability of the land surface processes.

In addition to auxiliary variables such as cloud frequency, the use of aggregated, upscaled representations of EO datasets may be useful to help performing inference faster. In classical geospatial applications, the use of coarser representations of raster data (pyramids) is a common technique to reduce time required for data processing at coarser scale. However, the same approach can be used to generate aggregated representations for EO datasets. This should help to perform image classification much faster when compared to existing methods. Eventhough this is more a technical question, it can be frequently crucial to perform some data operations very fast, allowing researchers to focus on more fundamental questions.

In the present thesis I have investigated applicability of the new methods to freely available EO data. But nowadays, many commercial datasets, with much higher resolution (spatial, temporal or spectral) can be acquired. Therefore, one of the future directions can be to investigate the optimal type of dataset required for every time of the study. In many cases, the use of free EO data may be sufficient to answer many questions. But for other studies, the resolution may be insufficient to represent variability of the observed natural or human-made processes. In some cases, a hybrid approach may be feasible, where a limited amount of high-resolution data can be used to improve predictive quality of the algorithms. Finding an optimal strategy depends strongly on the goal of the study.

For some natural processes, mixing of EO-derived variables with the best physically-based numerical models can be useful to restore the sequence of surface water changes. One of the challenges in this case will be a compromise between computational and storage resources required to use numerical models. In some cases, the use of modern machine learning techniques may help to get better insights.

ACKNOWLEDGEMENTS

THis journey would be impossible without the support of my family, professors and mentors, and friends.

I would like to thank Delft University of Technology, Deltares, the FP7 European Commission project earth2Observe, and the SERVIR-Mekong project for funding the research presented in this thesis. I would like to express my most profound gratitude to my supervisors Nick van de Giesen, Elmar Eiseman and Bert Jagers for guiding me throughout this research project. Nick, I could not have imagined having a better advisor and mentor for my Ph.D. study. I am grateful to my colleagues Hessel Winsemius, Jaap Schellekens, Fedor Baart, Jaap Kwadijk, Joost Icke, Gerard van de Kolff, Arthur Baart for their patience and support in overcoming numerous obstacles I have been facing through my research. I am also grateful to Google for providing Cloud Compute Services credits which helped me to get results faster and of a better quality. Working with the Google Earth Engine community and software engineers was a great pleasure. Noel Gorelick, Nicolas Clinton, Tyler Ericson, Simon Ilyushchenko, Matt Hancher, George Azzari, Guido Lemoine, Sam Murphy and many other active participants of the Google Earth Engine Developers Forum - I've learned a lot from you during sleepless nights while working on this research. I like to express my special gratitude to Dr. Mark Zheleznyak who gave the start to my scientific career in Ukraine. To my wife Olga Donchyts and son Daniel Alexandr Donchyts, thank you for supporting and inspiring me in all of my pursuits.

A

SURFACE WATER DETECTION SOURCE CODE

A

A.1. LOCAL OTSU THRESHOLDING USING CANNY EDGE FILTER

```
/***
 * Computes the threshold that maximizes interclass variance.
 */

var otsu = function otsu(histogram) {
    histogram = ee.Dictionary(histogram);

    var counts = ee.Array(histogram.get('histogram'));
    var means = ee.Array(histogram.get('bucketMeans'));
    var size = means.length().get([0]);
    var total = counts.reduce(ee.Reducer.sum(), [0]).get([0]);
    var sum = means.multiply(counts).reduce(ee.Reducer.sum(), [0]).get([0]);
    var mean = sum.divide(total);

    var indices = ee.List.sequence(1, size);

    // Compute between sum of squares, where each mean partitions the data.
    var bss = indices.map(function (i) {
        var aCounts = counts.slice(0, 0, i);
        var aCount = aCounts.reduce(ee.Reducer.sum(), [0]).get([0]);
        var aMeans = means.slice(0, 0, i);
        var aMean = aMeans.multiply(aCounts).reduce(ee.Reducer.sum(), [0]).get([0]).divide(aCount);
        var bCount = total.subtract(aCount);
        var bMean = sum.subtract(aCount.multiply(aMean)).divide(bCount);
        return aCount.multiply(aMean.subtract(mean).pow(2)).add(bCount.multiply(bMean.subtract(mean).pow(2)));
    });

    // Return the mean value corresponding to the maximum BSS.
    return means.sort(bss).get([-1]);
};

/***
 * Compute local threshold assuming bimodal distribution.
 */

function computeThresholdUsingOtsu(image, scale, bounds, th, g, skipShort, weightGradient, minValue) {
    // clip image edges
    var mask = image.mask().gt(0).focal_min(ee.Number(scale).multiply(3), 'circle', 'meters');

    // detect sharp changes
    var edge = ee.Algorithms.CannyEdgeDetector(image, th, g);
    edge = edge.multiply(mask);

    // take the largest changes, estimate gradient around edge and use that as a weight
    if (weightGradient) {
        var gradient = image.gradient().abs();
        var edgeGradient = gradient.select(0).max(gradient.select(1)).mask(edge.gt(th)).reproject(image.projection().scale(2, 2));

        // take the upper percentiles only
        var mode = ee.Number(ee.Dictionary(edgeGradient.reduceRegion(ee.Reducer.mode(), bounds, scale)).values().get(0));
        var _buckets = 50;

        var significantEdgesMask = ee.Algorithms.If(
            ee.Algorithms.IsEqual(mode, null),
            edge.mask(),
            edgeGradient.gt(mode)
        );

        edge = ee.Image(ee.Algorithms.If(
            ee.Algorithms.IsEqual(mode, null),
            edge,
            edge.updateMask(significantEdgesMask))
        );

        if (debug) {
            // gradient around edges
            if (edgeGradient) {
                print(ui.Chart.image.histogram(edgeGradient, bounds, scale, _buckets));
                Map.addLayer(edgeGradient, {}, 'edge gradient', false);
                Map.addLayer(significantEdgesMask.mask(significantEdgesMask), {}, 'significant edges', false);

                print('Mode: ', mode);
            }
        }
    }

    // advanced, detect edge lengths
    var coonnectedVis = void 0;
    if (skipShort) {
        var connected = edge.mask(edge).lt(0.8).connectedPixelCount(50, true);
        var edgeLong = connected.gte(50);
        edge = edgeLong;
        coonnectedVis = connected.updateMask(edgeLong).visualize({ palette: ['ffffff', 'ff0000'], min: 0, max: 50 });
    }

    // buffer around NDWI edges
    var edgeBuffer = edge.focal_max(ee.Number(scale).multiply(1), 'square', 'meters');
```

```
    var imageEdge = image.mask(edgeBuffer);

    // compute threshold using Otsu thresholding
    var buckets = 100;
    var hist = ee.Dictionary(ee.Dictionary(imageEdge.reduceRegion(ee.Reducer.histogram(buckets), bounds, scale)).values().get(0));

    var threshold = ee.Algorithms.If(hist.contains('bucketMeans'), otsu(hist), 0);
    threshold = ee.Number(threshold); //.add(0.05)

    if (debug) {
        Map.addLayer(edge.mask(edge), { palette: ['ff0000'] }, 'edges', false);

        if (skipShort) {
            Map.addLayer(coonnectedVis, {}, 'edges (connected)', false);
        }

        print('Threshold: ', threshold);

        print(ui.Chart.image.histogram(image, bounds, scale, buckets));
        print(ui.Chart.image.histogram(imageEdge, bounds, scale, buckets));
        Map.addLayer(mask.mask(mask), { palette: ['000000'] }, 'image_mask', false);
    }

    return {
      threshold: minValue ? threshold.max(minValue) : threshold,
      edge: edge
    }
}
```

