

Article

# Automated Extraction of Consistent Time-Variable Water Surfaces of Lakes and Reservoirs Based on Landsat and Sentinel-2

Christian Schwatke <sup>\*ID</sup>, Daniel Scherer <sup>ID</sup> and Denise Dettmering <sup>ID</sup>

Deutsches Geodätisches Forschungsinstitut, Technische Universität München, Arcisstraße 21, 80333 München, Germany; daniel.scherer@tum.de (D.S.); denise.dettmering@tum.de (D.D.)

\* Correspondence: christian.schwatke@tum.de; Tel.: +49-89-23031-1109

Received: 7 March 2019; Accepted: 23 April 2019; Published: 28 April 2019



**Abstract:** In this study, a new approach for the automated extraction of high-resolution time-variable water surfaces is presented. For that purpose, optical images from Landsat and Sentinel-2 are used between January 1984 and June 2018. The first part of this new approach is the extraction of land-water masks by combining five water indexes and using an automated threshold computation. In the second part of this approach, all data gaps caused by voids, clouds, cloud shadows, or snow are filled by using a long-term water probability mask. This mask is finally used in an iterative approach for filling remaining data gaps in all monthly masks which leads to a gap-less surface area time series for lakes and reservoirs. The results of this new approach are validated by comparing the surface area changes with water level time series from gauging stations. For inland waters in remote areas without in situ data water level time series from satellite altimetry are used. Overall, 32 globally distributed lakes and reservoirs of different extents up to 2482.27 km<sup>2</sup> are investigated. The average correlation coefficients between surface area time series and water levels from in situ and satellite altimetry have increased from 0.611 to 0.862 after filling the data gaps which is an improvement of about 41%. This new approach clearly demonstrates the quality improvement for the estimated land-water masks but also the strong impact of a reliable data gap-filling approach. All presented surface area time series are freely available on the Database of Hydrological Time Series of Inland (DAHITI).

**Keywords:** surface area time series; spatial data gaps; lakes; reservoirs; Landsat; Sentinel-2; MNDWI; NWI; AWEI<sub>sh</sub>; AWEI<sub>nsh</sub>; TC<sub>wet</sub>; satellite altimetry; DAHITI; AWAX

## 1. Introduction

Monitoring and modeling of the Earth's water cycle has become increasingly important in the last few years, especially in the context of climate change. Though only 0.013% [1] of the Earth's water is stored in lakes and reservoirs, the knowledge about storage changes is of great importance for the development of hydrological models. In the 1970s, the first remote sensing satellites such as Landsat-1 for mapping the Earth's surface and Seasat for measuring the water level from space were launched. In the following decades until the present day, numerous further remote sensing satellites carrying sensors of improved quality have been launched. Presently, the usage of multi-mission approaches allows scientists to create time series of surface areas or water levels over a period of more than 30 years for lakes and reservoirs.

Time series give an insight in the dynamics of lakes and reservoirs which are fundamental for risk assessment of natural disasters. The number of reported natural disasters increased from about 150 in 1980 to more than 400 in 2008, where in 38% of the events can be assigned to wet mass movements

such as floods [2]. Also, the number of dams used worldwide for irrigation, hydroelectricity, etc. has increased over the last few decades. This can lead to water shortage in countries such as Iran which can end up in political crisis [3]. Over the last few decades, the number of in situ data used for local monitoring has strongly decreased. The reasons are mainly that no data exists or is not accessible. Therefore, remote sensing is a valuable tool for observing water bodies independently from space. An example for decreasing of in situ data is the Global Runoff Data Center (GRDC) where the number of river discharge stations decreased from about 7900 to 600 between 1979 and 2018 [4].

The extraction and identification of water bodies is essential for applications such as monitoring flooded regions [5] and wetlands [6,7], but also for the water storage quantification of lakes and reservoirs [8]. Remote sensing is also capable of observing long-term geomorphological changes of rivers and their meandering [9].

The main objective of using optical images is the classification of land cover types in categories such as water, vegetation, ice, etc. In this research field, one of the most common approaches is the land-water classification which is mainly based on the calculation of water indexes using different spectral information from optical images. Many water indexes such as Modified Normalized Difference Water Index (MNDWI) [10], New Water Index (NWI) [11], Automated Water Extraction Index for Non-Shadow Areas ( $AWEI_{nsh}$ ) [12], Automated Water Extraction Index for Shadow Areas ( $AWEI_{sh}$ ) [12], and Tasseled Cap for Wetness ( $TC_{wet}$ ) [13] have been developed and used in different modifications, combinations, and approaches over the last few decades. Based on the calculated land-water indexes, the more challenging task is the delineation of land and water in optical images. Gómez et al. [14] gives a clear overview of various algorithms such as artificial neural networks, clustering, decision trees, Gaussian maximum likelihood, or Support Vector Machines which have already been used for the classification of land cover. This also includes the separation of land and water pixels. All these approaches have a common weakness which is that they require training data for their classification. This is not the case in our study. For the separation of land and water pixels, the proven Otsu threshold algorithm [15] is often used. However, this requires a bimodal histogram distribution. In this paper, a new approach for the threshold computation based on five histograms derived from different land-water indexes is demonstrated.

A variety of global surface land-water products using optical images such as MODIS, Landsat, or Sentinel-2 has already been available for more than 30 years. Nevertheless, the spatial resolutions differ from 10 m for Sentinel-2, 30 m for Landsat and 250 m for MODIS which has a strong impact on the quality, especially on the resulting delineation of land and water masks at the shores. The Global Land Cover Facility (GLFC) developed the GLS2000 product which is a global surface water data set using Landsat and MODIS data for the year 2000 and with a spatial resolution of 30 m [16]. Another high-resolution global surface water data set based on Landsat and Google Earth Engine providing long-term information has been developed by Pekel et al., without pixel-wise accuracy and remaining data gaps [17]. The global WaterPack is a daily land-water data set based on MODIS data with a spatial resolution of 250 m where data gaps have been filled by a temporal interpolation [18].

The reconstruction of data gaps caused by voids, clouds, cloud shadows, or ice is an important step for the creation of homogeneous surface area time series. For that reason, a new approach has been developed in this paper. So far, several other approaches based on different data sets have already been developed. For example, data gaps in MODIS data with a high temporal resolution of 1–2 days can be reconstructed by applying a temporal interpolation [18]. Combining SAR with optical images is another approach to reduce data gaps caused by clouds [19].

Further multi-mission applications arise of using surface areas derived from optical images in combination with satellite altimetry for validation or computation of volume changes which has been already proven in several studies [8,20,21].

In this paper, a new approach for computing monthly surface area time series of lakes and reservoirs based on optical images from Landsat and Sentinel-2 is presented. This approach is divided into two major parts. The first part is the extraction of land-water masks with remaining data gaps by

combining five water indexes and an automated threshold computation. The combination of different water indexes leads to more accurate land-water masks. The second part of this approach is the step of filling data gaps caused by voids, clouds, cloud shadows, or snow. This is performed by estimating a long-term water probability mask based on data gained between January 1984 and June 2018. Then, this mask is used in an iterative approach for filling the data gaps in all monthly masks. And finally, a gap-less surface area time series is computed for 32 selected lakes and reservoirs. The new gap-filling approach leads to improved and more consistent surface area time series if data gaps are corrected.

The article is structured as follows. In Section 2 the data which serve as input for the processing and validation are described. In Section 3 the methodology for estimating surface area time series of inland waters based on classification and thresholds is explained. Section 4 begins with the introduction of 32 study areas. Then, three selected study areas are presented in detail followed by a validation and quality assessment of all study areas and its resulting surface area time series. The paper concludes with a summary of the results with discussion and an outlook.

## 2. Data

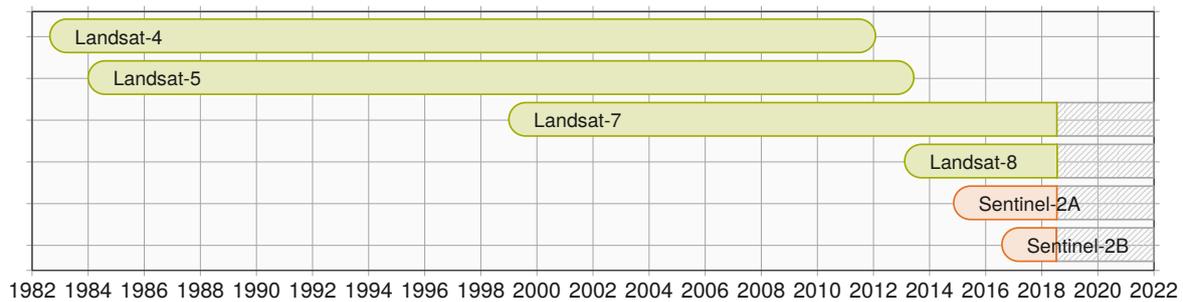
In this study, optical images from Landsat and Sentinel-2 are used to compute monthly land-water masks and finally surface area time series of lakes and reservoirs. To validate our results, we use water level time series from in situ stations and satellite altimetry.

### 2.1. Satellite Imagery

#### 2.1.1. Landsat

The Landsat program operated by the National Aeronautics and Space Administration (NASA) is the longest-running mission for providing optical images of the Earth. Landsat-1 was the first satellite launched in 1972. Since then, seven further Landsat satellites have been launched in the frame of the Landsat Data Continuity Mission (LDCM) [22], from Landsat-2 in 1975 to Landsat-8 in 2013. Figure 1 gives an overview of time spans of the used Landsat missions. The technology of Landsat satellites has improved over the last few decades. Starting with Landsat-4, the spatial resolution has been increased from about 75 m to 30 m which improves the extraction of smaller features in images. In this study, we use the six bands (red, green, blue, near-infrared, shortwave infrared 1, shortwave infrared 2) with a resolution of 30 m from Landsat-4, Landsat-5, Landsat-7, and Landsat-8 which have comparable characteristics. The Landsat-6 mission failed to reach the orbit. All Landsat satellites have a Sun-synchronous orbit with an inclination of  $98.2^\circ$  and a height of 705 km. The Landsat missions have a repeat cycle of 16 days. Combining Landsat-7 and Landsat-8 increases the temporal resolution to 8 days. The Landsat satellites are equipped with different sensors. Landsat-4 and Landsat-5 have a Multispectral Scanner (MSS) and the Thematic Mapper (TM) on board. Landsat-7 carries an Enhanced Thematic Mapper (ETM+) which is similar to the MSS and TM of previous Landsat missions. Landsat-8 is equipped with an Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS).

All Landsat data are freely available in the United States Geological Survey (USGS) archive. But the huge amount of data is the limiting factor in the processing of Landsat images. In order to decrease the amount of data, we use the Google Earth Engine [23], which allows us to extract the best cloud-free monthly composites of Landsat images. Google Earth Engine also provides tools for the cloud-processing of images. Furthermore, a quality band for cloud, shadow and snow coverage is already included in the Landsat data sets using the CFMASK (C Function of Mask) algorithm [24]. Since May 2003, a Scan Line Corrector Error has occurred on Landsat-7, leading to additional stripe-shaped data gaps which have to be considered in the data processing step. For this study, we use the Level-2 product "Surface Reflectance" for the Landsat satellites in the period between January 1984 and June 2018 for which the atmosphere has already been corrected.



**Figure 1.** Overview of the used satellites.

### 2.1.2. Sentinel-2

The European Space Agency (ESA) initiated the Copernicus program in which two new Earth observation satellites of Sentinel-2 have been developed [25]. The identical satellites are named Sentinel-2A and Sentinel-2B (Figure 1). Sentinel-2A was launched on 23 June 2015 and has a repeat cycle of 10 days. With the start of Sentinel-2B on 07 March 2017, the temporal resolution of the full Sentinel-2 constellation increased to 5 days. Both satellites have a Sun-synchronous orbit with an inclination of  $98.5^\circ$  and a height of 786 km. Sentinel-2 is equipped with a Multispectral Imager (MSI) providing 12 spectral bands similar to Landsat which is shown in Table 1. However, the spatial resolution has improved and varies between 10 m (red, green, blue), and 20 m (near-infrared, shortwave infrared 1, shortwave infrared 2) for the used bands.

In this study, the Level-1C product of Sentinel-2 retrieved from the Copernicus Open Access Hub maintained by ESA is used. This product contains as already mentioned top-of-atmosphere (TOA) reflectances which are referenced to a fixed cartographic geometry (combined UTM projection and WGS84 ellipsoid). Radiometric and geometric corrections have already been applied to all images. Each scene has an area of about  $100 \text{ km}^2$ .

Since recently, ESA provides Sentinel-2 Level-2A bottom-of-atmosphere (BOA) products on their Copernicus Open Access Hub platform. Additionally, ESA provides a processor “Sen2Cor” [26] for the generation of the Sentinel-2 Level-2A BOA product which performs a correction of atmosphere, terrain and cirrus clouds using the Sentinel-2 Level-1C TOA product. However, a scene classification for clouds, water, cloud shadows, etc. is created which is essential for our study. Unfortunately, the scene classification contains unclassified pixels near lakes shores but also in transition zones between clouds and cloud shadows which is therefore not suitable for our purpose.

The Level-1C product does not contain a quality band including flags for clouds, shadows, and snow which is required for the processing of this data set. In order to get a suitable quality band, the Fmask 4.0 algorithm is applied on daily Sentinel-2 scenes. Afterwards, all used daily Sentinel-2 scenes are merged by using the computed quality band to get the best cloud-free monthly composite.

The Fmask 4.0 algorithm [27] is an improvement of the CFMASK approach used for Landsat to compute a quality band. This is required for our approach as it provides information about the scene classification pixel-by-pixel. In Landsat, each pixel of the quality band is classified as valid, void, cloud, cloud shadow, or snow. The Fmask 4.0 algorithm works only with the Level-1C product of Sentinel-2. Therefore, we use the TOA Sentinel-2 Level-1C product to get a reliable quality band. Tests showed that the discrepancy between the TOA Level-1C product and BOA Level-2 product have no significant impact in our study.

**Table 1.** Overview of used individual and standard band names.

Band Description	Landsat-4,-5,-7 (TM and ETM+)	Landsat-8 (OLI)	Sentinel-2A,-2B	Standard Name
Blue	Band 1	Band 2	Band 2	<i>BLUE</i>
Green	Band 2	Band 3	Band 3	<i>GREEN</i>
Red	Band 3	Band 4	Band 4	<i>RED</i>
Near-Infrared (NIR)	Band 4	Band 5	Band 8	<i>NIR</i>
Short Wave Infrared 1 (SWIR <sub>1</sub> )	Band 5	Band 6	Band 11	<i>SWIR<sub>1</sub></i>
Short Wave Infrared 2 (SWIR <sub>2</sub> )	Band 7	Band 7	Band 12	<i>SWIR<sub>2</sub></i>
Mask (Clouds, Shadow, etc.)	CFmask	CFmask	Fmask4.0	<i>QUALITY</i>

## 2.2. In Situ Data

Water levels from in situ stations are the primary source for validation as the surface area and water level are highly correlated. The relationship between surface area and water level for lakes and reservoirs is always monotonic increasing caused by the surrounding topography. Therefore, we use water level data of 15 stations for validation provided by the USGS, Bavarian Environmental Agency (BEA) and Agência Nacional de Águas (ANA) shown in Table 2. In this study, 32 globally distributed study areas are investigated (see Section 4.1).

**Table 2.** List of used in situ stations.

Lake/Reservoir	Station Name	Station ID	Source
Claiborne	Aycock	07364840	USGS
Clear	Clarence	07352895	USGS
Cooper	Cooper	07342495	USGS
Forggen	Rosshaupten	12001301	BEA
Lewisville	Lewisville	08052800	USGS
Mead	n.a.	n.a.	USGS
Murray	Columbia	02168500	USGS
Nova Ponte	n.a.	n.a.	ANA
Poço da Cruz	n.a.	n.a.	ANA
Ray Roberts	Pilot Point, TX	08051100	USGS
Richland Chambers	Kerens, TX	08064550	USGS
Salton Sea	Westmorland	10254005	USGS
Sam Rayburn	Jasper	08039300	USGS
Tawakoni	Wills Point	08017400	USGS
Toledo Bend	Burkeville	08025350	USGS

## 2.3. Satellite Altimetry

Satellite altimeter data is an alternative data source for validation as it can be used also in remote areas where no gauging stations are available. For many years, satellite altimetry has been proven to have the capability of monitoring water level changes of inland waters [28–30]. Depending on the investigated inland water body, water level time series since 1992 can be achieved. In this study, we use altimeter missions ERS-2 (1995–2011), Envisat (2001–2013) and SARAL (since 2013) all characterized by the same orbit with a repeat cycle of 35 days. Additionally, the altimeter missions Topex/Poseidon (1992–2006), Jason-1 (2001–2013), Jason-2 (since 2008) and Jason-3 (since 2016) are used for the estimation of water level time series which have a repeat cycle of about 10 days. Sentinel-3A (since 2016) has a repeat cycle of 27 days and has a different orbit than the other two satellite constellations. The temporal resolution of the water level time series depends on the size and shape of the lakes. This fact influences the number of crossing satellite tracks. The temporal resolution varies between a few days for large inland waters crossed by several altimeter tracks and 35 days for small lakes crossed by only a single altimeter track (e.g., Envisat). Satellite altimetry has the potential to provide

water level time series also in remote areas where no in situ stations are available. The accuracy varies between a few centimeters for large lakes and several decimeters for smaller rivers.

