

Comparing thresholding with machine learning classifiers for mapping complex water

Bangira, Tsitsi; Alfieri, Silvia Maria; Menenti, Massimo; van Niekerk, Adriaan

DOI

[10.3390/rs11111351](https://doi.org/10.3390/rs11111351)

Publication date

2019

Document Version

Final published version

Published in

Remote Sensing

Citation (APA)

Bangira, T., Alfieri, S. M., Menenti, M., & van Niekerk, A. (2019). Comparing thresholding with machine learning classifiers for mapping complex water. *Remote Sensing*, 11(11), Article 1351. <https://doi.org/10.3390/rs11111351>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Article

Comparing Thresholding with Machine Learning Classifiers for Mapping Complex Water

Tsitsi Bangira ^{1,2,*} , Silvia Maria Alfieri ², Massimo Menenti ^{2,3} and Adriaan van Niekerk ¹ 

¹ Department of Geography and Environmental Studies, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa; avn@sun.ac.za

² Department of Geoscience and Remote Sensing, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands; S.M.alfieri-1@tudelft.nl (S.M.A.); M.Menenti@tudelft.nl (M.M.)

³ State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China

* Correspondence: T.Bangira@tudelft.nl or tsitsibangira@gmail.com; Tel.: +27-73-546-8920

Received: 11 February 2019; Accepted: 30 May 2019; Published: 5 June 2019



Abstract: Small reservoirs play an important role in mining, industries, and agriculture, but storage levels or stage changes are very dynamic. Accurate and up-to-date maps of surface water storage and distribution are invaluable for informing decisions relating to water security, flood monitoring, and water resources management. Satellite remote sensing is an effective way of monitoring the dynamics of surface waterbodies over large areas. The European Space Agency (ESA) has recently launched constellations of Sentinel-1 (S1) and Sentinel-2 (S2) satellites carrying C-band synthetic aperture radar (SAR) and a multispectral imaging radiometer, respectively. The constellations improve global coverage of remotely sensed imagery and enable the development of near real-time operational products. This unprecedented data availability leads to an urgent need for the application of fully automatic, feasible, and accurate retrieval methods for mapping and monitoring waterbodies. The mapping of waterbodies can take advantage of the synthesis of SAR and multispectral remote sensing data in order to increase classification accuracy. This study compares automatic thresholding to machine learning, when applied to delineate waterbodies with diverse spectral and spatial characteristics. Automatic thresholding was applied to near-concurrent normalized difference water index (NDWI) (generated from S2 optical imagery) and VH backscatter features (generated from S1 SAR data). Machine learning was applied to a comprehensive set of features derived from S1 and S2 data. During our field surveys, we observed that the waterbodies visited had different sizes and varying levels of turbidity, sedimentation, and eutrophication. Five machine learning algorithms (MLAs), namely decision tree (DT), k-nearest neighbour (k-NN), random forest (RF), and two implementations of the support vector machine (SVM) were considered. Several experiments were carried out to better understand the complexities involved in mapping spectrally and spatially complex waterbodies. It was found that the combination of multispectral indices with SAR data is highly beneficial for classifying complex waterbodies and that the proposed thresholding approach classified waterbodies with an overall classification accuracy of 89.3%. However, the varying concentrations of suspended sediments (turbidity), dissolved particles, and aquatic plants negatively affected the classification accuracies of the proposed method, whereas the MLAs (SVM in particular) were less sensitive to such variations. The main disadvantage of using MLAs for operational waterbody mapping is the requirement for suitable training samples, representing both water and non-water land covers. The dynamic nature of reservoirs (many reservoirs are depleted at least once a year) makes the re-use of training data unfeasible. The study found that aggregating (combining) the thresholding results of two SAR and multispectral features, namely the S1 VH polarisation and the S2 NDWI, respectively, provided better overall accuracies than when thresholding was applied to any of the individual features considered. The accuracies of this dual thresholding technique were comparable to those of machine learning and may thus offer a viable solution for automatic mapping of waterbodies.