A

A.2. PERONA-MALIK FILTER: ANISOTROPIC DIFFUSION

```
/***
 * Solves anisotropic diffusion equation using CFD:
 *
 *   I(n+1, i, j) = I(n, i, j) + lambda * (cN * dN(I) + cS * dS(I) + cE * dE(I) + cW * dW(I))
 *
 */

var peronaMalikFilter = function(I, iter, K, method) {
    var dxW = ee.Kernel.fixed(3, 3,
                              [[ 0,  0,  0],
                               [ 1, -1,  0],
                               [ 0,  0,  0]]);

  var dxE = ee.Kernel.fixed(3, 3,
                              [[ 0,  0,  0],
                               [ 0, -1,  1],
                               [ 0,  0,  0]]);

  var dyN = ee.Kernel.fixed(3, 3,
                              [[ 0,  1,  0],
                               [ 0, -1,  0],
                               [ 0,  0,  0]]);

  var dyS = ee.Kernel.fixed(3, 3,
                              [[ 0,  0,  0],
                               [ 0, -1,  0],
                               [ 0,  1,  0]]);

  var lambda = 0.2;

  if(method == 1) {
    var k1 = ee.Image(-1.0/K);

    for(var i = 0; i < iter; i++) {
      var dI_W = I.convolve(dxW)
      var dI_E = I.convolve(dxE)
      var dI_N = I.convolve(dyN)
      var dI_S = I.convolve(dyS)

      var cW = dI_W.multiply(dI_W).multiply(k1).exp();
      var cE = dI_E.multiply(dI_E).multiply(k1).exp();
      var cN = dI_N.multiply(dI_N).multiply(k1).exp();
      var cS = dI_S.multiply(dI_S).multiply(k1).exp();

      I = I.add(ee.Image(lambda).multiply(cN.multiply(dI_N).add(cS.multiply(dI_S))
            .add(cE.multiply(dI_E)).add(cW.multiply(dI_W))))
    }
  }
  else if(method == 2) {
    var k2 = ee.Image(K).multiply(ee.Image(K));

    for(var i = 0; i < iter; i++) {
      var dI_W = I.convolve(dxW)
      var dI_E = I.convolve(dxE)
      var dI_N = I.convolve(dyN)
      var dI_S = I.convolve(dyS)

      var cW = ee.Image(1.0).divide(ee.Image(1.0).add(dI_W.multiply(dI_W).divide(k2)));
      var cE = ee.Image(1.0).divide(ee.Image(1.0).add(dI_E.multiply(dI_E).divide(k2)));
      var cN = ee.Image(1.0).divide(ee.Image(1.0).add(dI_N.multiply(dI_N).divide(k2)));
      var cS = ee.Image(1.0).divide(ee.Image(1.0).add(dI_S.multiply(dI_S).divide(k2)));

      I = I.add(ee.Image(lambda).multiply(cN.multiply(dI_N).add(cS.multiply(dI_S))
            .add(cE.multiply(dI_E)).add(cW.multiply(dI_W))))
    }
  }

  return I;
}
```

LIST OF PUBLICATIONS

LIST OF JOURNAL PUBLICATIONS

3. **Donchyts, G.**, Schellekens, J., Winsemius, H., Eisemann, E., van de Giesen, N. (2016). A 30 m Resolution Surface Water Mask Including Estimation of Positional and Thematic Differences Using Landsat 8, SRTM and OpenStreetMap: A Case Study in the Murray-Darling Basin, Australia. Remote Sensing, 8(5), 386.