In this study, water level time series of the “Database of Hydrological Time Series of Inland Waters” (DAHITI) developed by the “Deutsche Geodätisches Forschungsinstitut der Technischen Universität München” are used. The processing strategy of DAHITI is based on an extended outlier detection and a Kalman filtering approach [31]. Currently, DAHITI contains more than 1600 freely available water level time series of inland waters.

### 3. Methodology

In this section, we describe the new “Automated Water Area Extraction Tool” (AWAX) in detail. First, we start with the initialization, followed by a monthly pre-processing of the Landsat and Sentinel-2 images. Then, monthly land-water masks with remaining data gaps caused by clouds, cloud shadows, or snow are created. Also, data voids caused by sensor errors are considered. This is achieved by calculating five different water indexes and a classification to separate land and water pixels. Then, monthly land-water masks with remaining data gaps are computed by combining 5 water indexes (see Section 3.2.2). Based on all available land-water masks with data gaps between January 1984 and June 2018, a water probability mask is computed. In the next step, the long-term water probability mask is used in an iterative approach to fill remaining data gaps in the monthly data. Finally, based on all gap-filled land-water masks a surface area time series is computed. Figure 2 shows the flowchart of the AWAX approach in detail.

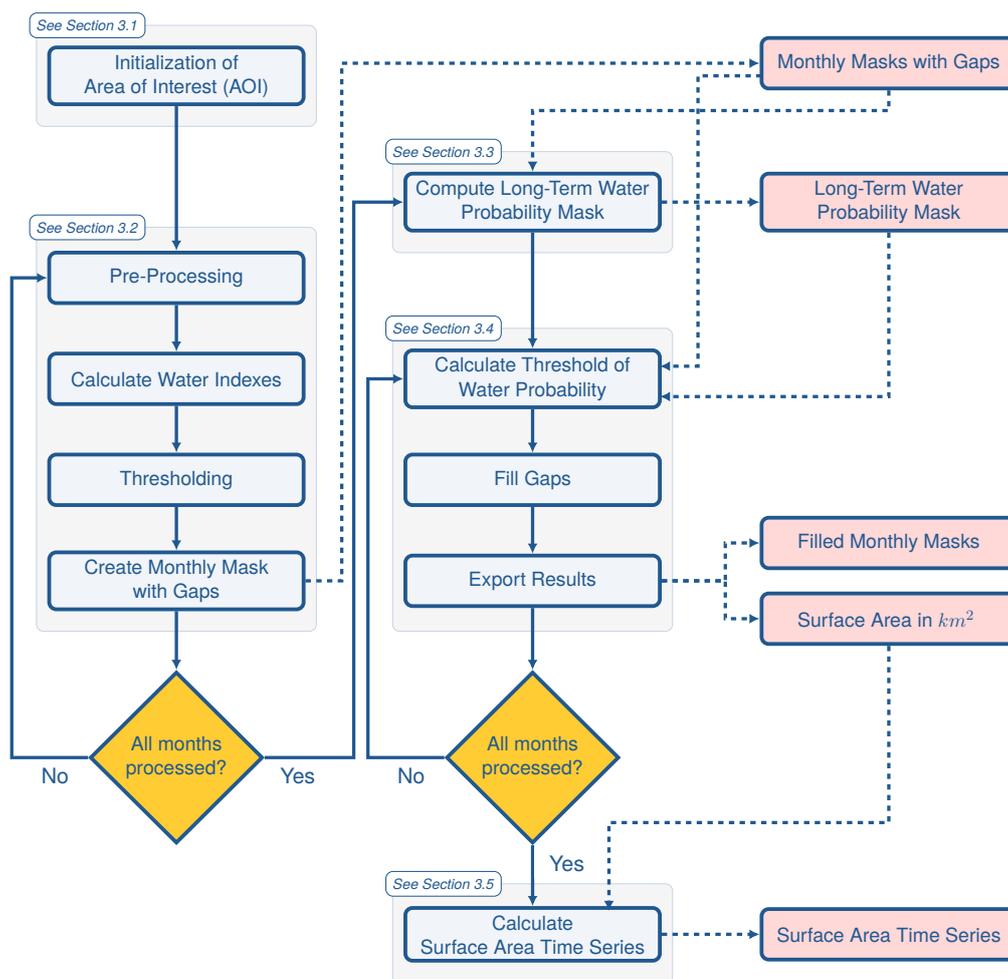


Figure 2. Flowchart of the AWAX with processing steps (blue) and resulting data sets (red).

### 3.1. Initialization

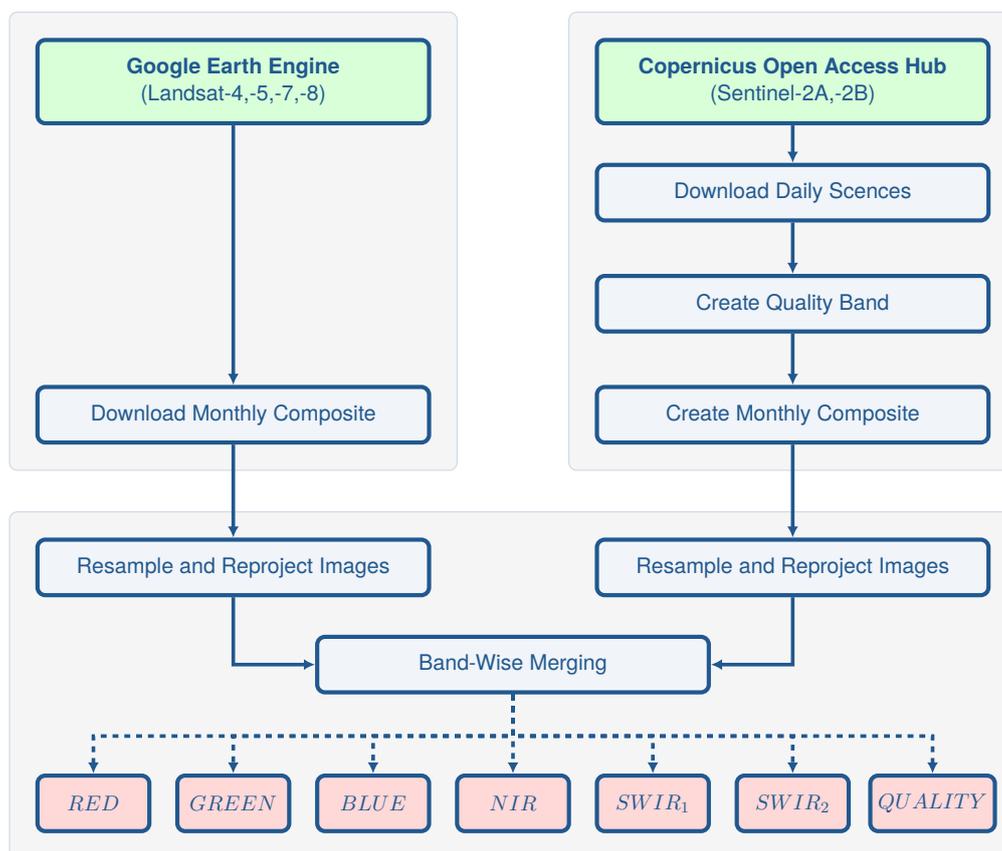
The automated processing of surface water level time series requires only two input parameters. First, a coarse boundary of the investigated inland water body is required to extract the Landsat data from the Google Earth Engine. Secondly, a reference point located within the inland water is necessary for the automated selection of the investigated lake or reservoir.

### 3.2. Computation of Monthly Land-Water Masks with Data Gaps

The computation of monthly land-water mask with remaining data gaps caused by clouds, cloud shadows, and snow/ice is divided into four processing steps. The first step is the pre-processing step in which the input data are downloaded. Afterwards, the data are homogenized by converting them to the same spatial resolution and image projection. In the second step, water indexes are computed followed by an automated classification which is finally used to create monthly land-water masks with remaining data gaps.

#### 3.2.1. Pre-Processing

In the pre-processing step, monthly composite images and quality flags derived from Landsat and Sentinel-2 are created. Figure 3 gives an overview of the applied steps to achieve monthly composites.



**Figure 3.** Flowchart of the AWAX-Pre-processing.

#### Acquisition and Pre-Processing of Landsat Imagery

The coarse boundary of the inland water body from the initialization is used as input data to extract the data of interest using the Google Earth Engine. First, available scenes of the different Landsat missions within the area and month of interest are selected and stored in an image collection. Every scene contains additional information such as cloud coverage in percent and a quality band with flags for clouds, cloud shadows, and snow calculated by the CFMASK algorithm. These are used

for merging all scenes to create the best gap-free composite image. The creation of composite images is performed pixel-wise in the Google Earth Engine by iterating through all scenes sorted by cloud coverage, searching for the smallest quality flag. Considering the order of cloud coverage leads to a more homogeneous cloud-free composite image.

In this study, we use the red (*RED*), green (*GREEN*) and blue (*BLUE*) bands as well as the near-infrared (*NIR*), shortwave infrared 1 (*SWIR<sub>1</sub>*) and shortwave infrared 2 (*SWIR<sub>2</sub>*) bands. Additionally, existing masks of voids, clouds, cloud shadows and snow are extracted for managing data gaps and are stored in a quality (*QUALITY*) band. All seven bands are downloaded as monthly composites from the Google Earth Engine.

#### Acquisition and Pre-Processing of Sentinel-2 Imagery

The data source for downloading Sentinel-2 images is the Copernicus Open Access Hub provided by the ESA. In this study, the Level-1C data set of Sentinel-2A and Sentinel-2B are used. These scenes are stored as tiles with an extent of 100 km × 100 km which are TOA reflectances in cartographic geometry.

Initially, all available Sentinel-2A and Sentinel-2B tiles within the boundary of the inland water body defined in the initialization step are downloaded. Naturally, Level-1C are only raw data sets, without additional information such as clouds, cloud shadows, or snow flags which are required in this study. Therefore, we apply the FMask algorithm [27] in advance to create these flags. Based on the computed flags and the daily images, we create a monthly composite which is as free of gaps as possible by averaging all gap-free pixels. Since the spatial resolution for the bands differ (10 m for *RED*, *GREEN* and *BLUE*, 20 m for *NIR*, *SWIR<sub>1</sub>* and *SWIR<sub>2</sub>*), we resample the infrared bands to 10 m spatial resolution.

#### Combination of Landsat and Sentinel-2 Imagery

In this step, the monthly Landsat and Sentinel-2 composites are merged to achieve a multi-mission composite. A monthly temporal resolution was selected to reduce the amount of data. But this approach can also be adapted on daily images. Since both data sets have different spatial resolutions ('10 m' or '30 m') and projections ('WGS 84' or 'WGS 84/UTM'), a homogenization is necessary. In our case, all the monthly Landsat and Sentinel-2 composites are resampled to a spatial resolution of 0.0001° which is about 10 m and reflects the spatial resolution of Sentinel-2. Also, we reproject all images to the WGS84 coordinate system. Then, the images are cut out with respect to the polygon of the investigated water body to get two images with an identical shape.

Finally, we merge each band of both images pixel-wise. This is performed for the *RED*, *GREEN*, *BLUE*, *NIR*, *SWIR<sub>1</sub>*, *SWIR<sub>2</sub>* and *QUALITY* band. If both pixels are without data gaps, both pixels of each data set are averaged. In case of a data gap in one of the two data sets, the gap-less value is taken. If both pixels have data gaps, the data gap remains in the final merged composite. By averaging all valid pixels, the resulting composite image represents a quasi-averaged surface area at which initial scenes are made with an irregular temporal resolution within a month.

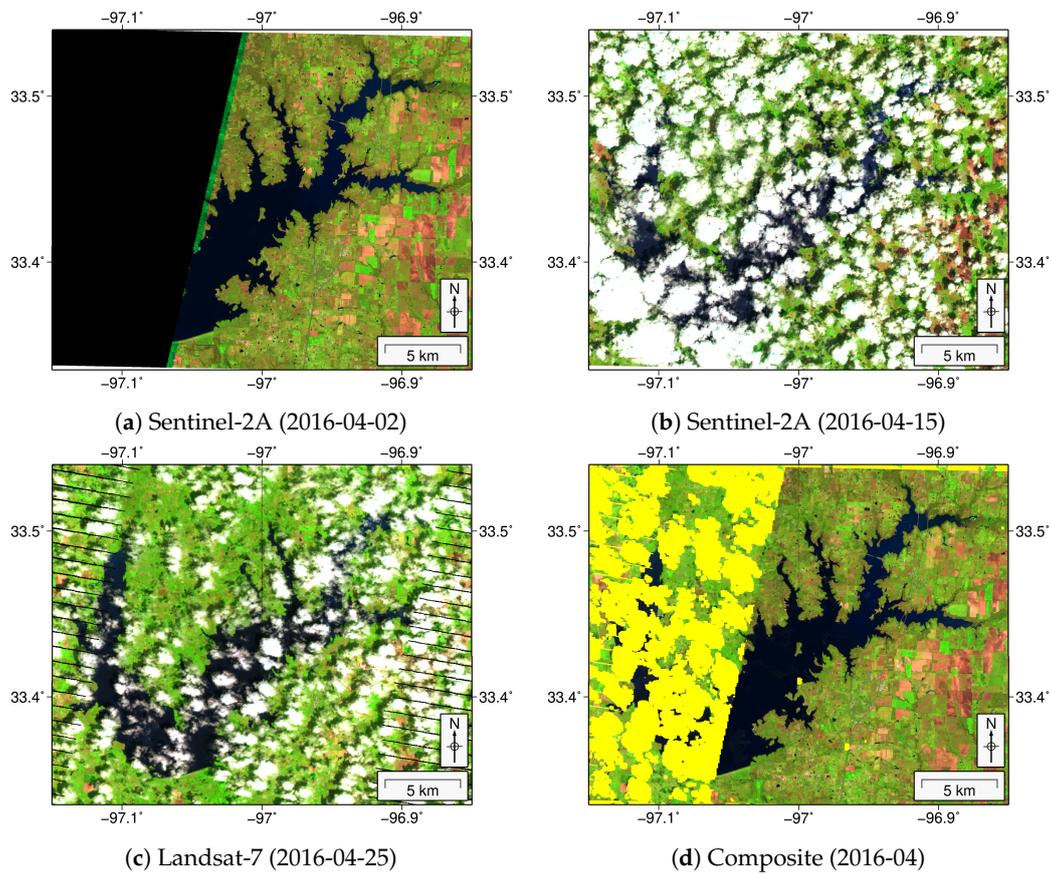
Figure 4 shows an example for the image combination for Lake Ray Roberts in April 2016. For this month, three images are available, two from Sentinel-2 and one from Landsat. Figure 4d shows the combined monthly composite. It contains optical images of Landsat-7 from the 2016-04-25 which has a cloud coverage of 27%. The monthly Sentinel-2 composite contains two optical images of Sentinel-2A from the 2016-04-02 (0% cloud coverage) and 2016-04-15 (27% cloud coverage). Both daily Sentinel-2A images are taken from different satellite tracks. This causes additional no-data gaps in the daily scene from Sentinel-2A on 2016-04-02 (Figure 4a) as the satellite stripes the study area on the edge only.

Despite the usage of several scenes, clouds, cloud shadows, and snow gaps remain in the resulting composite. One can clearly see that the cloud-free Sentinel-2A image has the largest impact on the resulting combined composite because the other images are covered with clouds. The remaining data gap in the cloud-free image is then filled by combining the other cloud-covered scenes. Therefore, the monthly composite is divided into a cloud-free part on the right and a cloud-covered part on the

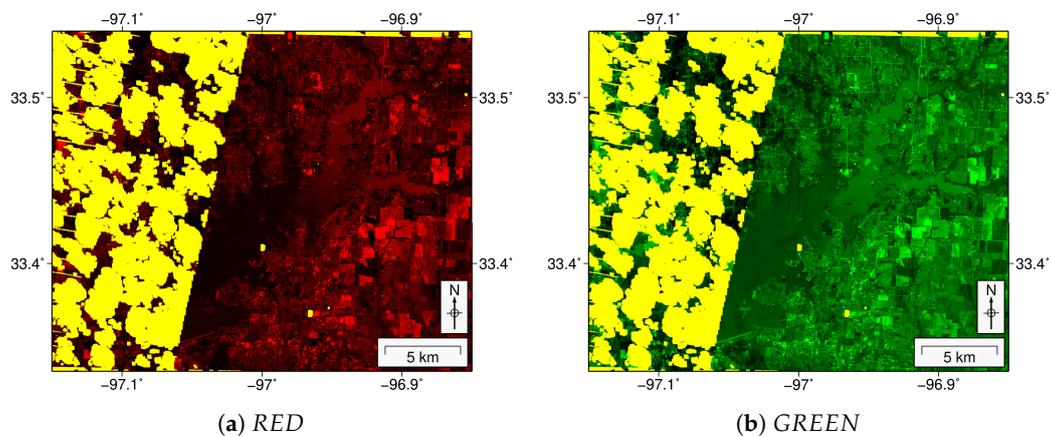
left. In the cloud-covered area on the left, one can also see thin lines of data gaps caused by the SLC error in Landsat 7 which are filled with cloudy scenes from Sentinel-2A.