Keywords: waterbody mapping; machine learning; thresholding; optically complex; remote sensing; Sentinel-1; Sentinel-2

1. Introduction

Communities in developing countries rely on freshwater stored in small waterbodies for agricultural, domestic, mining, and industrial use [1]. These water resources are highly susceptible to climate variations and are often not sufficient to withstand long periods of drought. Recently, the water resources of the Cape Winelands District of South Africa have been under severe pressure due to drought conditions brought about by the El Niño weather cycle [2]. Agriculture plays a critical role in this region's economy [3], with wine production alone contributing to more than 30% of its regional gross domestic product (RGDP). Furthermore, the wine production industry provides more than 8% of the employment in the Western Cape Province [4]. The district is well-known for irrigated perennial crop production, mainly grapes (mostly for wine production) and fruits (apples, pears, peaches, olives, and citrus) [2]. In contrast to other parts of southern Africa, the area has a semi-arid Mediterranean climate with a mean annual rainfall of about 400 mm [5] and, as such, receives winter rainfall when demand for irrigation water is relatively low. In contrast, the growing season occurs during the dry and hot months when rainfall is low (about 20% of the total annual) and water demand for irrigation is at its apex [6].

During the recent drought (2015–2018), water reserves in the principal reservoirs were reduced to below 17% (April 2018), necessitating the implementation of drastic water restrictions by as much as 80% of normal usage for crop irrigation and industrial and domestic use [7]. Authorities were confronted with difficult decisions about how to best manage the limited available water resources and minimise the inevitable socio-economic impacts. Many limitations of existing procedures and gaps in available information sources were exposed. One of the biggest needs was to determine how resilient the agricultural industry, in particular the perennial crops sector, would be to severe water restrictions. This proved to be challenging given that no operational systems are in place to quantify and monitor how much water is stored in privately owned and managed reservoirs (dams). These reservoirs are of various sizes, ranging from 0.5–5 km². Most of these dams are ungauged and setting up, maintaining, and managing conventional in situ surveys, gauge stations, and telemetry networks would be prohibitively expensive and time-consuming [8].

Satellite remote sensing techniques have been shown to provide a viable alternative for monitoring water bodies. Satellite data can provide real-time, dynamic, and cost-effective information, and Earth observation procedures can be set up to provide operational (autonomous) monitoring of water resources [9,10]. Several methods have been proposed to classify surface water areas using either multispectral [9,11,12] or SAR remotely sensed data [13,14]. Popular techniques are image thresholding (rule-based classification) and supervised/unsupervised classification [15]. Image thresholding is easy to implement and thresholds that can be applied to images of different dates and areas can be automatically applied and are computationally inexpensive (not time-consuming) [9,16].

During thresholding, a single threshold value within the image scene is determined and all pixels below (or above) it are classified as water or non-water. According to Pierdicca et al. [17], the identification of a suitable threshold relies on a range of environmental factors, including atmospheric conditions, adjacency effects, mixed pixels, shadows and system factors such as viewing angle and pixel size [18–20]. Defining a robust threshold, one that will work effectively in different areas and on imagery acquired on different dates, has been cited by Feyisa et al. [21] as being a very challenging task, especially in optically complex (e.g., flooded vegetation and sedimented and turbid water) environments. An alternative approach to finding a single “optimal” threshold that will work in multiple situations is to make use of automated, image-specific, threshold identification methods. Several such techniques have been proposed, among which Otsu's simple and robust algorithm [22]

is one of the most utilised techniques for surface water mapping [9,12,15]. The Otsu algorithm finds a threshold by maximising the inter-class variance and minimising the weighted within-class variance [22].

Supervised and unsupervised classification techniques have also been popular for mapping water features using remotely sensed data [23,24]. For instance, using 30 m multispectral Landsat TM imagery, Xie et al. [25] obtained an accuracy of 96%, whereas Pradhan et al. [26] achieved an accuracy of 58% using 3 m TerraSAR-X data to retrieve water (flooded) pixels based on iterative self-organizing data analysis technique (ISODATA) unsupervised classification. The relatively low accuracy of the latter study was attributed to the presence of vegetation in the flooded area. Feng et al. [27] employed supervised classification to map surface waterbodies with 30 m multispectral HJ-1B imagery and achieved 94% overall accuracies. Similarly, Verpoorter et al. [28] achieved an accuracy of 95% using Landsat 7 ETM+ imagery. Although many authors agree that supervised classification is an efficient (accurate and fast) approach to map waterbodies, many highlight the need for prior definitions (training sites) to construct models capable of classifying unknown sites. The generation and collection of training samples is time-consuming, expensive, and tedious, often requiring extensive field visits. Nevertheless, recent implementations of non-parametric MLAs, including SVM, RF and DT, have demonstrated their value for mapping surface water. MLAs have the ability to classify unknown sites accurately using relatively small training sets and can handle large numbers of features [29].