2. **Donchyts, G.**, Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., van de Giesen, N. (2016). Earth's surface water change over the past 30 years. Nature Climate Change, 6(9), 810-813.

1. Luijendijk, A., Hagenaars, G., Baart, F., **Donchyts, G.**, Aarninkhof, S., Ranasinghe, F. (2017). Global sandy shorelines changes derived from satellite images, Nature (submitted)

LIST OF CONFERENCES

6. **Donchyts, G.**, van de Giesen, N., Gorelick, N. (2017). Reconstruction of reservoir and lake surface area dynamics from optical and SAR satellite imagery. 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Bruges, Belgium.

5. **Donchyts, G.**, Baart, F., Gorelick, N., Eisemann, E., van de Giesen, N. (2017). Fast segmentation of satellite images using SLIC, WebGL and Google Earth Engine. EGU General Assembly Conference Abstracts, 19, p.18283.

4. **Donchyts, G.**, Winsemius, H.C.; Schellekens, J.; Erickson, T.; Gao, H.; Savenije, H.; van de Giesen, N. (2016). Global 30m Height Above the Nearest Drainage. EGU General Assembly Conference Abstracts, 18 p.17445.

3. **Donchyts, G.**, Winsemius, H.C.; Schellekens, J.; Erickson, T.; Gao, H.; Savenije, H.; van de Giesen, N. (2016). Global 30m Height Above the Nearest Drainage. EGU General Assembly Conference Abstracts, p.17445.

2. **Donchyts, G.**, Jagers, B.; van de Giesen, N.; Baart, F; van Dam, A. (2015). Automated estimation of river bathymetry using change detection based on Landsat imagery and river morphological models. Proceedings of the AGU fall meeting.

1. **Donchyts, G.**, Schellekens, J.; Winsemius, H.C.; van de Giesen, N. (2015). Quality assessment of OpenStreetMap water mask using LANDSAT 8 imagery and Google Earth Engine. EGU General Assembly Conference Abstracts, 17, p.12021.

BIBLIOGRAPHY

R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

D. Amitrano, G. D. Martino, A. Iodice, F. Mitidieri, M. N. Papa, D. Riccio, and G. Ruello. Sentinel-1 for monitoring reservoirs: a performance analysis. *Remote Sensing*, 6(11): 10676–10693, 2014.

A. Angal, N. Mishra, X. Xiong, and D. Helder. Cross-calibration of landsat 5 tm and landsat 8 oli with aqua modis using pics. In *SPIE Optical Engineering+ Applications*, pages 92180K–92180K. International Society for Optics and Photonics, 2014.

F. O. Annor, N. Van De Giesen, J. Liebe, P. Van de Zaag, A. Tilmant, and S. Odai. Delineation of small reservoirs using radar imagery in a semi-arid environment: A case study in the upper east region of ghana. *Physics and Chemistry of the Earth, Parts A/B/C*, 34(4):309–315, 2009.

T. P. Barnett and D. W. Pierce. When will Lake Mead go dry? *Water Resources Research*, 44 (3):n/a–n/a, mar 2008. ISSN 00431397. doi: 10.1029/2007WR006704.

C. Barron, P. Neis, and A. Zipf. A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 18(6):877–895, 2014. ISSN 13611682. doi: 10.1111/tgis.12073. URL http://doi.wiley.com/10.1111/tgis.12073.

V. R. Barros, C. B. Field, D. J. Dokke, M. D. Mastrandrea, K. J. Mach, T. E. Bilir, M. Chatterjee, K. L. Ebi, Y. O. Estrada, R. C. Genova, B. Girma, E. S. Kissel, A. N. Levy, S. MacCracken, P. R. Mastrandrea, and L. L. White. Climate change 2014: impacts, adaptation, and vulnerability. Part B: regional aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, 2015.

C. Bemmelen, M. Mann, M. Ridder, M. Rutten, and N. Giesen. Determining water reservoir characteristics with global elevation data. *Geophysical Research Letters*, 2016.

A. Berk, L. S. Bernstein, and D. C. Robertson. Modtran: A moderate resolution model for lowtran. Technical report, DTIC Document, 1987.

M. F. Bierkens, V. A. Bell, P. Burek, N. Chaney, L. Condon, C. H. David, A. de Roo, P. Döll, N. Drost, J. S. Famiglietti, et al. Hyper-resolution global hydrological modelling: what is next. *Hydrol. Process*, 29:310–320, 2015.

C. Birkett, C. Reynolds, B. Beckley, and B. Doorn. From research to operations: the usda global reservoir and lake monitor. In *Coastal altimetry*, pages 19–50. Springer, 2011.

A

T. Blaschke. Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16, 2010.

L. Breiman, J. Friedman, C. Stone, and R. Olshen. *Classification and regression trees*. Wadsworth Statistics/Probability, 1st edition, 1984. URL https://scholar.google.nl/scholar?hl=en{&}q=breiman+classification+regression{&}btnG={&}as{_}sdt=1{%}252C5{&}as{_}sdtp={#}0.