Figure 5, finally shows all merged monthly bands with data gaps of the composite image using Landsat and Sentinel-2 for Lake Ray Roberts in April 2016.

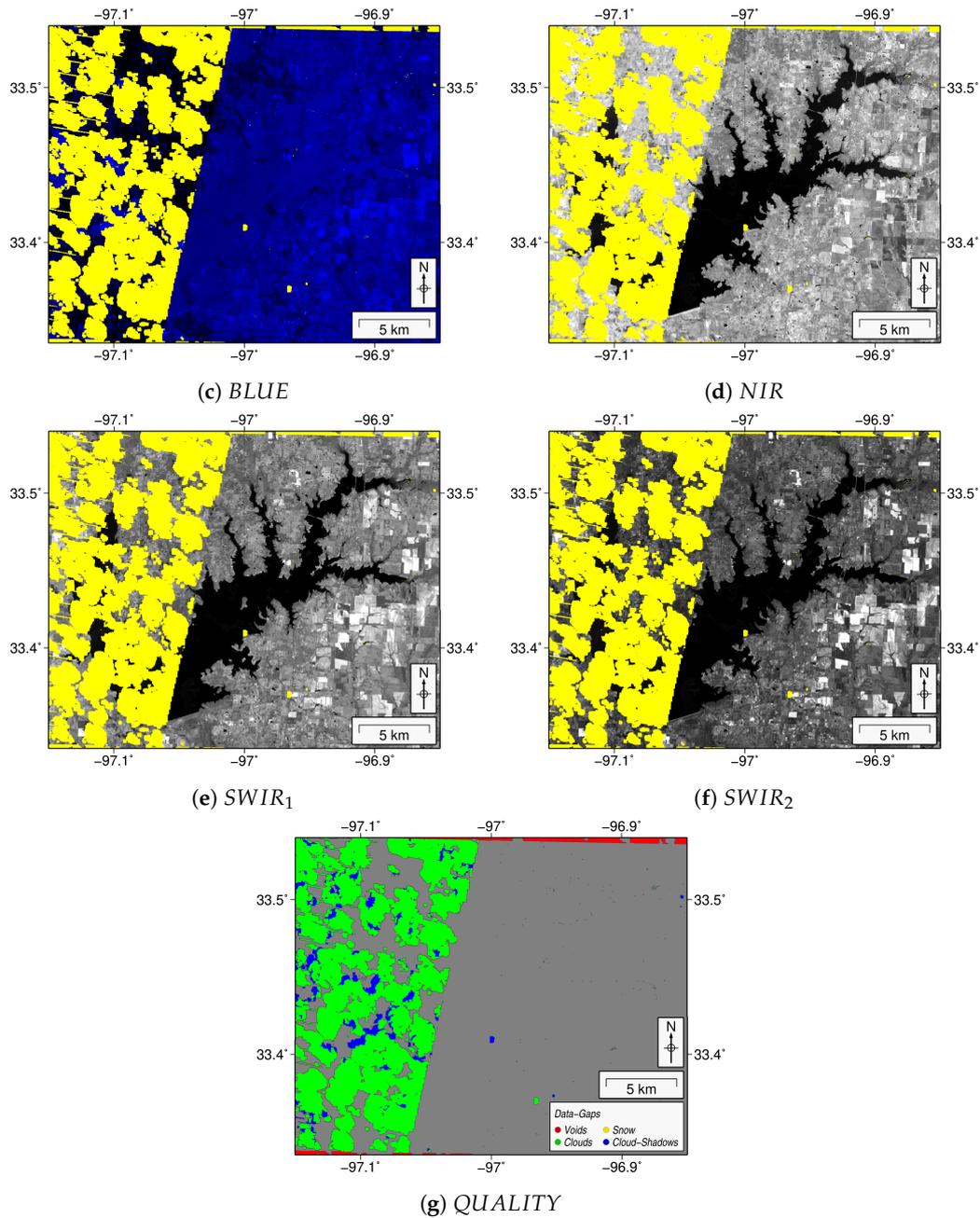
Even if the different band behavior of the two missions is slightly influencing the index computation (see Section 3.2.2), the threshold estimation is not impacted, so that the combination can be safely done by simple averaging.



**Figure 4.** Used daily scenes from Landsat and Sentinel-2 and resulting composite image for Lake Ray Roberts in April 2016. Data gaps in the composite image are highlighted in yellow.



**Figure 5.** Cont.



**Figure 5.** Combined Landsat and Sentinel-2 bands (*RED*, *GREEN*, *BLUE*, *NIR*, *SWIR*<sub>1</sub>, *SWIR*<sub>2</sub>) and *QUALITY* mask (voids, clouds, cloud shadows, snow) used for the computation of indexes for Lake Ray Roberts in April 2016. Data gaps in the Figure 5a–f are highlighted in yellow.

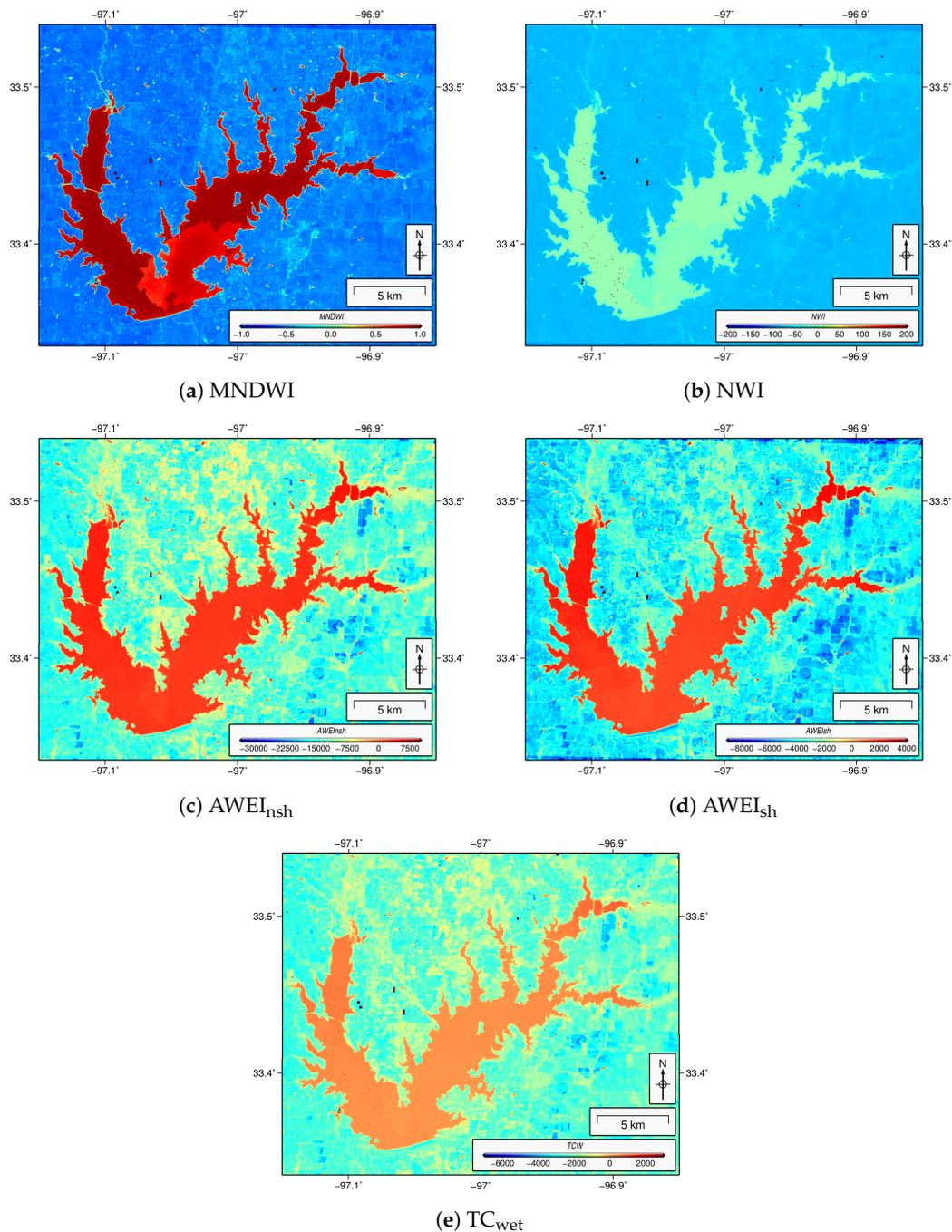
### 3.2.2. Calculation of Water Indexes

Over the last few decades, many water indexes used for classifying water bodies have been developed. However, no optimal water index exists working in all areas. Especially near shores water indexes react with a different sensitivity. Each water index is developed to separate land and water pixels as good as possible. The resulting histograms of the water indexes have bimodal-like shapes, wherein land and water must be separated. Therefore, we combine five indexes to use their individual strengths for the identification of water bodies. In detail, we use the

- Modified Normalized Difference Water Index (MNDWI) [10],
- New Water Index (NWI) [11],

- Automated Water Extraction Index for Non-Shadow Areas ( $AWEI_{nsh}$ ) [12],
- Automated Water Extraction Index for Shadow Areas ( $AWEI_{sh}$ ) [12] and
- Tasseled Cap for Wetness ( $TC_{wet}$ ) [13].

For calculating the indexes, different combinations of visible bands (*RED*, *GREEN*, *BLUE*) and infrared bands (*NIR*, *SWIR<sub>1</sub>*, *SWIR<sub>2</sub>*) are used. All five water indexes used in this study are now presented in detail. Figure 6 shows as an example the results of all five indexes for the Lake Ray Roberts in January 2017.



**Figure 6.** The used water indexes, Modified Normalized Difference Water Index (MNDWI), New Water Index (NWI), Automated Water Extraction Index for Non-Shadow Areas ( $AWEI_{nsh}$ ), Automated Water Extraction Index for Shadow Areas ( $AWEI_{sh}$ ) and Tasseled Cap for Wetness ( $TC_{wet}$ ) are shown for Lake Ray Roberts in January 2017.

### Modified Normalized Difference Water Index (MNDWI)

The MNDWI [10] is a modified version of the well-known Normalized Difference Water Index (NDWI) [32] shown in Equation (1). Here, the *NIR* band has been replaced by the *SWIR*<sub>1</sub> band. The value range of the MNDWI is always scaled between  $-1.0$  and  $1.0$ . A threshold of  $0.0$  can be selected to get a good but not the best separation of land and water pixels.

$$MNDWI = \frac{GREEN - SWIR_1}{GREEN + SWIR_1} \quad (1)$$

### New Water Index (NWI)

The NWI, shown in Equation (2), is a water index using spectral rationing, considering more bands in the infrared spectrum and the blue instead of the *GREEN* band [11]. Additionally, the result is multiplied by a constant value *C* (e.g., 100) to scale the index, to avoid low decimal numbers. NWI has already been used in different studies such as [33].

The NWI is expressed as:

$$NWI = \frac{BLUE - (NIR + SWIR_1 + SWIR_2)}{BLUE + (NIR + SWIR_1 + SWIR_2)} \times C \quad (2)$$

### Automated Water Extraction Index for Non-Shadow Areas (AWEI<sub>nsh</sub>)

The Automated Water Extraction Index for Non-Shadow Areas (AWEI<sub>nsh</sub>) has been developed to maximize separability of water and other dark areas [12] which is shown in Equation (3). In contrast to the previous methods no ratio is used, but weighted indicators are subtracted from each other. The AWEI<sub>nsh</sub> is specifically formulated to separate water from dark surfaces such as buildup areas, disregarding shadows. It is expressed as:

$$AWEI_{nsh} = 4.0 \times (GREEN - SWIR_1) - (0.25 \times NIR + 2.75 \times SWIR_2) \quad (3)$$

### Automated Water Extraction Index for Shadow Areas (AWEI<sub>sh</sub>)

The Automated Water Extraction Index for Shadow Areas (AWEI<sub>sh</sub>) (Equation (4)) works similar to the AWEI<sub>nsh</sub> but is more suited for regions where shadows are a major problem such as valleys in alpine regions [12]. The index additionally includes the *BLUE* band and uses a different weighting of the bands.

The AWEI<sub>sh</sub> is expressed as:

$$AWEI_{sh} = BLUE + 2.5 \times GREEN - 1.5 \times (NIR + SWIR_1) - 0.25 \times SWIR_2 \quad (4)$$

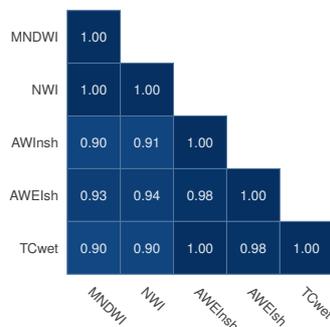
### Tasseled Cap for Wetness (TC<sub>wet</sub>)

The Tasseled Cap index is based on a method to transform spectral information of satellite data into spectral indicators [13,34]. Depending on the application, three different types of Tasseled Cap are available, optimized for brightness, greenness, and wetness. In this study, we use the Tasseled Cap for wetness TC<sub>wet</sub> because the interest lies in the detection of open water. The algorithm, shown in Equation (5), consists of a weighted sum of all six bands from Landsat or Sentinel-2.

$$TC_{wet} = 0.1509 \times BLUE + 0.1973 \times GREEN + 0.3279 \times RED + 0.3406 \times NIR - 0.7112 \times SWIR_1 - 0.4572 \times SWIR_2 \quad (5)$$

In summary, one can say that all indexes perform similar to all bands are highly correlated with each other. However, selecting the different combination of bands for the index calculation results in different correlations between them. For example, the correlation coefficients between all indexes

have been computed for Lake Ray Roberts in January 2017. Figure 7 shows the resulting correlations matrix with varying correlation coefficients between 0.90 and 1.00. Despite a high correlation as it is between MNDWI and NWI but also between  $AWEI_{nsh}$ ,  $AWEI_{sh}$  and  $TC_{wet}$ , the resulting histogram separating land and water behaves differently which will be shown in detail in the next processing step. In this study, we use the advantage of combining information of five different water indexes in order to estimate only one uniform threshold for separating land and water pixels.



**Figure 7.** Correlation matrix between all indexes for Lake Ray Roberts in January 2017 (MNDWI, NWI,  $AWEI_{nsh}$ ,  $AWEI_{sh}$ ,  $TC_{wet}$ ).

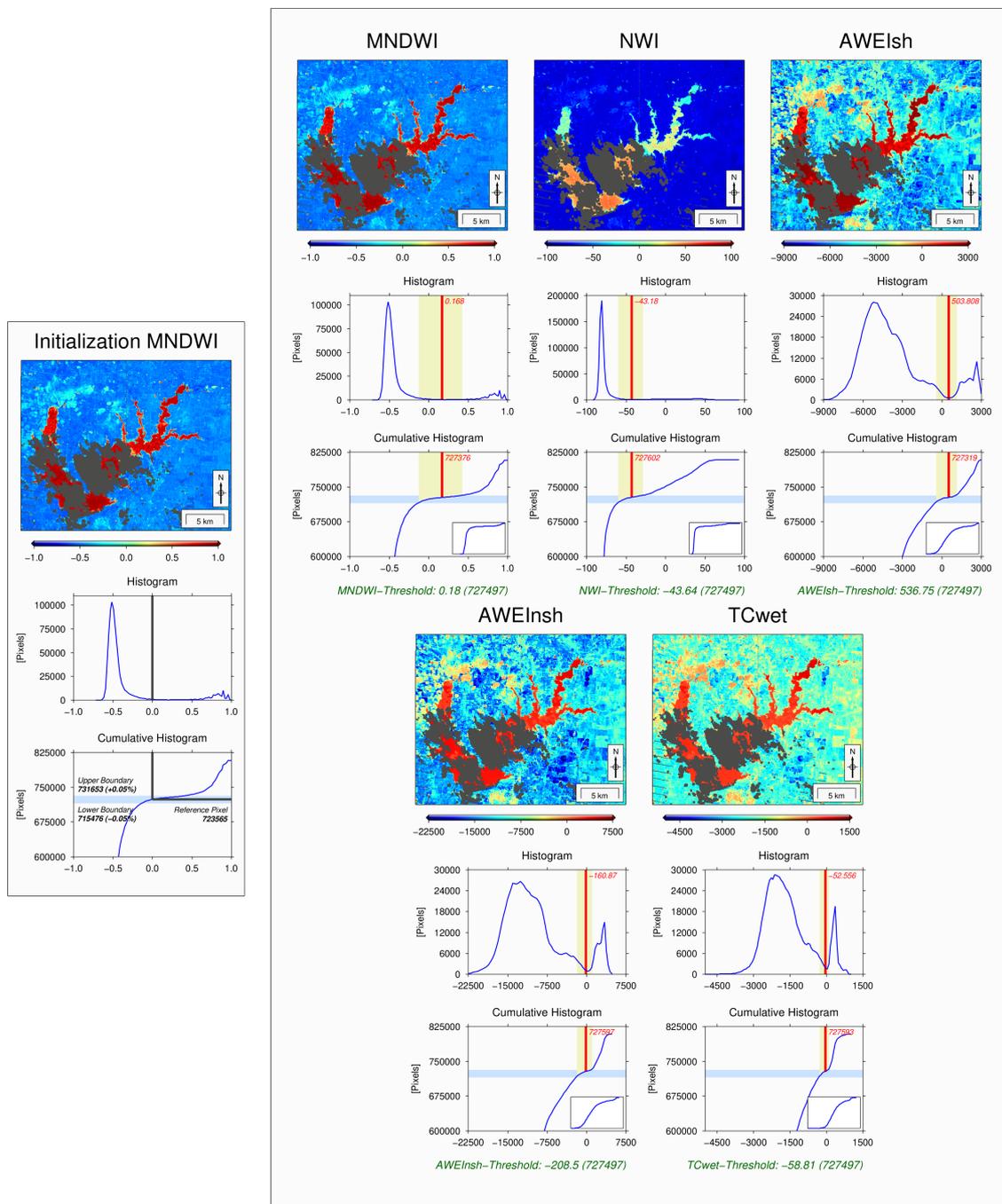
### 3.2.3. Thresholding

After computing the five water indexes, the resulting histograms show different land and water peaks in each histogram which must be separated as good as possible in this step. In general, MNDWI and NWI show dominant peaks for land on the right side of the histogram but almost none for water. For the  $AWEI_{sh}$ ,  $AWEI_{nsh}$  and  $TC_{wet}$ , the histograms show wider land accumulations on the right side and smaller water peaks on the left side. One can see in Figure 8 that the resulting histograms do not clearly have bimodal shapes. This is required for common approaches such as the Otsu method [15] that would lead to unsatisfying results in the threshold computation. Therefore, we developed a new approach for an own optimized automated threshold computation by combining the information of all five indexes.