In general, SAR and multispectral techniques are capable of accurately extracting water features if there is a significant contrast between water and non-water features in the data. However, the optical complexity of water affects the reflected spectral profile and backscatter values. For instance, the waterbodies in the Cape Winelands are characterised by varying concentrations of suspended sediments (turbidity), algae (e.g., chlorophylls, carotenoids), chemicals (e.g., nutrients, pesticides, and metals), dissolved organic matter, and aquatic plants [30,31]. This makes the implementation of supervised remote sensing-based water extraction methods difficult, as training data needs to be universally applicable and frequently updated, especially in the case of water bodies that are highly dynamic (reservoirs may be full or empty). To date, the remote sensing research community has given very little attention to how these variations affect waterbody mapping. Notable exceptions include Hong et al. [32], Frazier and Page [33], and Yang and Chen [34], who used RADARSAT-1 (16 m), Landsat TM (30 m), and S2 (10 m) data to map optically complex waterbodies. The latter study mapped optically complex waterbodies in urban areas and concluded that it is necessary to find the most appropriate and practical water identification methods regardless of the physical and chemical characteristics of waterbodies. In the studies done by Hong, Jang, Kim, and Sohn [32] and Yang and Chen [34], the water properties and conditions were not characterised. However, Frazier and Page [33] mapped the waterbodies with a defined turbidity of 90 mg/L.

The ESA recently launched a constellation of high spatial and temporal resolution satellites, namely S1 and S2, carrying C-band SAR and multispectral sensors, respectively [35]. Thanks to their dual-satellite-per-orbit configurations, S1 and S2 have relatively high revisit times of six and five days respectively. To our knowledge, no research has evaluated how data from these satellites can be combined to improve classification accuracies of waterbodies in complex environments, such as the Winelands District of South Africa [36,37].

Taking into account the challenges of mapping waterbodies with diverse physical and chemical characteristics, the objective of this study is as follows:

- To compare the performance of simple rule-based methods, i.e., the application of dynamic thresholds that can be easily incorporated into operational workflows, to the performance of supervised learning approaches (i.e., MLAs).

Thresholding and MLAs were applied to a range of features derived from Sentinel-1 (SAR) and Sentinel-2 (multispectral) data. This included a range of existing and new water indices and texture measures. Five popular MLAs, namely DTs, RF, k-NN, c-SVM, and SVM, were considered. The study

concludes by assessing the value of combining SAR and multispectral thresholding rules for mapping optically complex waterbodies.

2. Materials and Methods

2.1. Study Area

The study area is located in the Cape Winelands district of South Africa (Figure 1.). The focus area is about 40×45 km in size. The Cape Winelands district is the major wine and fruit producing region in South Africa. The area was chosen because of the optical complexity of the dams and reservoirs located therein.

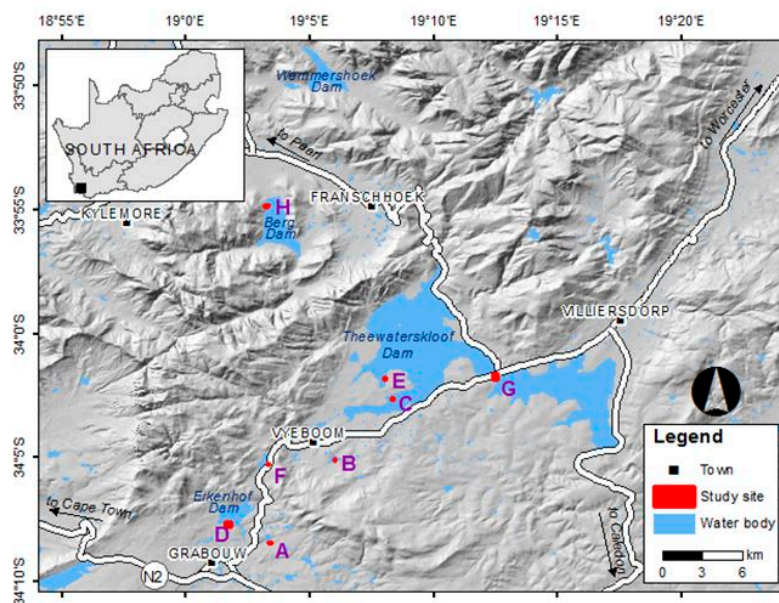


Figure 1. Study area and location of field survey sites.