J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

M. Carroll, J. R. Townshend, C. M. DiMiceli, P. Noojipady, and R. Sohlberg. A new global raster water mask at 250 m resolution. *International Journal of Digital Earth*, 2(4): 291–308, 2009.

G. Chander, B. L. Markham, and D. L. Helder. Summary of current radiometric calibration coefficients for landsat mss, tm, etm+, and eo-1 ali sensors. *Remote sensing of environment*, 113(5):893–903, 2009.

P. S. Chavez. Image-based atmospheric corrections-revisited and improved. *Photogrammetric engineering and remote sensing*, 62(9):1025–1035, 1996.

Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6):2094–2107, 2014.

Y. Chen, W. Li, H. Deng, G. Fang, and Z. Li. Changes in central asia's water tower: Past, present and future. *Scientific reports*, 6, 2016.

W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368):829–836, 1979.

J.-F. Cretaux, A. Kouraev, M. Berge-Nguyen, A. Cazenave, and F. Papa. Satellite altimetry for monitoring lake level changes. *Transboundary Water Resources: Strategies for Regional Security and Ecological Stability*, pages 141–146, 2005.

O. Crossman, S. & Li. Surface Hydrology Lines (Regional). 2 ed., Australia, 2015. URL http://www.ga.gov.au/metadata-gateway/metadata/record/83130.

Z. Dai, J. T. Liu, W. Wei, and J. Chen. Detection of the Three Gorges Dam influence on the Changjiang (Yangtze River) submerged delta. *Scientific reports*, 4:6600, jan 2014. ISSN 2045-2322. doi: 10.1038/srep06600.

C. Davidson. Dubai: foreclosure of a dream. *Middle East report.*, 251(summer):8–13, 2009. ISSN 08992851 (ISSN). URL http://dro.dur.ac.uk/7099/.

G. Donchyts. ee-runner GitHub web page, 2016. URL https://github.com/gena/ee-runner.

G. Donchyts. Fast segmentation of satellite images using SLIC, WebGL and Google Earth Engine. volume 19, page 18283, 2017. URL https://www.researchgate.net/publication/316608607_Segmentation_of_satellite_images_using_SLIC_and_Google_Earth_Engine.

G. Donchyts, F. Baart, H. Winsemius, N. Gorelick, J. Kwadijk, and N. van de Giesen. Earth's surface water change over the past 30 years. *Nature Climate Change*, 6(9):810–813, 2016a.

G. Donchyts, J. Schellekens, H. Winsemius, E. Eisemann, and N. van de Giesen. A 30 m resolution surface water mask including estimation of positional and thematic differences using landsat 8, srtm and openstreetmap: A case study in the murray-darling basin, australia. *Remote Sensing*, 8(5):386, 2016b.

G. Donchyts, H. Winsemius, J. Schellekens, T. Erickson, H. Gao, H. Savenije, and N. van de Giesen. Global 30m height above the nearest drainage. *EGU General Assembly*, 18, 2016c.

M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote Sensing of Environment*, 120:25–36, 2012.

Z. Duan and W. Bastiaanssen. Estimating water volume variations in lakes and reservoirs from four operational satellite altimetry databases and satellite imagery data. *Remote Sensing of Environment*, 134:403–416, 2013.

D. Eilander, F. O. Annor, L. Iannini, and N. van de Giesen. Remotely sensed monitoring of small reservoir dynamics: A Bayesian approach. *Remote Sensing*, 6(2):1191–1210, 2014. ISSN 20724292. doi: 10.3390/rs6021191.

O. Elmi, M. J. Tourian, and N. Sneeuw. Dynamic river masks from multi-temporal satellite imagery: An automatic algorithm using graph cuts optimization. *Remote Sensing*, 8(12):1005, 2016.

R. E. Emerton, E. M. Stephens, F. Pappenberger, T. C. Pagano, A. H. Weerts, A. W. Wood, P. Salamon, J. D. Brown, N. Hjerdt, C. Donnelly, et al. Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews: Water*, 2016.

FAO. Water monitoring: Mapping existing global systems and initiatives. Technical report, Prepared by FAO on behalf of the UN-Water Task Force on Monitoring, 2006.

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

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

A

G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud. Automated water extraction index: A new technique for surface water mapping using landsat imagery. *Remote Sensing of Environment*, 140:23–35, 2014.

A. Fisher, N. Flood, and T. Danaher. Comparing landsat water index methods for automated water classification in eastern australia. *Remote Sensing of Environment*, 175: 167–182, 2016.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

N. Flood. Continuity of reflectance data between landsat-7 etm+ and landsat-8 oli, for both top-of-atmosphere and surface reflectance: A study in the australian landscape. *Remote Sensing*, 6(9):7952–7970, 2014.

P. S. Frazier, K. J. Page, et al. Water body detection and delineation with landsat tm data. *Photogrammetric engineering and remote sensing*, 66(12):1461–1468, 2000.

B. C. Gao. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3):257–266, 1996. ISSN 00344257. doi: 10.1016/S0034-4257(96)00067-3.

B.-C. Gao, C. Davis, and A. Goetz. A review of atmospheric correction techniques for hyperspectral remote sensing of land surfaces and ocean color. In *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, pages 1979–1981. IEEE, 2006.