In the initialization step, we first calculate the histogram for MNDWI. Also, a cumulative histogram is computed. Then, an initial threshold of 0.0 for the MNDWI is selected which can be used as a first rough threshold. Based on this threshold, the corresponding reference pixel (#723565) in the cumulative histogram is selected. Then, a search range for the final thresholds has to be defined. This is done by applying a manually selected boundary of  $\pm 0.05\%$  of all pixels on the reference pixel. This boundary is selected as a compromise between a large search range that can lead to ambiguous thresholds and a small search range in which the optimal threshold is not captured (per water index). This leads to the upper (#731653) and lower (#715476) boundary of the search range in the cumulative histogram. This restriction on the search range leads to an improved and more homogeneous threshold computation of the five indexes.

In the next step, the pixel boundaries are applied on all cumulative histograms of all indexes as the number of land and water pixels are equally distributed. Then, the pixel boundaries are transferred from the cumulative histograms to the normal histograms of each index. The threshold is now searched in the search range which is highlighted in yellow of Figure 8. Instead of just using the minimum within the search range, we compute a standard deviation within a sliding window because of the histogram noise. The index value at the minimum standard deviation within each search range of each index is then selected as threshold and also transferred to a pixel value in the cumulative histogram which is shown as a red line in Figure 8. Then, all five pixel-based thresholds are averaged to get only one pixel-based threshold ( $727497 \pm 124$ ) for all five indexes. The small standard deviation of 124 pixels already shows that the non-averaged thresholds are very accurate. Finally, the averaged

pixel threshold is transferred again to all histograms in order to get the final index threshold which will be used in the next step to derive the land-water masks.



**Figure 8.** Calculation of the thresholds for the land-water classification and its derived land-water masks using five water indexes shown exemplary for the Lake Ray Roberts in January 2007.

### 3.2.4. Masking

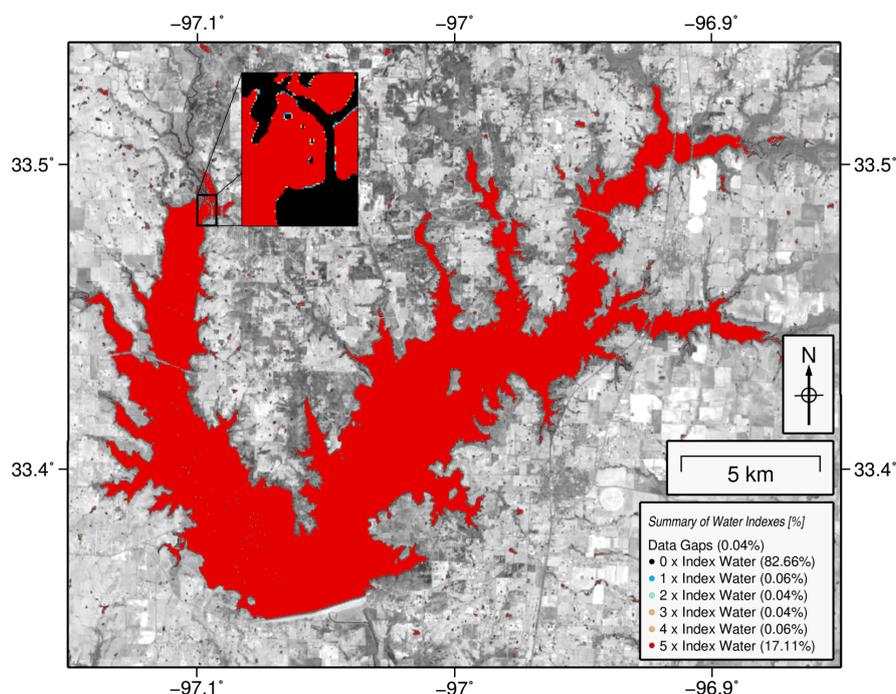
Based on the performed classification and threshold computation of the five indexes, the resulting land-water masks are now combined to achieve a final monthly land-water mask including remaining data gaps caused by voids, clouds, cloud shadows, or snow. Therefore, all five binary land-water masks (0: land, 1: water) are accumulated leading to values between 0 (land in all 5 indexes) and

5 (water in all 5 indexes) which is shown in the first column of Table 3. Figure 9 shows the result for Lake Ray Roberts in January 2017 where only pixels which were classified by at least one index as water ( $\geq 1$ ) are shown. For the case that all five indexes are classified as land, the pixels are not highlighted which affects 82.66% of the pixels. 17.11% of the pixels have been classified as water by all indexes shown in red. There are also 0.04% data gaps. Only 0.2% of all pixels were not classified uniformly by all indexes as water or land. Figure 9 clearly shows that only smaller sections near the lake shore are not uniformly classified by land or water. The reason therefore is mainly the different color in shallow water. This influence is very small as we work with a spatial resolution of about 10 m.

**Table 3.** Look up table for creation of final land-water mask and index error based on 5 water indexes.

Input Value (based on 5 Water Indexes)	Monthly Land-Water Mask	Index Error
5 Water pixels	Water (1)	No
4 Water pixels	Water (1)	Yes
3 Water pixels	Data gap	No
2 Water pixels	Data gap	No
1 Water pixel	Land (0)	Yes
0 Water pixels	Land (0)	No
Data gap	Data gap	No

Then, the final binary land-water mask with remaining data gaps is created by using Table 3. All pixels which are indicated as water by four or five indexes will be set to 1 in the final mask. All pixels which are classified as water by zero or one index are set to 0 in the final mask. In cases of disagreements between land and water where two or three indexes indicate water, the pixels are set to data gaps which will be filled in the next step. Existing data gaps are taken over in the land-water mask. In the final area error assessment, also errors caused by the indexing are considered. Therefore, pixels which do not match clearly (one or four water indexes) go into the final area error as index error.



**Figure 9.** Combination of all land-water masks using five different water indexes for Lake Ray Roberts in January 2017.

### 3.3. Calculation of a Long-Term Water Probability Mask

The filling of data gaps in all monthly files is performed by using a long-term water probability mask. Therefore, all available monthly water masks between January 1984 and June 2018 are used. In the best case, 414 monthly land-water masks are available. In our study case of Lake Ray Roberts 359 of 414 (86.7%) monthly land-water masks with data gaps are available.

The procedure for computing the long-term water probability mask is the following: First, all water pixels of all monthly masks are accumulated (Figure 10a). Secondly, all water and land pixels of all monthly masks are accumulated (Figure 10b). Then, a long-term water probability mask (Figure 10c) is computed by dividing the accumulated water pixels (Figure 10a) by the accumulated water and land pixels (Figure 10b). The water probability mask can still contain smaller surrounding water bodies which are not connected to the water body of interest and therefore cannot be used for the step of filling the data gaps. For finding the area of interest a simple connected-component labeling algorithm [35] was applied to find all connected water pixels. Figure 10d shows the 10322 labeled water bodies. Finally, we use the reference location from the initialization step to find the water body of interest (Figure 10e) which is stored as binary mask. Then, we merge the long-term water probability mask (Figure 10c) and the area of interest mask (Figure 10e). Also, we compute a dependency function between water probability and area extent shown in Figure 10f. This function is derived directly from the long-term water probability mask. For each existing water probability value, the surface area of all pixels that have a higher probability is computed in order to create the dependency function shown in Figure 10f. Both are used in the next step of filling the data gaps in monthly masks. For Lake Ray Roberts, the maximum water probability is 93.9% within the period between January 1984 and June 2018. The reason is that a dam was constructed in 1987 and filled up until 1990. Therefore, no pixel has a maximum probability of 100% which is not an error in the water classification step.

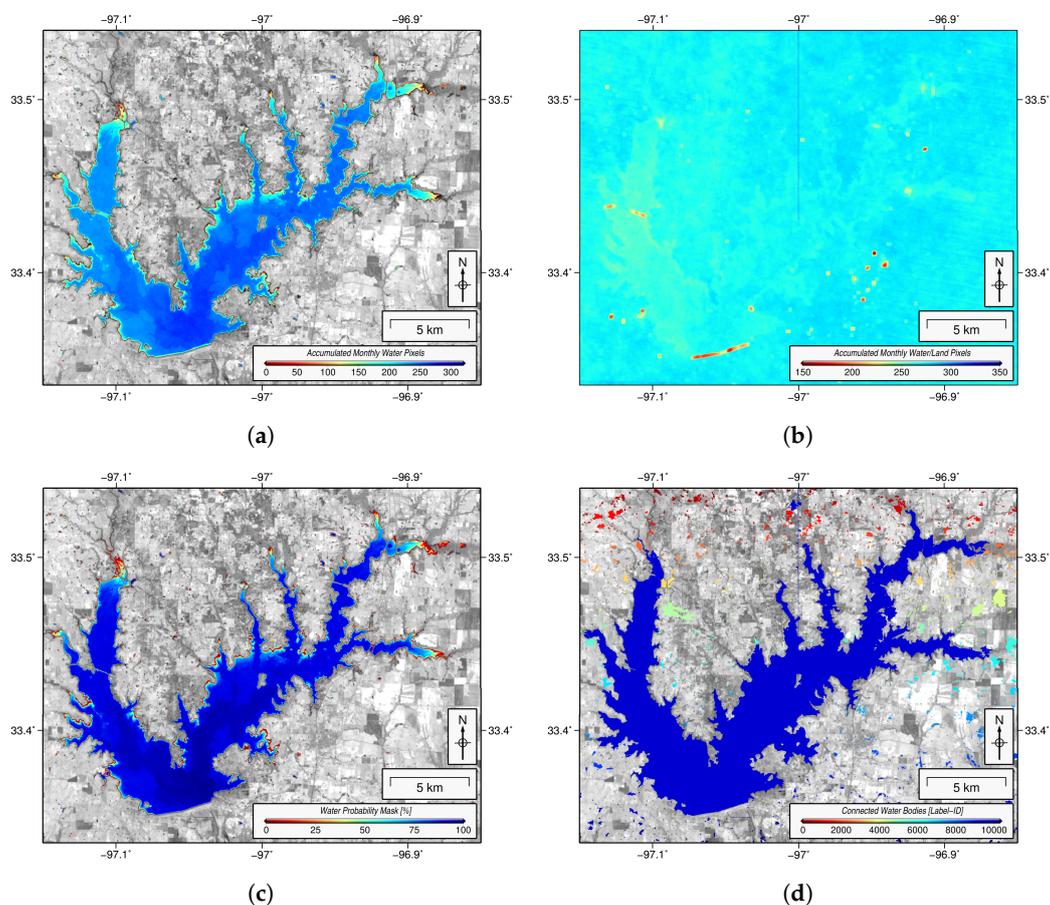
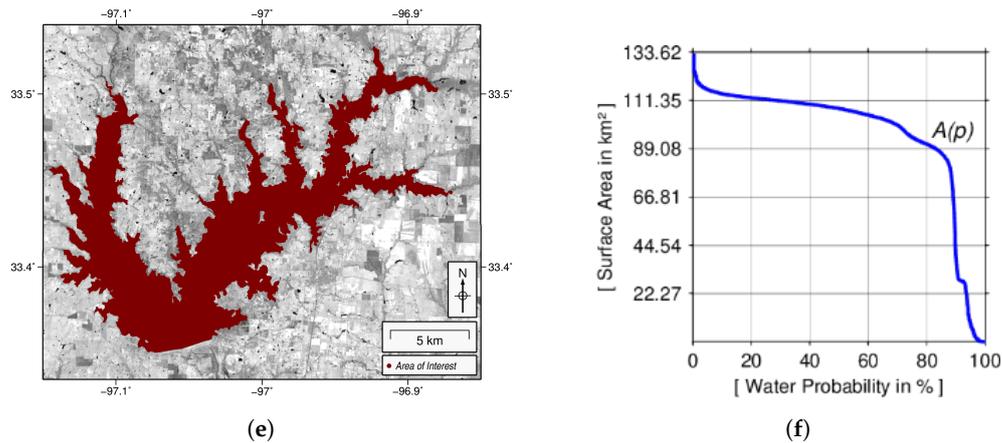


Figure 10. Cont.



**Figure 10.** (a) Number of water pixels of all accumulated monthly files (b) Number of water and land pixels of all accumulated monthly files (c) Water probability mask (d) Labeling of connected surfaces (e) Mask of interested water body (f) Dependency between water probability and area extent.

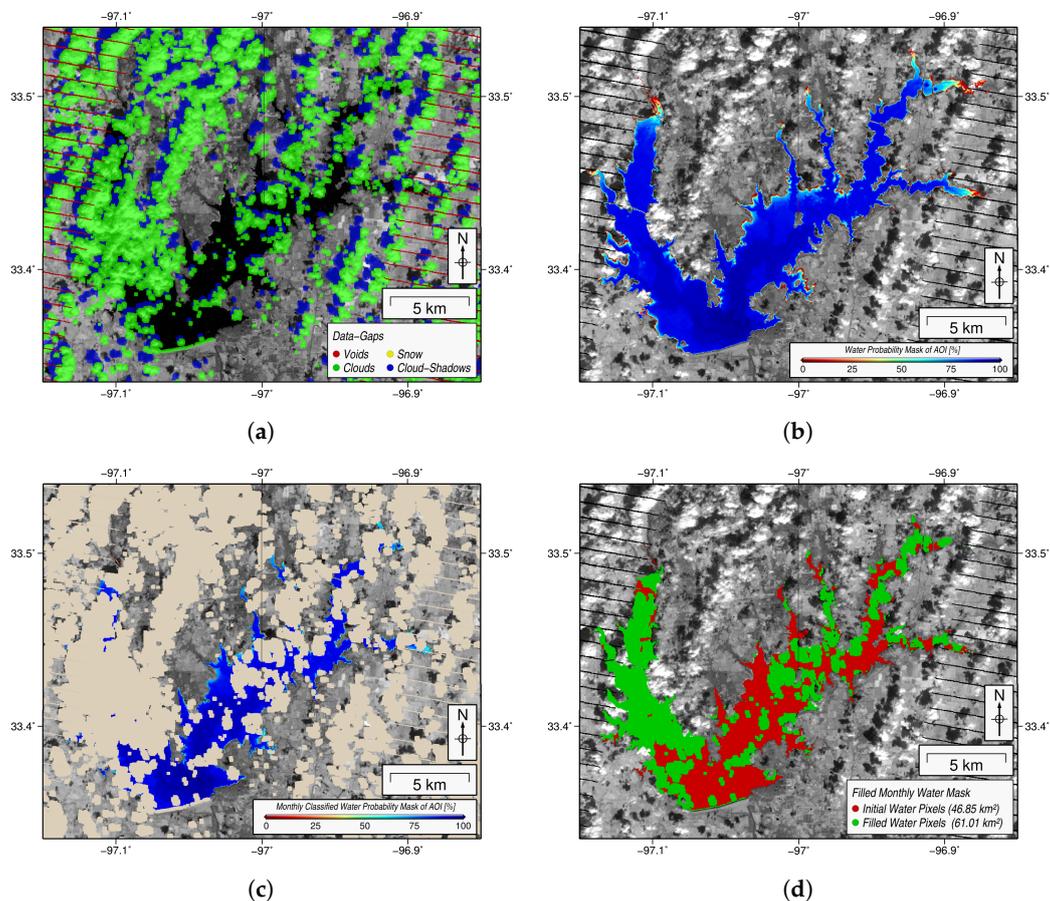
### 3.4. Filling Data Gaps of Monthly Land-Water Mask