The study area has a Mediterranean climate, characterised by warm, dry summers and cool, wet winters [5]. It receives a mean annual rainfall of about 400 mm and has a mean annual minimum and maximum temperature of 11 °C and 22 °C respectively. The high mountain ranges receive rainfall of up to 2000 mm per annum. The resulting runoff is collected by reservoirs located in the valleys. The suitable climate and presence of rivers and dams have led to agricultural activities and urbanisation. Fertilisers containing phosphorous and nitrogen are widely used to increase crop yields. These nutrients are carried by runoff from agricultural areas to waterbodies, resulting in eutrophication.

2.2. Data Collection and Preparation

2.2.1. Test Sites and Data Collection

Eight test sites (Table 1)) located in areas with diverse land cover/use featuring different types of waterbodies (Table 1) were chosen to evaluate the performance of the proposed methods.

Water edge (i.e., transition between water and non-water) reference points, each representing a 10×10 m plot to correspond with a Sentinel-2 image pixel, were collected using a handheld global positioning system (GPS) receiver (three metres accuracy). The GPS measurements were taken along the water edge at each site. Four GPS surveys at different dates were carried out to record water edge changes (due to water level fluctuations). The dates of the surveys were chosen to closely match the dates of satellite acquisitions (Table 2). Since the GPS points were collected along the edge of the reservoirs, they represent mixed pixels (i.e., they contained both water and non-water

components). However, the locations were selected so that the majority of land cover in each plot is water. These samples were consequently labelled as water.

Table 1. Description of the physical characteristics of the survey sites.

Site	Description	Size (km ²)
Site A	Very shallow and turbid	0.8
Site B	Shallow with moderate turbidity and eutrophication	0.4
Site C	Clear with moderate eutrophication	1.6
Site D	Shallow and humic-rich (black) water	2.7
Site E	Very shallow and eutrophied	2.3
Site F	Shallow, sediment and humic-rich (black) water	1.7
Site G	Shallow and moderate turbidity	3.1
Site H	Shallow, clear, and wind-induced turbulence	4.88

Table 2. Sentinel image acquisition and field visit dates.

Field Visit	Sentinel-1 Image	Sentinel-2 Image
27 October 2016	27 October 2016	23 October 2016
26 November 2016	25 November 2016	22 November 2016
28 January 2017	31 January 2017	31 January 2017
25 February 2017	24 February 2017	03 March 2017

Reference points representing pure (not-mixed) water pixels were difficult to obtain during field surveys as they required access to open water (e.g., using a boat). Instead, pure water samples were collected using visual interpretation of the Sentinel-2 and Google Earth imagery. Point distributions were random, although some points were excluded in cases where they were deemed to be mixed (i.e., if they occurred near other land covers). Non-water reference samples were collected in a similar manner. A broad four-class (grass, bare and built up, shadow, trees and shrubs) classification scheme was adopted for the non-water samples to ensure diversity and to gain a better understanding of which non-water classes are most frequently confused for water. Shadow was included as a separate class, as it is well known to be misclassified as water. Table 3 summarizes the samples collected per land cover class.

Table 3. Reference samples collected per land cover type.

Class	%	No. of Samples from Imagery	No. of GPS Samples	Total
Water	26	3350	1045	4395
Grass	16	2660	0	2660
Shadow	10	1695	0	1695
Bare & built up	28	4660	0	4660
Trees & shrubs	20	3330	0	3330
Total	100			16,640

2.2.2. Multispectral Images Pre-Processing

Four cloud-free Sentinel-2 level-1C images were downloaded from ESA's Scientific Data Hub (<https://scihub.copernicus.eu/dhus/>). The Sentinel-2 images have 13 bands, of which, four bands (blue, green, red and NIR) have a spatial resolution of 10 m; six bands (including SWIR) have a spatial resolution of 20 m; and three have a 60 m resolution (coastal aerosol, water vapour, and SWIR-Cirrus bands). The images were atmospherically corrected using the Sen2cor algorithm, available in the Sentinel Application Platform (SNAP) toolbox, which uses the Climate Change Initiative (CCI) land cover data to characterize atmospheric conditions at the time of acquisition. The atmospheric correction was done at 10 m, resulting in the output excluding the 60 m bands (Bands 1, 9, and 10) and resampling the 20 m bands to 10 m [38]. Thus, ten bands at 10 m spatial resolution were preserved for further analysis.