Y. Gao and W. Zhang. A simple empirical topographic correction method for etm+ imagery. *International Journal of Remote Sensing*, 30(9):2259–2275, 2009.

L. Garcia, J. D. Rodriguez, M. Wijnen, and I. Pakulski. *Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector*. The World Bank, Washington D. C., apr 2016. ISBN 978-1-4648-0475-5. doi: 10.1596/978-1-4648-0475-5.

A. Garzelli. A review of image fusion algorithms based on the super-resolution paradigm. *Remote Sensing*, 8(10):797, 2016.

GDAL. *GDAL - Geospatial Data Abstraction Library 2.0*. Open Source Geospatial Foundation, 2017. URL `http://www.gdal.org`.

C. Geoscience Australia. Digital elevation model (dem) of australia derived from lidar 5 metre grid, 2016. URL `http://dx.doi.org/10.4225/25/5652419862E23`. Accessed: Apr 5, 2016.

D. Gesch, M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. The national elevation dataset. *Photogrammetric engineering and remote sensing*, 68(1):5–32, 2002.

L. Giosan, J. Syvitski, S. Constantinescu, and J. Day. Climate change: protect the world's deltas. *Nature*, 516(7529):31–3, dec 2014. ISSN 1476-4687. doi: 10.1038/516031a.

J. F. Girres and G. Touya. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 14(4):435–459, 2010. ISSN 13611682. doi: 10.1111/j.1467-9671.2010.01203.x.

M. H. Glantz. *Creeping environmental problems and sustainable development in the Aral Sea basin*. Cambridge University Press, Cambridge, 1999. ISBN 0-511-03506-3.

M. F. Goodchild and G. J. Hunter. A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science*, 11(3):299–306, 1997. ISSN 1365-8816. doi: 10.1080/136588197242419.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

N. Gorelick. Google Earth Engine. *AGU Fall Meeting Abstracts*, 15:11997, 2012. URL http://adsabs.harvard.edu/abs/2012AGUFM.U31A..04G.

M. Haklay. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design*, 37(4):pp. 682–703, 2010. ISSN 02658135. doi: 10.1068/b35097.

D. K. Hall, G. A. Riggs, and V. V. Salomonson. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote sensing of Environment*, 54(2):127–140, 1995.

M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. a. Turubanova, a. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, a. Kommareddy, a. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-resolution global maps of 21st-century forest cover change. *Science (New York, N.Y.)*, 342(6160):850–3, 2013. ISSN 1095-9203. doi: 10.1126/science.1244693. URL http://www.ncbi.nlm.nih.gov/pubmed/24233722.

Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae. Global flood risk under climate change. *Nature Climate Change*, 3(9):816–821, 2013.

T. Hoberg, F. Rottensteiner, R. Q. Feitosa, and C. Heipke. Conditional random fields for multitemporal and multiscale classification of optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2):659–673, 2015.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

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

A

R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson. Characterization of the landsat-7 etm+ automated cloud-cover assessment (acca) algorithm. *Photogrammetric engineering & remote sensing*, 72(10):1179–1188, 2006.

S. Jackson and A. Sleigh. Resettlement for China's Three Gorges Dam: socio-economic impact and institutional tensions. *Communist and Post-Communist Studies*, 33(2): 223–241, jun 2000. ISSN 0967067X. doi: 10.1016/S0967-067X(00)00005-2.

S. K. Jain, R. Singh, M. Jain, and A. Lohani. Delineation of flood-prone areas using remote sensing techniques. *Water Resources Management*, 19(4):333–347, 2005.

L. Ji, L. Zhang, and B. Wylie. Analysis of Dynamic Thresholds for the Normalized Difference Water Index. *Photogrammetric Engineering & Remote Sensing*, 75 (11):1307–1317, 2009. ISSN 0099-1112. doi: 10.14358/PERS.75.11.1307. URL http://essential.metapress.com/openurl.asp?genre=article{&}id=doi:10.14358/PERS.75.11.1307.

J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.

A. Karpatne, A. Khandelwal, X. Chen, V. Mithal, J. Faghmous, and V. Kumar. Global monitoring of inland water dynamics: state-of-the-art, challenges, and opportunities. In *Computational Sustainability*, pages 121–147. Springer, 2016.

D. Karssenberg and O. Schmitz. A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling & Software*, 25(4):489–502, 2010. URL http://www.sciencedirect.com/science/article/pii/S1364815209002643.

N. Keshava. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal*, 14 (1):55–78, 2003.

A. Khandelwal, A. Karpatne, M. E. Marlier, J. Kim, D. P. Lettenmaier, and V. Kumar. An approach for global monitoring of surface water extent variations in reservoirs using modis data. *Remote Sensing of Environment*, 2017.

M. L. Kirwan, S. Temmerman, E. E. Skeehan, G. R. Guntenspergen, and S. Fagherazzi. Overestimation of marsh vulnerability to sea level rise. *Nature Climate Change*, 6(3): 253–260, feb 2016. ISSN 1758-678X. doi: 10.1038/nclimate2909.

D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

C. Lamarche, M. Santoro, S. Bontemps, R. dAndrimont, J. Radoux, L. Giustarini, C. Brockmann, J. Wevers, P. Defourny, and O. Arino. Compilation and validation of sar and optical data products for a complete and global map of inland/ocean water tailored to the climate modeling community. *Remote Sensing*, 9(1):36, 2017.