Based on the long-term water probability mask and the area of interest (AOI) mask of Lake Ray Roberts, the remaining data gaps of all monthly mask are filled. The method for filling data gaps is exemplary demonstrated for Lake Ray Roberts in October 2012 (Figure 11a). First, the binary monthly mask (1: water, 0: land) with data gaps is multiplied with the long-term water probability mask (Figure 11b) of Lake Ray Roberts. Then, Figure 11c shows the water probability mask for all water pixels with existing data gaps. The Lake Ray Roberts has a maximum extent of 132.16 km<sup>2</sup>. The monthly water mask of October 2012 contains data gaps of voids (0.72 km<sup>2</sup>), clouds (57.26 km<sup>2</sup>) and cloud shadows (14.99 km<sup>2</sup>) which are about 55% of the maximum extent. The area of all water pixels in the AOI which are not covered by data gaps is called initial area  $A_{initial}$  (46.85 km<sup>2</sup>). The dependency between surface area and water probability shown in Figure 10f is used in the gap-filling process. It provides information about the surface area for each percentage of the long-term water probability. This function reflects the surrounding topography and is always monotone decreasing. Based on all water pixels larger than a selected percentage of water, a new mask of the land-water distribution is created. Then, this is used for filling the data gaps. Each long-term water probability  $p$  is assigned to a fixed final surface area in the curve. For example, all water pixels with a water probability larger than 60% always cover a surface area larger than 104.59 km<sup>2</sup>.

$$|(A_{initial} + A_{fill}(p)) - A(p)| \stackrel{!}{=} Min \quad (6)$$

This information is used in an iterative approach where Equation (6) is minimized. Here, the initial area  $A_{initial}$  is introduced as constant value for the current month. Then, the fill area  $A_{fill}$  is computed iteratively using different water probabilities  $p$ . Therefore, the water probability  $p$  is applied pixel-wise on the long-term water probability mask where all water pixels with a higher probability than the estimated water probability  $p$  are removed. Then, the remaining water pixels are intersected with the mask of data gaps in order to get the fill area  $A_{fill}$  which is computed again in each iteration step. Then,  $A_{fill}$  is added to the initial area  $A_{initial}$ . This results in a monthly area land-water mask and a surface area. This is subtracted by the final area  $A(p)$  derived from Figure 10f in which the same water probability  $p$  is used as for the calculation of  $A_{fill}$ . The remaining area residuals are then iteratively minimized with respect to the water probability  $p$ . Finally, the areas  $A_{initial}$  (red) and  $A_{fill}$  (green) are merged which is shown in Figure 11d. The resulting surface area  $A_{final}$  is 107.86 km<sup>2</sup> based on the water probability  $p$  of 49.0%. One can clearly see that using the optimized water probability  $p$  leads to a very good agreement at the lake shore without inconsistencies.

For filling the data gaps, the percentage or size of the data gaps in the AOI is of minor relevance. The main requirement is that parts of the lake shore are visible as this information is used in the iterative gap-filling step. One can conclude, the less data gaps near the shore are available, the better is our approach to derive correctly filled masks and accurate surface areas.



**Figure 11.** (a) Monthly mask with data gaps for Lake Ray Roberts in October 2012 (b) Long-term water probability mask  $\cap$  AOI mask (c) Long-term water probability mask  $\cap$  AOI mask  $\cap$  water pixels of monthly mask (d) Filled monthly land-water mask with  $A_{initial}$  (red) and  $A_{fill}$  (green).

Finally, an error assessment is performed for each gap-filled monthly land-water mask, respectively surface area. This is essential in order to quantify the quality of each resulting surface area. Therefore, a surface area error is computed which includes the index error from Section 3.2.4 but also error information from the gap-filling step.

The gap-filling error uses the water probability threshold  $p$  as additional information in the error assessment step. Initially, the complete filled area is assumed to be an error. But this is very pessimistic as the errors of gap-filling occur mainly near the lake shores where the probability threshold  $p$  is computed. Within the lakes or reservoirs, where the long-term water probability is increased the gap-filling error decreases. Therefore, we compute a confidence fill derived from the computed probability threshold  $p$  plus 5% which is the limit of 54%. This corresponds to a quasi-error band along the lake shores. The areas of filled water pixels with a water probability larger than 54% are summarized and used as confidence fill which is assumed to be correctly filled. The value of 60.22 km<sup>2</sup> is then subtracted from the initial area error 72.25 km<sup>2</sup> which results in a more realistic gap-filling error of 12.03 km<sup>2</sup>. Finally, we add an index error of 2.11 km<sup>2</sup> computed in the classification step where not all water indexes show the same results. The outcome of filling the data gap is a surface area with error and a land-water mask without data gaps.

### 3.5. Computation of Surface Area Time Series

In the last processing step, a surface area time series of all available months is created. Therefore, all monthly surface areas and their errors are used. In order to achieve the best surface area time series, we finally apply an outlier detection. The resulting surface area time series of Lake Ray Roberts contains monthly areas and errors between January 1984 and June 2018. In Section 4, the results of Lake Ray Roberts will be validated and compared with other data sets in detail.

## 4. Results and Validation

In this section, surface area time series based on our new approach are presented and validated. Since no in situ measurements for surface areas are available, we compare our results with water level time series as both have a linear or monotone relationship depending on the surrounding topography. Primarily, we use water level time series from in situ stations for validation. But in remote areas where no in situ data is available, we use water level time series from satellite altimetry which are available at the DAHITI (<http://dahiti.dgfi.tum.de>). Since a global validation is not possible, we focus on 32 selected study areas with different characteristics introduced in Section 4.1. Then, three of the study areas (Ray Roberts, Poço da Cruz, Tharthar) are presented and validated in detail followed by a quality assessment of all 32 investigated inland water bodies.

### 4.1. Study Areas

For the demonstration and validation of our new approach, we selected 32 globally distributed lakes or reservoirs shown in Figure 12. The major requirement for selecting inland water bodies was that there is a water level time series of a gauging station or at least from satellite altimetry available for validation. Then, the study areas were selected in different climatic regions with varying cloud coverage and rainfall causing data gaps. Also, regions with data gaps caused by ice-coverage are selected. In order to demonstrate the impact of our approach, we selected lakes and reservoirs which show variations in the water level but also in the surface area. Table 4 gives an overview of all study areas and their characteristics. In the first column, the maximum water level variations based on long-term in situ data and satellite altimetry are shown. Furthermore, the maximum surface area and shore length derived from the long-term water probability mask are presented. Finally, the ratio between surface area and shore length, climate zone, annual rainfall and annual clouds are given.

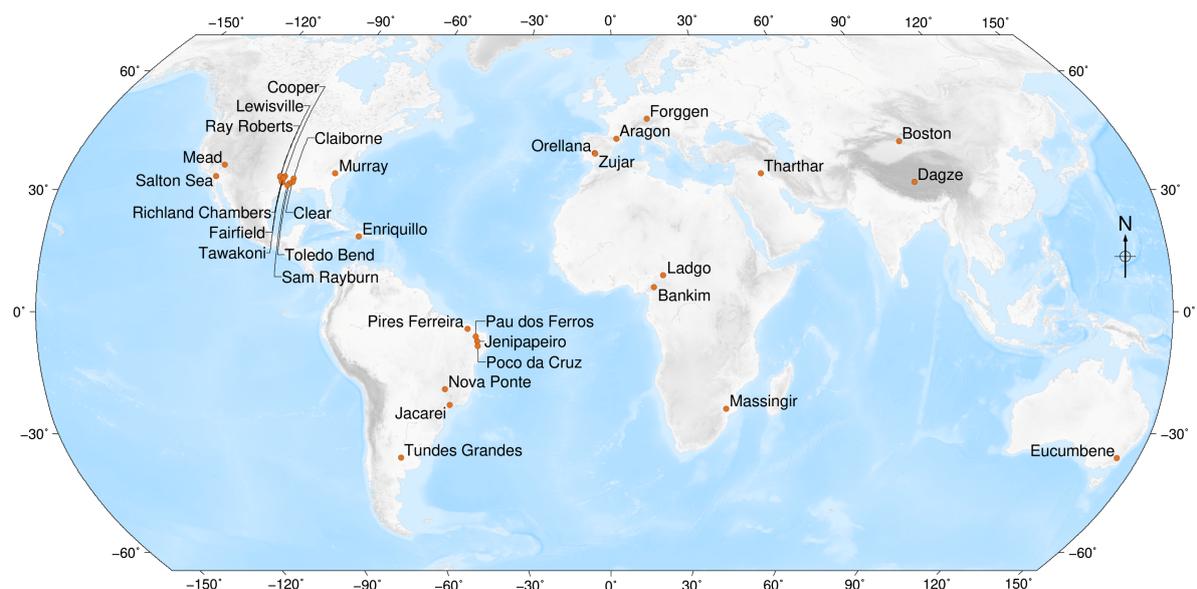


Figure 12. Map of selected and globally distributed study areas.

**Table 4.** Overview of characteristics for 32 selected study areas. For each study area, the water level variations from altimetry or gauge are given. Furthermore, the maximum surface area, length of the shore, and its ratio derived from the long-term water probability mask are shown. Additionally, information about the climate zone, annual rainfall and annual cloud coverage are shown.

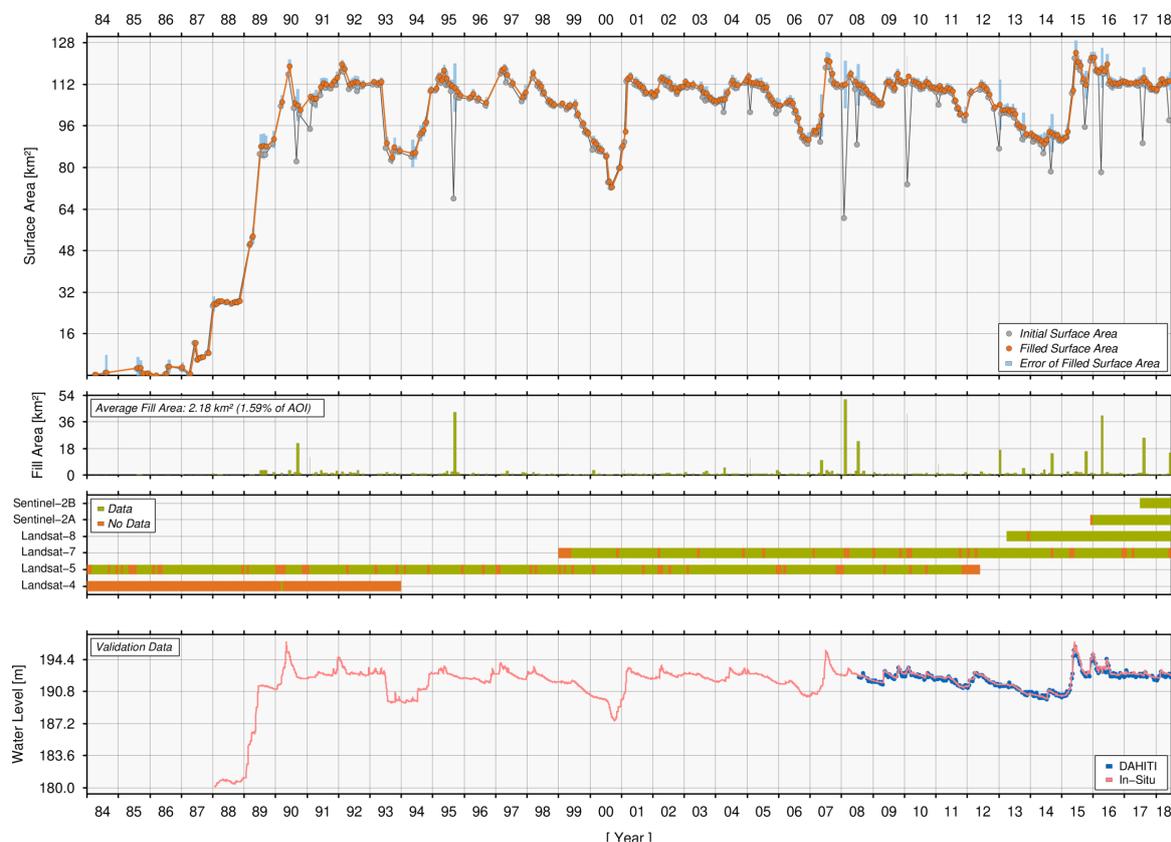
Target Name, Country (DAHITI ID)	Max. Water Level Var. [m]	Max. Surface Area [km <sup>2</sup> ]	Max. Shore Length [km]	Ratio Area/Shore	Climate Zone [36]	Annual Rainfall [37] [mm/y]	Annual Clouds [38] [%]
Aragon, Spain (10297)	33.45	22.14	93.53	0.24	Cfb	738	52
Bankim, Cameroon (3560)	14.17	336.56	2154.36	0.16	Aw	1749	53
Boston, China (226)	4.36	1226.56	1558.10	0.79	BWk	79	47
Claiborne, USA (10472)	4.30	25.54	138.31	0.18	Cfa	1340	53
Clear, USA (10496)	2.26	49.94	267.63	0.19	Cfa	1375	52
Cooper, USA (10505)	8.07	82.68	186.64	0.44	Cfa	1122	49
Dagze, China (10425)	8.12	326.71	234.66	1.39	BSk	260	47
Enriquillo, Dom. Rep. (11521)	10.12	354.14	190.65	1.86	BSh	566	35
Eucumbene, Australia (64)	37.57	140.98	458.60	0.31	Cfb	799	50
Fairfield, USA (2672)	3.05	9.84	60.23	0.16	Cfa	1021	50
Forggen, Germany (10341)	9.86	16.73	47.97	0.35	Dfb	1320	65
Jacarei, Brazil (10345)	28.22	49.39	323.28	0.15	Cfb	1453	56
Jenipapeiro, Brazil (3581)	15.76	11.05	101.89	0.11	Aw	914	59
Lagdo, Cameroon (1472)	9.33	759.42	1327.21	0.57	Aw	1006	51
Lewisville, USA (11327)	4.38	130.33	533.76	0.24	Cfa	938	44
Massingir, Mozambique (606)	20.81	189.09	249.58	0.76	BSh	472	42
Mead, USA (204)	43.56	598.29	1350.23	0.44	BWh	133	24
Murray, USA (8852)	3.92	196.46	1038.15	0.19	Cfa	1181	50
Nova Ponte, Brazil (10351)	32.98	400.05	1936.99	0.21	Aw	1540	51
Orellana, Spain (11415)	6.41	52.76	282.09	0.19	Csa	535	40
Pau dos Ferros, Brazil (8692)	13.90	11.58	67.27	0.17	Aw	853	56
Pires Ferreira, Brazil (8671)	19.91	85.05	509.66	0.17	Aw	974	57
Poço da Cruz, Brazil (8702)	24.62	58.07	382.99	0.15	BSh	552	65
Ray Roberts, USA (10146)	5.74	132.16	487.12	0.27	Cfa	990	45
Richland Chambers, USA (8814)	3.70	182.87	410.50	0.45	Cfa	1004	49
Salton Sea, USA (71)	2.78	989.59	417.88	2.37	BWh	71	22
Sam Rayburn, USA (10246)	5.97	448.34	1204.62	0.37	Cfa	1260	50
Tawakoni, USA (8813)	3.70	160.85	405.58	0.40	Cfa	1074	44
Tharthar, Iraq (122)	21.07	2491.37	1183.21	2.11	BWh	130	27
Toledo Bend, USA (10247)	4.20	659.44	1733.77	0.38	Cfa	1289	49
Tundes Grandes, Argentina (4475)	3.05	331.27	809.82	0.41	Cfa	840	45
Zujar, Spain (10301)	20.26	153.56	730.30	0.21	BSk	517	40

## 4.2. Selected Results

### 4.2.1. Ray Roberts, Lake

Lake Ray Roberts is the first study area which is presented in detail. In Section 3, Lake Ray Roberts was already selected to demonstrate our new approach for estimating surface area time series. The reservoir is located in Texas, United States of America. This reservoir has been created in the Mid 1980s by building a dam retaining the Trinity river. The surface area has been varying strongly up to a maximum of about 123 km<sup>2</sup> over the last few decades. In this area, the climate zone is defined as temperate without dry season and hot summer [36]. The annual cloud coverage is about 45% [38] and precipitation average is about 990 mm/y [37].

Figure 13 shows the resulting surface area time series (orange) and their errors (blue) of Lake Ray Roberts. Also, the initial surface areas including data gaps (gray) are shown in order to demonstrate the impact of this new approach filling the data gaps. For each month, the size of filled areas (green bars) is also given. Additionally, the availability of optical images from Landsat and Sentinel-2 is shown.



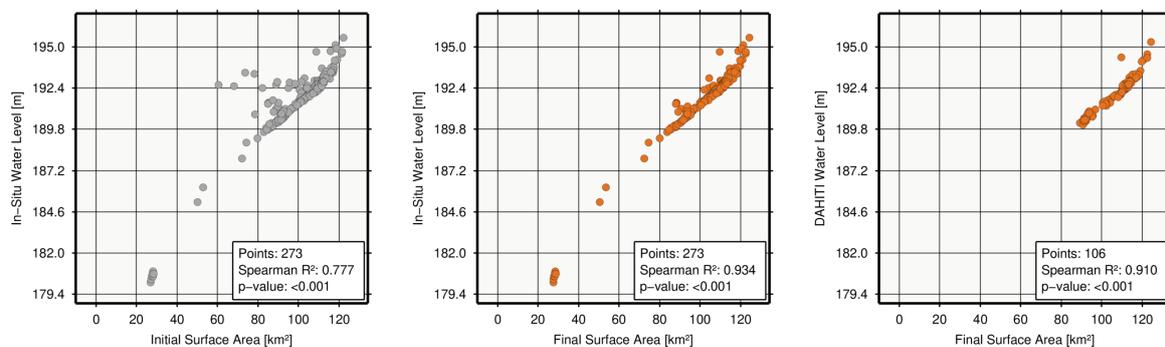
**Figure 13.** Water surface area time series (orange) and its errors (blue) of Lake Ray Roberts. Additionally, the initial surface area with data gaps (gray) and the filled area (green) are shown. Also, the data availability of monthly optical images from Landsat and Sentinel-2 is presented. Finally, the water level time series from in situ stations and DAHITI which are used for validation are shown.

In the study period between January 1984 and June 2018 theoretically 414 resulting months of surface area are possible. In the case of Lake Ray Roberts, there are 36 months (8.7%) where no optical images from any of the six satellites are available. In detail, the following monthly data of Landsat-4 (0.8%), Landsat-5 (83.6%), Landsat-7 (88.5%), Landsat-8 (98.4%), Sentinel-2A (96.8%) and Sentinel-2B (100.0%) are available. In 20 other months (4.8%), the processing of final surface areas failed. The main reason for this is that the optical images contain too big data gaps over the AOI leading to an invalid indexing or threshold computation. 358 months (86.5%) could be calculated successfully, wherein

another 68 months (16.4%) have been rejected because of too big indexing errors, cloud coverage, etc. Finally, 290 months (70.0%) of valid surface areas remain in the resulting time series.

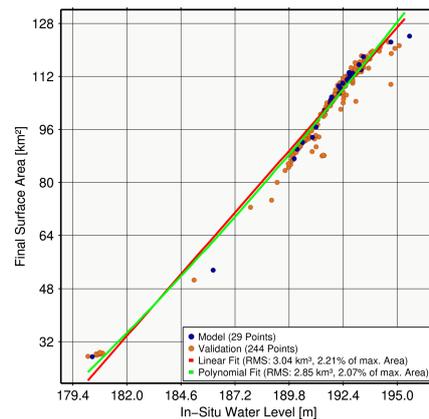
For the validation of surface area time series external water level data from in situ stations and satellite altimetry are used. In remote areas where no in situ data is available satellite altimetry provides reliable water levels for validation. On the example of Lake Ray Roberts we use both data sets for the validation to show the quality of both independent water level time series. We use the in situ data from the station Pilot Point which is provided by USGS. This time series contains 10,945 daily measurements since January 1988 which are merged on a monthly basis for validating monthly surface areas. The time series from satellite altimetry consists of 252 points and is based on data from Jason-2, Jason-3, and SARAL which have been available since July 2008. The water level time series from satellite altimetry has a lower temporal resolution than in situ data as measurements are only available when the satellite is crossing the water body. Figure 13 shows the used water level time series from the in situ station Pilot Point (red) and satellite altimetry (blue). Visual comparisons between water levels and surface areas show a very good agreement. The dam construction in the 1980s and fluctuations of the water levels (e.g., 1994, 2000, 2014) are clearly visible in both data sets.

For a more accurate quality assessment, we compute the squared Spearman correlation coefficient  $R^2$  between surface areas and water level.  $R^2$  is recommended for monotonic function as the dependency between surface area and water level is. Figure 14 shows the resulting scatter plots of the correlations. In order to demonstrate the improvement of our new approach, we compute the squared correlation coefficient between initial surface areas with data gaps and water levels from the in situ station first. The squared correlation coefficient  $R^2$  between initial surface areas and water levels from gauge is 0.777 by using 273 months of data.  $R^2$  increases up to 0.934 between the final filled surface areas and the water levels from gauge. By filling the data gaps, the quality improved by about 20%. An additional comparison between final filled surface areas and water levels from satellite altimetry yields a  $R^2$  of 0.910 which is similar to the in situ data. One can clearly see that our new approach leads to an improved surface area time series. In addition to in situ data, satellite altimetry is an alternative data source for validation of ungauged water bodies.



**Figure 14.** Validation of surface area time series for Lake Ray Roberts using water levels from in situ stations and satellite altimetry. The improvement of our new approach is shown by comparing initial surface areas with data gaps (left) and final filled surface areas (middle, right).

The validation of surface areas is a challenging task as there are no in situ data available as for water levels. In order to provide not only correlations but also RMS errors for quality assessment, we carry out an additional cross-validation based on surface areas computed from the given water levels using a fixed stage-area relationship. For this purpose, we randomly selected 10% of points where surface area and water levels are available in order to compute a hypsometry using two functions (linear and 2nd order polynom). This is shown in Figure 15 for Lake Ray Roberts. Then, the heights of the remaining 90% of points are used to predict the surface areas by using both fitted functions derived from the hypsometry. Based on them, RMS errors with respect to the final surface areas are computed.



**Figure 15.** Relation between water level and surface area (hypsoetry) used for cross-validation at Lake Ray Roberts. The blue points (10% of all points) are used for the function fitting, the orange points for the validation. The two fitted curves are plotted in red (linear fit) and green (polynom).

For Lake Ray Roberts, the cross-validation using the linear fit yields an RMS of  $3.04 \text{ km}^2$  which is 2.21% of the maximum area. By using the polynomial fit, the resulting RMS is  $2.85 \text{ km}^2$  which is 2.07% of the maximum area. Both results perform similarly, but one can clearly see for Lake Ray Roberts that the commonly used linear and polynomial function for the calculation of the hypsoetry do not clearly describe the dependency between surface areas and water levels as the influence of the surrounding topography is unpredictable.

#### 4.2.2. Poço da Cruz, Reservoir

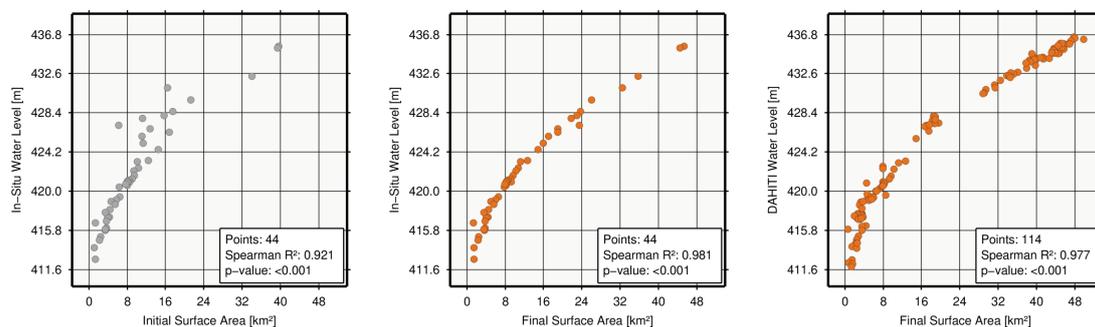
Poço da Cruz is a reservoir in Brazil, more precisely in the state Pernambuco near Ibimirim which is located about 350 km west of Recife. The climate zone around Poço da Cruz Reservoir is arid steppe with hot temperatures [36]. The averaged yearly precipitation is about  $550 \text{ mm/y}$  [37] which is about the half as for Lake Ray Roberts. On the other hand, the annual cloud coverage increased up to about 65% [38]. This leads to more cloud coverage data gaps in the optical images which have to be considered. The average monthly data gap is  $9.73 \text{ km}^2$  representing about 16% of the AOI.

Figure 16 shows the resulting surface area time series (orange) with their errors (blue) and initial surface areas including data gaps (gray) of Poço da Cruz Lake. This reservoir shows very strong variations of the surface area over the last three decades. The water level as well as the surface area vary between minimum (1998–2000), where the reservoir is almost empty, and maximum (1986, 1989, 2004, 2008) where the reservoir is filled. In between, the reservoir is subjected to strong variations. In the study period of 414 months, there are 39 months (9.4%) without any optical images available. In detail, the following monthly data of Landsat-4 (0.0%), Landsat-5 (79.5%), Landsat-7 (85.0%), Landsat-8 (100.0%), Sentinel-2A (96.8%) and Sentinel-2B (100.0%) are available. In the case of 14 months (3.4%), the processing of the final surface areas failed. For 361 months (87.2%) the final surface areas could be computed successfully, wherein afterwards 133 months (32.1%) have been rejected in the post-processing step. The final surface area time series consists of data for 228 months (55.1%). One can clearly see that the area error increased compared to Lake Ray Roberts. This is caused by the higher cloud coverage and by the shape of the lake that has much longer shores compared to its surface which also affects the gap-filling process and the resulting error.



**Figure 16.** Water surface area time series (orange) and their errors (blue) of Poço da Cruz Reservoir. Additionally, the initial surface area with data gaps (gray) and the filled area (green) are shown. Also, the data availability of monthly optical images from Landsat and Sentinel-2 is presented. Finally, the water level time series from in situ stations and DAHITI which used for validation is shown.

For the validation of surface area time series, in situ data and satellite altimetry are used. The water level time series from satellite altimetry is based on ERS-2, Envisat, SARAL, and Sentinel-3A. The time series contains 180 data points in the period between 1995 and 2018. Figure 16 shows the used water level time series from in situ data (red) and satellite altimetry (blue). Comparisons between water levels and surface areas (Figure 16) show a very good agreement with high correlations (Figure 17).

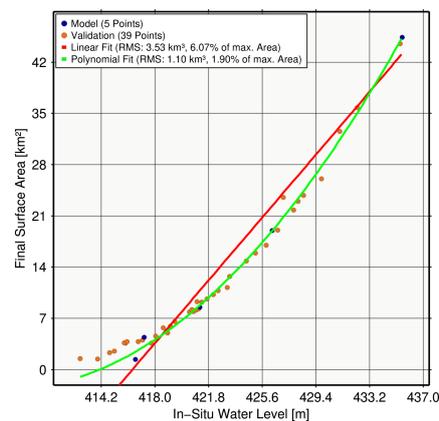


**Figure 17.** Validation of surface area time series for Poço da Cruz Lake using water levels from in situ stations and satellite altimetry. The improvement of our new approach is shown by comparing initial surface areas with data gaps (left) and final filled surface areas (middle, right).

For a more accurate quality assessment, we compute  $R^2$  between surface areas and water level. Figure 17 shows the resulting scatter plots of the correlations between surface area and in situ data, respectively satellite altimetry. For in situ data,  $R^2$  increases from 0.921 for the initial surface area (based on 44 months) to 0.981 for filled surface areas. This is a quality improvement of about 7%.  $R^2$  for

satellite altimetry also increased significantly up to 0.977. By filling the data gaps, the quality improved by about 8%. The new approach clearly shows the capability of reconstructing data gaps, especially in the period between 2004 and 2012 where more data gaps occur. In this period, the averaged fill area is  $6.25 \text{ km}^2$  which is about 11% of the AOI.

The cross-validation for Poço da Cruz Reservoir clearly shows the problematic of the function selection for the computation of the relationship between surface area and water levels. The linear and polynomial fitting by using again 10% of available points show clear differences as shown in Figure 18. The linear fit yields an RMS of  $3.53 \text{ km}^2$  which is 6.07% of the maximum area. By using the polynomial fit, the resulting RMS strongly decreases to  $1.10 \text{ km}^2$  which is only 1.90% of the maximum area. The selection of the fitting function has a strong impact on the resulting RMS as the Poço da Cruz Reservoir shows a clear polynomial dependency between surface areas and water levels.



**Figure 18.** Relation between water level and surface area (hypsoetry) used for cross-validation at Poço da Cruz Lake. The blue points (10% of all points) are used for the function fitting, the orange points for the validation. The two fitted curves are plotted in red (linear fit) and green (polynom).

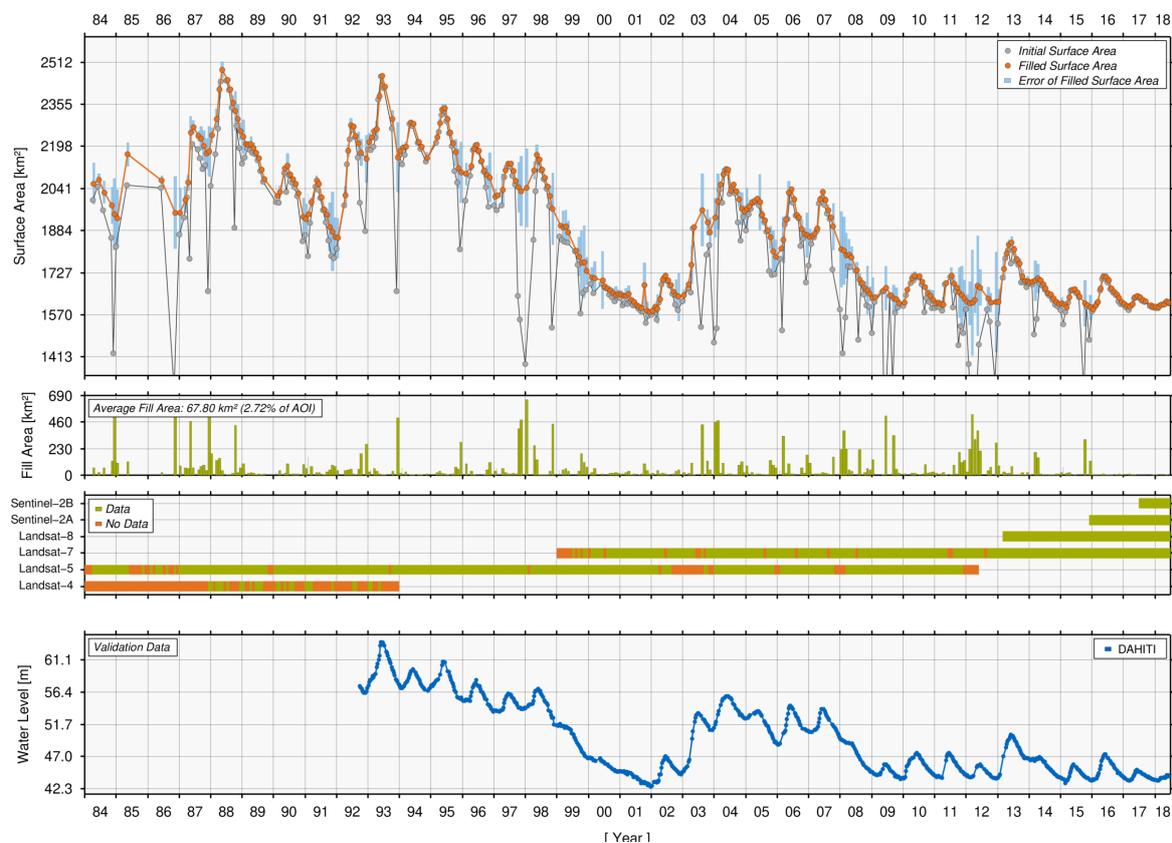
#### 4.2.3. Tharthar, Lake

Lake Tharthar located in Iraq is an artificial lake created in 1956. It is located 100 km Northwest of Baghdad between the Euphrates and Tigris rivers. The climate in this area is arid and mostly dry. The averaged yearly precipitation is about  $130 \text{ mm/y}$  [37] which is much less than for Lake Ray Roberts and Poço da Cruz Reservoir. The annual cloud coverage of 27% [38] is also decreasing. Figure 19 shows the estimated surface area time series and area errors of Lake Tharthar. Additionally, the initial surface areas with data gaps and the filled area are shown in order to demonstrate the potential of the new gap-filling approach presented in this paper. The average monthly data gap is about  $189.68 \text{ km}^2$  which is about 7% of the AOI.

Figure 19 shows the resulting surface area time series (orange) with their errors (blue) of Tharthar Lake which shows seasonal variations and long-term trends. The surface area and water level time series clearly show the drought between 1998 and 2001, wherein also the seasonal variations disappeared. In the study period of 414 months, there are 22 months (5.3%) without any optical images available. In detail, the following monthly data of Landsat-4 (21.7%), Landsat-5 (86.5%), Landsat-7 (91.0%), Landsat-8 (100.0%), Sentinel-2A (100.0%) and Sentinel-2B (100.0%) are available. The processing of final surface areas failed in only 6 months (1.4%). Further 59 monthly data sets (14.2%) have been rejected because of quality deficits which leads to a final surface area time series of 327 months (79.0%). The lower cloud coverage and annual rainfall, leads finally to a more complete time series and a smaller average area error of about 2.72% of the AOI.