C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

J.-S. Lee. Speckle analysis and smoothing of synthetic aperture radar images. *Computer graphics and image processing*, 17(1):24–32, 1981.

B. Lehner and P. Döll. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296(1):1–22, 2004.

B. Lehner and G. Grill. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15):2171–2186, 2013. URL `http://onlinelibrary.wiley.com/doi/10.1002/hyp.9740/full`.

B. Lehner, K. Verdin, and A. Jarvis. New global hydrography derived from spaceborne elevation data. *Eos*, 89(10):93–94, 2008. ISSN 00963941. doi: 10.1029/2008EO100001.

B. Lehner, C. R. Liermann, C. Revenga, C. Vörösmarty, B. Fekete, P. Crouzet, P. Döll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J. C. Robertson, R. Rödel, N. Sindorf, and D. Wisser. High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9(9):494–502, nov 2011. ISSN 1540-9295. doi: 10.1890/100125.

W. Li, Z. Du, F. Ling, D. Zhou, H. Wang, Y. Gui, B. Sun, and X. Zhang. A comparison of land surface water mapping using the normalized difference water index from TM, ETM+ and ALI. *Remote Sensing*, 5(11):5530–5549, 2013. ISSN 20724292. doi: 10.3390/rs5115530.

J. Liebe, N. Van De Giesen, and M. Andreini. Estimation of small reservoir storage capacities in a semi-arid environment: A case study in the upper east region of ghana. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(6):448–454, 2005.

J. R. Liebe, N. Van De Giesen, M. S. Andreini, T. S. Steenhuis, and M. T. Walter. Suitability and limitations of envisat asar for monitoring small reservoirs in a semiarid area. *IEEE Transactions on Geoscience and Remote Sensing*, 47(5):1536–1547, 2009.

J. Lira. Segmentation and morphology of open water bodies from multispectral images. *International Journal of Remote Sensing*, 27(18):4015–4038, 2006.

D. Liu and J. Yu. Otsu method and K-means. *Proceedings - 2009 9th International Conference on Hybrid Intelligent Systems, HIS 2009*, 1(2):344–349, 2009. doi: 10.1109/HIS.2009.74.

A. Lopes, E. Nezry, R. Touzi, and H. Laur. Maximum a posteriori speckle filtering and first order texture models in sar images. In *Geoscience and Remote Sensing Symposium, 1990. IGARSS'90.'Remote Sensing Science for the Nineties'., 10th Annual International*, pages 2409–2412. IEEE, 1990.

A

M. F. McCabe, M. Rodell, D. E. Alsdorf, D. G. Miralles, R. Uijlenhoet, W. Wagner, A. Lucieer, R. Houborg, N. E. Verhoest, T. E. Franz, et al. The future of earth observation in hydrology. *Hydrol. Earth Syst. Sci*, 2017.

R. I. McDonald, P. Green, D. Balk, B. M. Fekete, C. Revenga, M. Todd, and M. Montgomery. Urban growth, climate change, and freshwater availability. *Proceedings of the National Academy of Sciences*, 108(15):6312–6317, 2011. doi: 10.1073/pnas.1011615108. URL http://www.pnas.org/content/108/15/6312.abstract.

S. K. McFeeters. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432, 1996. ISSN 0143-1161. doi: 10.1080/01431169608948714.

M. P. Mello, J. Risso, C. Atzberger, P. Aplin, E. Pebesma, C. A. O. Vieira, and B. F. T. Rudorff. Bayesian networks for raster data (baynerd): Plausible reasoning from observations. *Remote Sensing*, 5(11):5999–6025, 2013.

P. Micklin. The future Aral Sea: hope and despair. *Environmental Earth Sciences*, 75(9): 844, may 2016. ISSN 1866-6280. doi: 10.1007/s12665-016-5614-5.

N. Mishra, M. O. Haque, L. Leigh, D. Aaron, D. Helder, and B. Markham. Radiometric cross calibration of landsat 8 operational land imager (oli) and landsat 7 enhanced thematic mapper plus (etm+). *Remote Sensing*, 6(12):12619–12638, 2014.

Y. A. Mohamed, B. J. J. M. V. den Hurk, H. H. G. Savenije, and W. G. M. Bastiaanssen. Impact of the Sudd wetland on the Nile hydroclimatology. *Water Resour. Res.*, 41(8): W08420, aug 2005.

P. Mooney, P. Corcoran, and A. C. Winstanley. Towards quality metrics for OpenStreetMap. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems GIS 10*, pages 514–517, 2010. doi: 10.1145/1869790.1869875. URL http://portal.acm.org/citation.cfm?doid=1869790.1869875.

N. Mueller, A. Lewis, D. Roberts, S. Ring, R. Melrose, J. Sixsmith, L. Lymburner, A. McIntyre, P. Tan, S. Curnow, and A. Ip. Water observations from space: Mapping surface water from 25years of Landsat imagery across Australia. *Remote Sensing of Environment*, 174:341–352, nov 2015. ISSN 00344257. doi: 10.1016/j.rse.2015.11.003. URL http://www.sciencedirect.com/science/article/pii/S0034425715301929.

M. Mulligan, L. Saenz-Cruz, A. van Soesbergen, V. Smith, and L. Zurita. Global dams database and geowiki. version 1, 2009. URL http://geodata.policysupport.org/geowiki-databases.