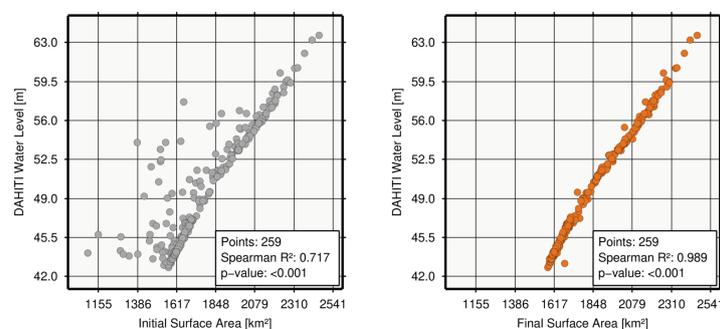
For Lake Tharthar, only satellite altimetry can be used for validation as no in situ data is available. Therefore, we use the water level time series from satellite altimetry from missions Topex, Jason-1, Jason-2, and Jason-3. The time series contains 835 data points in the period between 1992 and 2018

which are available every 10 days on average. Figure 19 shows the used water level time series from satellite altimetry (blue).



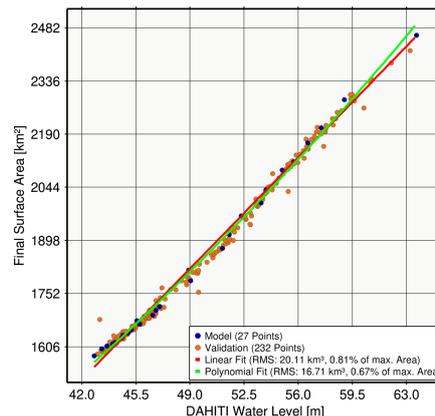
**Figure 19.** Water surface area time series (orange) and its errors (blue) of Tharthar Lake. Additionally, the initial surface area with data gaps (gray) and the filled area (green) are shown. Also, the data availability of monthly optical images from Landsat and Sentinel-2 are presented. Finally, the water level time series from DAHITI which is used for validation is shown.

Finally, we compute  $R^2$  between surface areas and water level. Figure 20 shows the resulting scatter plots of the correlations between surface area and satellite altimetry.  $R^2$  between initial surface areas and water levels from satellite is computed showing  $R^2$  0.717 by using 259 months of data. Then,  $R^2$  increases up to 0.989 between the final filled surface areas and the water levels from satellite altimetry. After filling the data gaps, the quality improved by about 38%.



**Figure 20.** Validation of surface area time series for Tharthar Lake using water levels from satellite altimetry. The improvement of our new approach is shown by comparing initial surface areas with data gaps (left) and final filled surface areas (right).

For Tharthar Lake, we also perform a cross-validation shown in Figure 21. At first glance, the dependency between surface areas and water levels seems to be mainly linear but there are also smaller variations caused by the surrounding topography included. The cross-validation for Tharthar Lake yields for the linear fit an RMS of 20.11 km<sup>2</sup> which is 0.81% of the maximum area. By using the polynomial fit, the resulting RMS slightly improves to 16.71 km<sup>2</sup> which is 0.67% of the maximum area. It can be expected that there will be a further improvement in the RMS by using a higher degree polynomial fitting. But overall, the resulting RMS are very good as they are smaller than 1.0% of the maximum surface area.



**Figure 21.** Relation between water level and surface area (hypsoetry) used for cross-validation at Tharthar, Lake. The blue points (10% of all points) are used for the function fitting, the orange points for the validation. The two fitted curves are plotted in red (linear fit) and green (polynom).

#### 4.3. Quality Assessment and Discussion

In this paper, 32 study areas shown in Table 4 have been used for validating the proposed approach. After a detailed discussion of three selected water bodies (Ray Roberts, Poço da Cruz, Tharthar), a general quality assessment of all 32 globally distributed study areas is performed in this Section.

All study areas of lakes and reservoirs have different characteristics. The long-term water level variations are between 2.26 m for Clear Lake and 43.56 m for Mead Lake. The smallest investigated inland water body is the Jenipapeiro reservoir whose surface area varies between 0.0 km<sup>2</sup> and 9.15 km<sup>2</sup>. Here, a dam has been constructed at the turn of the millennium. Lake Tharthar is the largest study area which has the strongest area variations between 1579.39 km<sup>2</sup> and 2476.11 km<sup>2</sup>. In addition to the surface area, the lake-shore lengths of the study area and its ratio are shown in Table 4. For study areas in mountainous regions which are mainly reservoirs, the ratio between surface area and shore length decreases. The ratio of natural lakes is therefore higher. For example, the shape of Salton Sea is more circular which leads to a higher ratio of 2.37 between surface area and shore length. In contrast, Jenipapeiro Lake is a reservoir located in the mountains whose topography is more frayed as the reservoir filled up the valley which leads to a longer shore line compared to its surface area. The ratio between surface area and lake shore varies between 0.11 and 2.37. In theory, the shape of lakes and reservoirs has a strong impact on the resulting quality of the surface area time series. In the land-water classification process indexing errors mainly are located at shores and not inside the lake. This fact has the potential for higher errors of reservoirs with longer shores than for natural lakes. But comparisons between ratio of surface area and shore length (Table 4) and the resulting correlations of the validation (Table 5) show no direct link.

**Table 5.** Results for 32 selected study areas. For each study area, the data availability of the Landsat and Sentinel-2 and number of resulting monthly surface area masks. Furthermore, the validation results with in situ data and satellite altimetry (Number of used months, correlation with initial surface areas with gaps and with final surface areas, improvement of the new approach ) are presented in detail. Additionally, the average area estimates for each study area is shown.

Target Name, Country (ID)	Data Availability						Monthly	Validation Gauge			Validation Altimetry			Area Error w.r.t AOI			
	L4 [%]	L5 [%]	L7 [%]	L8 [%]	S2A [%]	S2B [%]	Masks [#]	Points [#]	Initial R <sup>2</sup>	Final R <sup>2</sup>	Impr. ΔR <sup>2</sup>	Points [#]	Initial R <sup>2</sup>	Final R <sup>2</sup>	Impr. ΔR <sup>2</sup>	km <sup>2</sup>	[%]
Aragon, Spain (10297)	3.3	90.6	88.9	100.0	100.0	100.0	310		no data			125	0.698	0.823	0.125	1.21	6.08
Bankim, Cameroon (3560)	4.2	5.3	81.2	100.0	100.0	100.0	114		no data			80	0.464	0.922	0.458	30.50	9.68
Boston, China (226)	5.0	80.1	94.9	100.0	100.0	100.0	208		no data			115	0.402	0.896	0.494	38.95	3.37
Claiborne, USA (10472)	0.8	90.9	95.7	100.0	96.8	100.0	298	212	0.071	0.267	0.196 <sup>1</sup>		no data			0.66	2.93
Clear, USA (10496)	0.8	84.5	96.6	96.8	96.8	100.0	214	151	0.199	0.343	0.144 <sup>1</sup>		no data			2.52	6.31
Cooper, USA (10505)	0.8	87.1	96.2	100.0	96.8	100.0	259	172	0.779	0.940	0.161 <sup>1</sup>		no data			1.84	2.36
Dagze, China (10425)	2.5	59.8	90.6	100.0	96.8	100.0	151		no data			67	0.462	0.900	0.438	8.98	2.78
Enriquillo, Dom. Rep. (11521)	7.5	24.0	67.1	100.0	96.8	100.0	196		no data			95	0.641	0.949	0.308	14.26	4.09
Eucumbene, Australia (64)	0.0	68.9	96.6	98.4	100.0	100.0	289		no data			95	0.936	0.966	0.030	4.84	3.53
Fairfield, USA (2672)	0.8	78.6	84.6	87.3	96.8	100.0	307		no data			96	0.212	0.440	0.228	0.30	3.45
Forggen, Germany (10341)	10.8	79.5	82.5	100.0	100.0	100.0	183	179	0.595	0.770	0.175 <sup>2</sup>		no data			0.77	5.11
Jacarei, Brazil (10345)	0.8	69.8	83.3	98.4	100.0	100.0	282		no data			65	0.847	0.970	0.123	1.87	4.01
Jenipapeiro, Brazil (3581)	0.0	80.1	84.6	100.0	96.8	100.0	219		no data			82	0.751	0.967	0.216	0.68	7.47
Lagdo, Cameroon (1472)	5.8	6.2	83.3	98.4	100.0	100.0	161		no data			101	0.062	0.796	0.734	56.89	7.63
Lewisville, USA (11327)	0.8	83.9	88.5	96.8	96.8	100.0	291	265	0.736	0.949	0.213 <sup>1</sup>	36	0.410	0.672	0.262	2.93	2.43
Massingir, Mozambique (606)	6.7	41.9	79.1	96.8	100.0	100.0	240		no data			64	0.589	0.957	0.368	6.72	4.99
Mead, USA (204)	7.5	91.5	97.0	100.0	100.0	100.0	347	145	0.949	0.984	0.035 <sup>1</sup>	151	0.979	0.993	0.014	11.56	1.95
Murray, USA (8852)	0.8	92.4	94.0	95.2	96.8	100.0	302	297	0.601	0.863	0.262 <sup>1</sup>	73	0.647	0.710	0.063	4.91	2.69
Nova Ponte, Brazil (10351)	0.8	82.1	94.4	100.0	96.8	100.0	189	189	0.956	0.993	0.037 <sup>3</sup>	88	0.912	0.983	0.071	11.52	2.75
Orellana, Spain (11415)	2.5	85.9	85.0	100.0	100.0	100.0	330		no data			57	0.293	0.836	0.543	2.07	4.23
Pau dos Ferros, Brazil (8692)	0.0	69.2	70.9	98.4	100.0	100.0	201		no data			59	0.907	0.971	0.064	0.60	5.59
Pires Ferreira, Brazil (8671)	1.7	62.8	70.9	100.0	100.0	100.0	141		no data			67	0.772	0.913	0.141	6.36	8.03
Poço da Cruz, Brazil (8702)	0.0	79.5	85.0	100.0	96.8	100.0	228	44	0.921	0.981	0.060 <sup>3</sup>	114	0.912	0.977	0.065	3.23	5.56
Ray Roberts, USA (10146)	0.8	83.6	88.5	98.4	96.8	100.0	290	273	0.777	0.934	0.157 <sup>1</sup>	106	0.591	0.910	0.319	2.18	1.59
Richland Chambers, USA (8814)	2.5	91.2	96.6	98.4	96.8	100.0	289	221	0.442	0.883	0.441 <sup>1</sup>	122	0.419	0.879	0.460	2.31	1.31
Salton Sea, USA (71)	8.3	91.8	97.0	100.0	100.0	100.0	352	326	0.536	0.911	0.375 <sup>1</sup>	127	0.565	0.954	0.389	8.62	0.89
Sam Rayburn, USA (10246)	1.7	93.0	96.2	100.0	96.8	100.0	283	246	0.472	0.907	0.435 <sup>1</sup>	94	0.489	0.779	0.290	11.97	2.77
Tawakoni, USA (8813)	1.7	93.0	96.6	100.0	96.8	100.0	292	207	0.580	0.912	0.332 <sup>1</sup>	89	0.505	0.903	0.398	2.10	1.38
Tharthar, Iraq (122)	21.7	86.5	91.0	100.0	100.0	100.0	327		no data			259	0.717	0.989	0.272	67.80	2.72
Toledo Bend, USA (10247)	1.7	93.0	96.2	100.0	96.8	100.0	257	137	0.522	0.876	0.354 <sup>1</sup>	89	0.428	0.802	0.374	14.80	2.32
Tundes Grandes, Argentina (4475)	15.0	71.3	94.4	100.0	93.5	83.3	206		no data			105	0.866	0.940	0.074	13.00	4.34
Zujar, Spain (10301)	2.5	85.9	85.0	100.0	100.0	100.0	350		no data			87	0.630	0.838	0.208	4.92	3.38

<sup>1</sup> USGS, <sup>2</sup> BEA, <sup>3</sup> ANA.

Based on the resulting monthly surface area time series, we validate our results by calculation square Spearman correlation  $R^2$  with respect to water level time series from gauging station respectively satellite altimetry depending on the availability. In order to demonstrate the improvement of our new approach for filling the data gaps, we also compute  $R^2$  for all initial monthly masks with data gaps with water level data sets. Depending on the length of the different water level time series, the number of used points differs for each study area. But the comparison of initial areas and final areas is always based on the same number of data. Figure 5 shows the detailed validation of the results for all 32 study areas with water level time series from in situ data and satellite altimetry. For each study area, the number of used points, correlations coefficient  $R^2$  with initial areas, correlations coefficient  $R^2$  with final areas and the performance are given for each used validation data set.

In addition to the study areas geometry, the location and its climate condition it is expected to have an impact on the results. All study areas are located in different climate zones which are shown in Table 4. The climate zones are defined mainly through precipitation and temperature which finally leads to a different cloud coverage and rainfall. In the data processing, a higher cloud coverage leads to more data gaps which have to be reconstructed. In our study, the annual cloud coverage varies between 22% for Salton Sea and 65% for Poço da Cruz Lake and Forgggen Lake. The annual rainfall is not directly connected to cloud coverage as this depends also on the temperature. The annual rainfall varies between 71 mm/y for Salton Sea up to 1749 mm/y for Bankim Lake.

Beside lake geometry and climate conditions, the data holding and its availability can be another limiting factor in the processing of the surface area time series. Depending on the data availability, the length and quality of the resulting surface area time series performs differently. The data availability of Landsat and Sentinel scenes varies globally [39]. For example, not all taken Landsat-5 scenes have been finally processed by the agencies and made available yet. In particular, in Africa, the lack of Landsat-5 scenes can be seen for Lake Bankim and Lake Lagdo located in Cameroon/Africa where only around 6% of all Landsat 5 images are available. Table 5 gives an overview of the used data and its availability for the six missions for each study area. The average data availability for Landsat-4 (3.18%), Landsat-5 (66.63%), Landsat-7 (88.51%), Landsat-8 (98.86%), Sentinel-2A (98.19%) and Sentinel-2B (99.48%) increases the newer the satellite mission is. For the study areas, one can clearly see that data availability especially for Landsat-5 and Landsat-7 varies with respect to the location. In this study, the time span is defined between January 1984 and June 2018. In the best case, the resulting surface area time series can contain 414 monthly data points. Depending on different factors such as input data, data gaps, land-water indexing, and threshold computation, the number of monthly data points decreases in reality. For the selected study areas, the resulting number of monthly masks varies between 114 for Lake Bankim and 352 for Salton Sea.

For validating 15 of 32 study areas water level time series from gauging stations are used. All in situ measurements are given on a daily basis which are averages on a monthly basis as the monthly surface area time series is based on various daily scenes. For all study areas, the performance of the new gap-filling approach shows improvements of the correlation coefficient  $R^2$  between initial and final area time series. The correlation coefficient  $R^2$  improved between 0.035 and 0.441 for all comparisons with in situ data. In most cases, the average correlation coefficient  $R^2$  increases from 0.610 for initial areas to 0.834 which is an improvement of about 37%. In this study, Lake Claiborne and Lake Clear have a very low correlation coefficient but still a good improvement. The reason is that both reservoirs have only low variations in height and area except of a few months in the period of three decades. On the other hand, Lake Mead ( $\Delta R^2$ : 0.035) and Poço da Cruz ( $\Delta R^2$ : 0.060) only show small improvements. Here, the initial masks are already very good with correlations larger than 0.9.

We use water level time series from satellite altimetry for validation in remote areas where no in situ data is available as it can be used globally as long as the lake or reservoir is crossed by a satellite track. Compared to in situ data, the quality of water level time series from satellite altimetry is lower. They provide cm-accuracy for large lakes and dm-accuracy for smaller lakes [31]. Here, the initial resolution varies between about 10 and 35 days depending on the used altimeter missions which

are merged on a monthly basis for validation. In this study, water level time series from satellite altimetry are used for 28 study areas crossed by at least one satellite track. The performance of the new gap-filling approach shows improvements of the correlation coefficient  $R^2$  between initial and final area time series for all study areas also for satellite altimetry. The correlation coefficient  $R^2$  improved by between 0.014 and 0.734 for 28 comparisons with water level time series from satellite altimetry. Overall, the average correlation coefficient  $R^2$  increased from 0.611 for initial areas to 0.880 which is an improvement of about 44%. Despite different targets, the performance is similar by using in situ data for validation. The behavior of the correlation coefficient  $R^2$  resulting from satellite altimetry is comparable with correlation coefficients  $R^2$  using in situ data even though the lengths of the validation time series differ between both data sets.

One can conclude that validating surface area time series using in situ or satellite altimetry provides reliable results in both cases. In order to demonstrate the potential of the gap-filling approach one can say that an improvement of about 37–44% can be achieved. In general, the best improvement can be achieved when the initial correlation coefficients  $R^2$  with water levels are between 0.4 and 0.6. For those cases, the final correlations have been increased to more than 0.8 or even 0.9 for almost all study areas.