NASA. Shuttle Radar Topography Mission, 2016. URL http://www2.jpl.nasa.gov/srtm/.

NASA/NGA. Srtm water body data product specific guidance, version 2.0. Technical report, NASA, 2003. URL http://dds.cr.usgs.gob/srtm/version2_1/SWBD/SWBD_Documentation.

A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, 2:841–848, 2002.

A. Nobre, L. Cuartas, M. Hodnett, C. Rennó, G. Rodrigues, A. Silveira, M. Waterloo, and S. Saleska. Height Above the Nearest Drainage ? a hydrologically relevant new terrain model. *Journal of Hydrology*, 404(1-2):13–29, jun 2011. ISSN 00221694. doi: 10.1016/j.jhydrol.2011.03.051. URL http://www.sciencedirect.com/science/article/pii/S0022169411002599.

J. O'Callaghan and D. Mark. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, 28(3):323–344, 1984. URL http://www.sciencedirect.com/science/article/pii/S0734189X84800110.

OpenStreetMap. OpenStreetMap Planet File, 2016a. URL http://planet.osm.org/. Accessed: June 10, 2017.

OpenStreetMap. OpenStreetMap Taginfo, 2016b. URL http://taginfo.openstreetmap.org/tags. Accessed: June 10, 2017.

N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

J.-F. Pekel, C. Vancutsem, L. Bastin, M. Clerici, E. Vanbogaert, E. Bartholomé, and P. Defourny. A near real-time water surface detection method based on hsv transformation of modis multi-spectral time series data. *Remote sensing of environment*, 140:704–716, 2014.

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

P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990.

P. V. Potapov, S. a. Turubanova, M. C. Hansen, B. Adusei, M. Broich, A. Altstatt, L. Mane, and C. O. Justice. Quantifying forest cover loss in Democratic Republic of the Congo, 2000-2010, with Landsat ETM+ data. *Remote Sensing of Environment*, 122:106–116, 2012. ISSN 00344257. doi: 10.1016/j.rse.2011.08.027. URL http://dx.doi.org/10.1016/j.rse.2011.08.027.

C. D. Rennó, A. D. Nobre, L. A. Cuartas, J. V. Soares, M. G. Hodnett, J. Tomasella, and M. J. Waterloo. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, 112(9):3469–3481, 2008. ISSN 00344257. doi: 10.1016/j.rse.2008.03.018. URL http://www.sciencedirect.com/science/article/pii/S003442570800120X.

E. Rodriguez, C. Morris, and J. Belz. An assessment of the SRTM topographic products. *Photogrammetric Engineering and Remote Sensing*, 72(3):249–260, 2006. doi: 0099-1112/06/7203?0249/$3.00/0.

A

J. Rouse Jr, R. Haas, J. Schell, and D. Deering. Monitoring vegetation systems in the great plains with erts. 1974.

D. Roy, V. Kovalskyy, H. Zhang, E. Vermote, L. Yan, S. Kumar, and A. Egorov. Characterization of landsat-7 to landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, 185:57–70, 2016.

D. P. Roy, M. a. Wulder, T. R. Loveland, W. C.E., R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy, T. a. Scambos, C. B. Schaaf, J. R. Schott, Y. Sheng, E. F. Vermote, a. S. Belward, R. Bindschadler, W. B. Cohen, F. Gao, J. D. Hipple, P. Hostert, J. Huntington, C. O. Justice, a. Kilic, V. Kovalskyy, Z. P. Lee, L. Lymburner, J. G. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. H. Wynne, and Z. Zhu. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145:154–172, 2014. ISSN 00344257. doi: 10.1016/j.rse.2014.02.001. URL http://dx.doi.org/10.1016/j.rse.2014.02.001.

J.-H. Ryu, J.-S. Won, and K. D. Min. Waterline extraction from landsat tm data in a tidal flat: a case study in gomso bay, korea. *Remote Sensing of Environment*, 83(3):442–456, 2002.

C. Sang-Hun. South Korea Rejects North's Explanation of Dam Release, 2009.

J. Schellekens, R. Brolsma, R. Dahm, G. Donchyts, and H. Winsemius. Rapid setup of hydrological and hydraulic models using OpenStreetMap and the SRTM derived digital elevation model. *Environmental Modelling & Software*, 61:98–105, 2014. ISSN 13648152. doi: 10.1016/j.envsoft.2014.07.006. URL http://www.sciencedirect.com/science/article/pii/S1364815214002072.

C. Schwatke, D. Dettmering, W. Bosch, and F. Seitz. Dahiti–an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, 19(10):4345–4364, 2015.

J. Serra. *Image Analysis and Mathematical Morphology*, volume 1. 1982. ISBN 0126372403. doi: 10.2307/2531038. URL http://dx.doi.org/10.1016/j.procir.2013.08.008.

P. R. Sethre, B. C. Rundquist, and P. E. Todhunter. Remote detection of prairie pothole ponds in the devils lake basin, north dakota. *GIScience & Remote Sensing*, 42(4):277–296, 2005.

W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016.

J. Smith, T. L. Lin, K. Ranson, et al. The lambertian assumption and landsat data. *Photogrammetric Engineering and Remote Sensing*, 46(9):1183–1189, 1980.

S. A. Soenen, D. R. Peddle, and C. A. Coburn. Scs+ c: A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on geoscience and remote sensing*, 43(9):2148–2159, 2005.

C. Song, C. E. Woodcock, K. C. Seto, M. P. Lenney, and S. A. Macomber. Classification and change detection using landsat tm data: When and how to correct atmospheric effects? *Remote sensing of Environment*, 75(2):230–244, 2001.

C. D. Storlazzi, E. P. L. Elias, and P. Berkowitz. Many atolls may be uninhabitable within decades due to climate change. *Nature Scientific Reports*, pages 1–9, 2015. ISSN 2045-2322. doi: 10.1038/srep14546. URL http://dx.doi.org/10.1038/srep14546.

B. Tan, R. Wolfe, J. Masek, F. Gao, and E. F. Vermote. An illumination correction algorithm on landsat-tm data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 1964–1967. IEEE, 2010.

B. Tan, J. G. Masek, R. Wolfe, F. Gao, C. Huang, E. F. Vermote, J. O. Sexton, and G. Ederer. Improved forest change detection with terrain illumination corrected landsat images. *Remote Sensing of Environment*, 136:469–483, 2013.

P. Teillet, B. Guindon, and D. Goodenough. On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, 8(2):84–106, 1982.

P. Teillet, J. Barker, B. Markham, R. Irish, G. Fedosejevs, and J. Storey. Radiometric cross-calibration of the landsat-7 etm+ and landsat-5 tm sensors based on tandem data sets. *Remote sensing of Environment*, 78(1):39–54, 2001.

Toblerity. Fiona GitHub web page, 2016a. URL https://github.com/Toblerity/Fiona. Accessed: June 10, 2017.

Toblerity. Shapely GitHub web page, 2016b. URL https://github.com/Toblerity/Shapely.

C. J. Tucker. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150, 1979. ISSN 00344257. doi: 10.1016/0034-4257(79)90013-0.

UN. Transforming our world: the 2030 agenda for sustainable development. Technical report, United Nations, 2015. URL http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

F. R. Valovcin. Snow/cloud discrimination. Technical report, DTIC Document, 1976.

S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

W. Verhoef and H. Bach. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. *Remote Sensing of Environment*, 87(1):23–41, 2003.

E. F. Vermote, D. Tanré, J. L. Deuze, M. Herman, and J.-J. Morcette. Second simulation of the satellite signal in the solar spectrum, 6s: An overview. *IEEE transactions on geoscience and remote sensing*, 35(3):675–686, 1997.

W. Wagner and D. Fritsch. Big data infrastructures for processing sentinel data. In *Photogrammetric Week*, pages 93–104, 2015.

K. R. Weiss. Before we drown we may die of thirst. *Nature*, 526(7575):624–627, oct 2015. doi: 10.1038/526624a.

R. Westerhoff, H. Winsemius, H. Huizinga, G. Brakenridge, and C. Bishop. Automated global water mapping based on wide-swath orbital synthetic-aperture radar. *Hydrology and Earth System Sciences*, 17(2):651, 2013.

K. White and H. M. El Asmar. Monitoring changing position of coastlines using thematic mapper imagery, an example from the nile delta. *Geomorphology*, 29(1):93–105, 1999.

A. M. Wilson and W. Jetz. Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS Biol*, 14(3):e1002415, 2016.

E. F. Wood, J. K. Roundy, T. J. Troy, L. Van Beek, M. F. Bierkens, E. Blyth, A. de Roo, P. Döll, M. Ek, J. Famiglietti, et al. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring earth's terrestrial water. *Water Resources Research*, 47 (5), 2011.

H. Xu. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14): 3025–3033, 2006. ISSN 0143-1161. doi: 10.1080/01431160600589179.

D. Yamazaki and M. A. Trigg. Hydrology: The dynamics of earth's surface water. *Nature*, 540(7633):348–349, 2016.

D. Yamazaki, M. A. Trigg, and D. Ikeshima. Development of a global˜ 90m water body map using multi-temporal landsat images. *Remote Sensing of Environment*, 171:337–351, 2015.

K. Yang, M. Li, Y. Liu, L. Cheng, Y. Duan, and M. Zhou. River Delineation from Remotely Sensed Imagery Using a Multi-Scale Classification Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12):4726–4737, 2014. ISSN 1939-1404. doi: 10.1109/JSTARS.2014.2309707. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6777572`.

Y. Yang, Y. Liu, M. Zhou, S. Zhang, W. Zhan, C. Sun, and Y. Duan. Landsat 8 oli image based terrestrial water extraction from heterogeneous backgrounds using a reflectance homogenization approach. *Remote Sensing of Environment*, 171:14–32, 2015.

Y. Zhang, X. Wang, and Y. Chen. An improved 6s code for atmospheric correction based on water vapor content. 2012.

Z. Zhu and C. E. Woodcock. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118:83–94, 2012. ISSN 00344257. doi: 10.1016/j.rse.2011.10.028. URL http://www.sciencedirect.com/science/article/pii/S0034425711003853.

Z. Zhu and C. E. Woodcock. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment*, 152:217–234, 2014. ISSN 00344257. doi: 10.1016/j.rse.2014.06.012. URL http://dx.doi.org/10.1016/j.rse.2014.06.012.

Z. Zhu, S. Wang, and C. E. Woodcock. Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote Sensing of Environment*, 159:269–277, 2015.