In addition to the validation using external data sets, an area error is given for each month. This area error contains a composite of uncertainties from the land-water indexing and the gap-filling step. All pixels which are not clearly identified as water or land within the AOI go into the final error. Moreover, all pixels which are filled in the gap-filling step and have a smaller long-term water probability than the computed threshold plus 5% go also into the error. This usually affects pixels which are located near the lake shore as pixels with a higher water probability are located within the lake or reservoir. This leads to an average area error between 1.31% and 9.68% of the AOI for all 32 study areas. Additionally, comparisons showed that there is no direct dependency between final correlation coefficients  $R^2$  and the resulting surface areas.

To provide RMS values for validation, we performed a cross-validation for all study areas. This has been already shown in detail for the study areas Ray Roberts, Poço da Cruz and Tharthar. The methodology is described in Section 4.2.1. It is based on a relationship function between water levels and surface areas. Table 6 shows the cross-calibration results for all study areas. Moreover, it provides the number of points used for the regression and the number of points used for the validation of each study area. As can be seen from the table, the relative area errors remain below 8% for all 32 targets. The absolute RMS errors varies between less than 1 km<sup>2</sup> and about 40 km<sup>2</sup>. Of course these values not only include the area uncertainties but may also reflect errors coming from the water level measurements or the assumed hypsometry function. When comparing both fittings (linear and polynomial) one can clearly see the influence of this choice. The optimal fitting function depends on the lakes (unknown) bathymetry. Of course, the results will improve when more data points are used for the fitting process. However, then the validation is no longer done by independent input data.

When comparing the cross-validation results with the internal errors derived in the AWAX approach (provided in Table 5) a very good agreement can be seen. For Lake Ray Roberts, the area error derived from the AWAX approach is 2.18 km<sup>2</sup> (1.59%) where the cross-validation yields a RMS of 3.04 km<sup>2</sup> (2.21%) for the linear fit, respectively 2.85 km<sup>2</sup> (2.07%) for the polynomial fit. Another example is Poço da Cruz, Lake where the linear and polynomial fit behaves differently. Here, the cross-validation shows an RMS of 3.53 km<sup>2</sup> (6.07%) for the linear fit, respectively 1.10 km<sup>2</sup> (1.90%) for the polynomial fit. The surface area error derived from the AWAX approach is 3.23 km<sup>2</sup> (5.56%) which is within both results of the cross-validation. The external validation thus confirms the order of magnitude of the internal AWAX errors and proves their high reliability.

**Table 6.** Results of the cross-validation of 32 selected study areas. For each study area, the used number of points for the calculation of the hypsometry and the number of points for the validation are shown. Furthermore, the resulting RMS and relative RMS with respect to the maximum surface area using a linear and polynomial of degree 2 are presented.

Target Name, Country (ID)	Hypsometry (Linear)				Hypsometry (2nd Deg.)	
	Model Points [#]	Validation Points [#]	RMS [km <sup>2</sup> ]	wrt. Area [%]	RMS [km <sup>2</sup> ]	wrt. Area [%]
Aragon, Spain (10297)	14	111	1.14	5.16	1.14	5.16
Bankim, Cameroon (3560)	9	71	21.30	6.33	19.26	5.72
Boston, China (226)	13	102	18.70	1.52	18.20	1.48
Claiborne, USA (10472)	23	189	0.65	2.53	1.07	4.20
Clear, USA (10496)	16	135	3.92	7.85	3.74	7.48
Cooper, USA (10505)	19	153	2.03	2.48	1.94	2.35
Dagze, China (10425)	7	60	3.54	1.08	3.68	1.13
Enriquillo, Dom. Rep. (11521)	11	84	13.97	3.94	11.59	3.27
Eucumbene, Australia (64)	11	84	2.58	1.83	2.67	1.90
Fairfield, USA (2672)	11	85	0.28	2.85	0.28	2.88
Forggen, Germany (10341)	19	160	0.83	4.94	0.84	5.02
Jacarei, Brazil (10345)	8	57	1.56	3.17	1.75	3.55
Jenipapeiro, Brazil (3581)	10	72	0.26	2.33	0.26	2.33
Lagdo, Cameroon (1472)	11	90	39.19	5.16	40.07	5.28
Lewisville, USA (11327)	28	237	2.85	2.19	2.42	1.86
Massingir, Mozambique (606)	8	56	6.30	3.33	4.03	2.13
Mead, USA (204)	16	129	5.21	0.87	3.97	0.66
Murray, USA (8852)	31	266	2.64	1.34	2.34	1.19
Nova Ponte, Brazil (10351)	20	169	6.07	1.52	5.47	1.37
Orellana, Spain (11415)	7	50	1.26	2.39	1.26	2.38
Pau dos Ferros, Brazil (8692)	7	52	0.80	6.89	0.51	4.40
Pires Ferreira, Brazil (8671)	7	60	5.45	6.41	6.80	7.99
Poço da Cruz, Brazil (8702)	5	39	3.53	6.07	1.10	1.90
Ray Roberts, USA (10146)	30	255	2.88	2.18	2.74	2.07
Richland Chambers, USA (8814)	23	198	2.34	1.28	2.19	1.20
Salton Sea, USA (71)	34	292	4.58	0.46	4.52	0.46
Sam Rayburn, USA (10246)	26	220	10.07	2.25	5.96	1.33
Tawakoni, USA (8813)	22	185	2.74	1.70	2.36	1.47
Tharthar, Iraq (122)	27	232	20.11	1.59	16.71	1.21
Toledo Bend, USA (10247)	15	122	10.48	0.81	7.96	0.67
Tundes Grandes, Argentina (4475)	12	93	8.43	2.55	8.11	2.45
Zujar, Spain (10301)	10	77	5.75	3.74	5.75	3.75

Further investigations also showed that there are no clear dependencies between surface area, respectively final correlation coefficients  $R^2$  with other parameters such as lake shore, climate zone, annual cloud coverage, or annual rainfall. Finally, one can conclude that this new approach can be used for nearly all lakes and reservoirs as they are large enough to be observed successfully by Landsat or Sentinel-2.

## 5. Conclusions

This paper presents a new approach for the automated extraction of consistent time-variable water surfaces of lakes and reservoirs using optical images from Landsat and Sentinel-2. It is based on a careful data pre-processing step creating monthly combined Landsat and Sentinel-2 scenes which are used as input for computing land-water masks. This is performed by estimating five established water indexes namely Modified Normalized Difference Water Index (MNDWI), New Water Index (NWI), Automated Water Extraction Index for Non-Shadow Areas ( $AWEI_{nsh}$ ), Automated Water Extraction Index for Shadow Areas ( $AWEI_{sh}$ ) and Tasseled Cap for Wetness ( $TC_{wet}$ ). Then, an automated threshold separating land and water is computed based on mutual integration of five water indexes. All monthly land-water masks with data gaps caused by clouds, cloud shadows, snow/ice and voids are merged in order to compute a long-term water probability mask for the time period between January 1984 and June 2018 for the investigated inland water body. Based on the long-term water probability mask, all remaining data gaps in the monthly masks are filled using an iterative threshold approach. The results of this new method are gap-less surface area time series for lakes and reservoirs and their reliable area errors. The data holding of the DAHITI has been extended by that new product named surface area time series presented in this study.

The performance of this new approach is assessed for 32 globally distributed lakes and reservoirs by comparison with water level variations from gauges or satellite altimetry. The validation shows no direct connection between surface area accuracy and the shape of the water body and the cloud coverage even if the internal errors increased in these cases. Calculating the squared Spearman correlation  $R^2$  between gap-filled surface area time series with respect to water level from in situ data and satellite altimetry leads to an average  $R^2$  of 0.834 and 0.880, respectively. Compared to the initial land-water masks with data gaps this improved by 37% and 44% respectively. The performed cross-validation shows RMS errors smaller than 40 km<sup>2</sup> for all study areas. This corresponds to relative errors below 8% for all 32 targets. These values are in the same order of magnitude than the internal surface area errors derived from the AWAX approach.

The new method of computing gap-less surface area time series provides valuable results with a good quality for hydrological applications. But there is still potential for improvements and challenges for the future. In this study, the resulting surface area time series have a monthly temporal resolution. This can be increased by using daily scenes which can be also incomplete as the new approach can also fill images where only a part of the water body is covered. But this will be very challenging for the data management as the required data storage will increase dramatically. There are also possibilities to reduce the data gaps in the monthly scenes which will improve the quality of the resulting surface area time series. This can be achieved by introducing for example MODIS data which are available in lower spatial but higher temporal resolution. But also radar images from Sentinel-1 which are not affected by clouds can be used for creating improved land-water masks as an additional data set especially in very cloudy regions. Our approach has the capability to integrate further diverse data sets as long as they are reliable land-water masks. Finally, one can conclude that this is a very important data set for hydrological applications. Surface area time series have also the potential to be used for the creation of enhanced products such as storage variations of lakes and reservoirs. This method can also slightly be modified in order to derive changes of river widths from land-water masks and help in deriving river discharge.

## 6. Data Availability

All presented surface area time series and water level time series from satellite altimetry used for the validation as well as results for many additional targets are freely available in the DAHITI at <http://dahiti.dgfi.tum.de>. The supplementary material of this paper also contains results of all 32 study areas where surface area time series, used optical data, long-term water probability maps and scatter plots of the validation are presented in detail.

**Supplementary Materials:** Supplementary Materials are available online at <http://www.mdpi.com/2072-4292/11/9/1010/s1>.

**Author Contributions:** D.S. developed a first version of the approach in his bachelor thesis. C.S. improved the land-water classification, developed the gap-filling approach, performed the validation and wrote the paper. D.D. contributed to the manuscript writing and helped with the discussions of the applied methods and results.

**Funding:** This research received no external funding. The APC was funded by the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

**Acknowledgments:** We thank the Agência Nacional de Águas (ANA); US Geological Survey (USGS) and Bavarian Environmental Agency (BEA) for providing the in situ data. We thank three anonymous reviewers for their valuable comments that helped to improve the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Shiklomanov, I. World Fresh Water Resources. In *Water in Crisis—A Guide to the World's Fresh Water Resources*; Gleick, P., Ed.; Oxford University Press: Oxford, UK, 1993; Chapter 2; pp. 13–23.
- Guha-Sapir, D.; Vos, F. Quantifying Global Environmental Change Impacts: Methods, Criteria and Definitions for Compiling Data on Hydro-meteorological Disasters. In *Coping with Global Environmental Change, Disasters and Security: Threats, Challenges, Vulnerabilities and Risks*; Brauch, H.G., Oswald Spring, U., Mesjasz, C., Grin, J., Kameri-Mbote, P., Chourou, B., Dunay, P., Birkmann, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 693–717. [[CrossRef](#)]
- Hasan, M.; Moody, A.; Benninger, L.; Hedlund, H. How war, drought, and dam management impact water supply in the Tigris and Euphrates Rivers. *Ambio* **2018**. [[CrossRef](#)] [[PubMed](#)]
- Global Runoff Data Center: Statistics Based on GRDC Station Catalogue from 2018-10-19. Available online: [https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/grdc\\_stations\\_ftp.html?nn=201352](https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/grdc_stations_ftp.html?nn=201352) (accessed on 25 January 2019).
- Memon, A.A.; Muhammad, S.; Rahman, S.; Haq, M. Flood monitoring and damage assessment using water indices: A case study of Pakistan flood-2012. *Egypt. J. Remote Sens. Space Sci.* **2015**, *18*, 99–106. [[CrossRef](#)]
- Amani, M.; Salehi, B.; Mahdavi, S.; Brisco, B. Spectral analysis of wetlands using multi-source optical satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2018**, *144*, 119–136. [[CrossRef](#)]
- DeVries, B.; Huang, C.; Lang, M.W.; Jones, J.W.; Huang, W.; Creed, I.F.; Carroll, M.L. Automated Quantification of Surface Water Inundation in Wetlands Using Optical Satellite Imagery. *Remote Sens.* **2017**, *9*, 807. [[CrossRef](#)]
- Singh, A.; Seitz, F.; Schwatke, C. Inter-annual water storage changes in the Aral Sea from multi-mission satellite altimetry, optical remote sensing, and GRACE satellite gravimetry. *Remote Sens. Environ.* **2012**, *123*, 187–195. [[CrossRef](#)]
- Ghanavati, E.; Firouzabadi, P.Z.; Jangi, A.A.; Khosravi, S. Monitoring geomorphologic changes using Landsat TM and ETM+ data in the Hendijan River delta, southwest Iran. *Int. J. Remote Sens.* **2008**, *29*, 945–959. [[CrossRef](#)]
- Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033. [[CrossRef](#)]
- Ding, F. Study on Information Extraction of Water Body with a New Water Index (NWI). *Sci. Surv. Mapp.* **2009**, *34*, 155–157.
- Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [[CrossRef](#)]

13. Kauth, R.; Thomas, G. The Tasseled Cap—A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, IN, USA, 29 June–1 July 1976; Volume 4B, pp. 41–51.
14. Gomez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [[CrossRef](#)]
15. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
16. Feng, M.; Sexton, J.O.; Channan, S.; Townshend, J.R. A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic-spectral classification algorithm. *Int. J. Digit. Earth* **2016**, *9*, 113–133. [[CrossRef](#)]
17. Pekel, J.F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422. [[CrossRef](#)]
18. Klein, I.; Gessner, U.; Dietz, A.J.; Kuenzer, C. Global WaterPack—A 250m resolution dataset revealing the daily dynamics of global inland water bodies. *Remote Sens. Environ.* **2017**, *198*, 345–362. [[CrossRef](#)]
19. Markert, K.N.; Chishtie, F.; Anderson, E.R.; Saah, D.; Griffin, R.E. On the merging of optical and SAR satellite imagery for surface water mapping applications. *Results Phys.* **2018**, *9*, 275–277. [[CrossRef](#)]
20. Zhang, S.; Gao, H.; Naz, B.S. Monitoring reservoir storage in South Asia from multisatellite remote sensing. *Water Resour. Res.* **2014**, *50*, 8927–8943. [[CrossRef](#)]
21. Busker, T.; de Roo, A.; Gelati, E.; Schwatke, C.; Adamovic, M.; Bisselink, B.; Pekel, J.F.; Cottam, A. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 669–690. [[CrossRef](#)]
22. Irons, J.R.; Dwyer, J.L.; Barsi, J.A. The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sens. Environ.* **2012**, *122*, 11–21. [[CrossRef](#)]
23. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
24. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [[CrossRef](#)]
25. ESA. *SENTINEL-2 ESA's Optical High-Resolution Mission for GMES Operational Services, SP-1322/2*; Technical Report; ESA: Paris, France, 2012.
26. ESA. *Sen2Cor—Configuration and User Manual, S2-PDGS-MPC-L2A-SUM-V2.5.5*; Technical Report; ESA: Paris, France, 2018.
27. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277. [[CrossRef](#)]
28. Da Silva, J.S.; Calmant, S.; Seyler, F.; Filho, O.C.R.; Cochonneau, G.; Mansur, W.J. Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions. *Remote Sens. Environ.* **2010**, *114*, 2160–2181. [[CrossRef](#)]
29. Dettmering, D.; Schwatke, C.; Boergens, E.; Seitz, F. Potential of ENVISAT Radar Altimetry for Water Level Monitoring in the Pantanal Wetland. *Remote Sens.* **2016**, *8*, 596. [[CrossRef](#)]
30. Crétaux, J.F.; Jelinski, W.; Calmant, S.; Kouraev, A.; Vuglinski, V.; Bergé-Nguyen, M.; Gennero, M.C.; Nino, F.; Rio, R.A.D.; Cazenave, A.; et al. SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data. *Adv. Space Res.* **2011**, *47*, 1497–1507. [[CrossRef](#)]
31. Schwatke, C.; Dettmering, D.; Bosch, W.; Seitz, F. DAHITI—An innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 4345–4364. [[CrossRef](#)]
32. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [[CrossRef](#)]
33. Haibo, Y.; Zongmin, W.; Hongling, Z.; Yu, G. Water Body Extraction Methods Study Based on RS and GIS. *Procedia Environ. Sci.* **2011**, *10 Pt C*, 2619–2624. [[CrossRef](#)]
34. Crist, E. A TM Tasseled Cap Equivalent Transformation for Reflectance Factor Data. *Remote Sens. Environ.* **1985**, *17*, 301–306. [[CrossRef](#)]

35. Rosenfeld, A.; Pfaltz, J.L. Sequential Operations in Digital Picture Processing. *J. ACM* **1966**, *13*, 471–494. [[CrossRef](#)]
36. Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* **2018**, *5*. [[CrossRef](#)]
37. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [[CrossRef](#)]
38. Wilson, A.M.; Jetz, W. Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. *PLoS Biol.* **2016**, *14*, 1–20. [[CrossRef](#)] [[PubMed](#)]
39. Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The global Landsat archive: Status, consolidation, and direction. *Remote Sens. Environ.* **2016**, *185*, 271–283. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).