2.2.3. SAR Data Pre-Processing

The Sentinel-1 constellation consists of two SAR satellites (Sentinel-1A and Sentinel-1B) that record C-band (5.405 GHz) backscatter at incidence angles ranging from 29–46°. The study uses the ground range detected (GRD) interferometric wide (IW) images, which have large swath widths (250 km) and moderately high spatial resolutions (5 × 20 m). IW offers dual polarization capability, which can provide more information about ground surfaces, as compared to single polarizations. Only horizontal transmit, vertical receive (HV) and vertical transmit, vertical receive (VV) polarizations were available over the study area.

The Sentinel-1 toolbox (S-1 TBX), available in SNAP, was used for the pre-processing of the SAR dataset. Figure 2 shows the pre-processing chain that was followed.

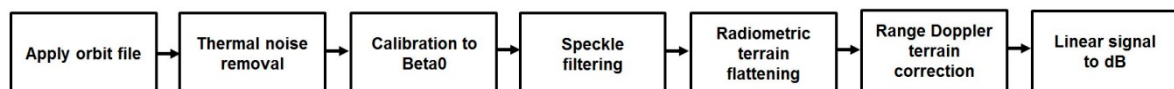


Figure 2. Pre-processing steps for Sentinel-1 data.

The images were projected and resampled using nearest-neighbor to 10 m resolution. The universal transverse Mercator (UTM) WGS84 coordinate system (zone 34 South) was used to allow for pixel-to-pixel comparison with the Sentinel-2 images.

2.3. Feature Set Generation for Classification

In addition to the ten Sentinel-2 spectral and two HV and VV Sentinel-1 polarizations, a range of supplementary features were generated and used as input to the classification methods. Table 4 outlines the 296 (74 per image capture date) features considered. To reduce the number of variables (feature dimensionality), Bands 5 (vegetation red-edge), 7 (vegetation red-edge), and 8a (narrow NIR) were excluded as the first two Bands were highly correlated with Band 6 and the latter with Band 8.

Table 4. Features used as input to the thresholding and MLAs.

Data	Type	Subtype	Description	Total Features
Sentinel-1	Speckle filters based on polarisations	HV	Boxcar, none, median (5 × 5), Lee-sigma, refined Lee, frost, gamma MAP, IDAN, and Lee	9
		VV	Boxcar, none, median (5 × 5), Lee-sigma, refined Lee, frost, gamma MAP, IDAN, and Lee	9
	Polarisation ratios	$\frac{HV}{VV}$	Boxcar, none, median (5 × 5), Lee-sigma, refined Lee, frost, gamma MAP, IDAN, and Lee	9
Sentinel-2	Spectral indices	Reflectance bands and mean of the six bands	B2, B3, B4, B6, B8, B11, and Mean	7
		Normalised difference spectral indices (NDSIs)	Band combinations from Sentinel-2 bands (B2, B3, B4, B6, B8, B11, and B12) e.g., (B2-B3)/(B2+B3)	21
	Pan-sharpening of SWIR (Band 11)		Band combinations P1 of B11	6
			Band combinations P2 of B11	6
	Textural features	Grey level co-occurrence matrix (GLCM)	Correlation, Homogeneity	2
Grey level difference vector (GLDV)		Contrast, Entropy, Mean	3	
Image transform	Principle components	PC1 and PC2	2	

Note: P1 = ATWT pan-sharpening, P2 = Gram Schmidt pan-sharpening, PC = principal component, B = band, MAP = maximum a posteriori, and IDAN = intensity driven adaptive neighbourhood.

The S2 spectral bands were used to develop normalised difference spectral indices (NDSIs), which include the normalised difference water index (NDWI) [39], normalized difference moisture index (NDMI) [40], modified normalized difference water index (MNDWI) [41], and water ratio index (WRI) [42] indices. Table 5 shows the calculation of the popular indices. Band 11 was up scaled from 20 m to 10 m (i.e., for generating a 10 m resolution SWIR band) to produce MNDWI at 10 m spatial resolution. Two popular pan-sharpening algorithms, namely Gram–Schmidt (GS) [43] and À Trouis Wavelet Transform (ATWT) [44] were used, where Band 8 was employed as the panchromatic (PAN) band, as suggested by [9]. Five bands (B2, B3, B4, B6, and B8) were used to develop the NDSIs at 10 m resolution.

Table 5. Calculation of the most popular indices-based on Sentinel-2 reflectance bands at 10 m spatial resolution.

Index	Equation
Normalized difference water index (NDWI)	$NDWI = \frac{Band\ 3 - Band\ 8}{Band\ 3 + Band\ 8}$
Normalized difference moisture index (NDMI)	$NDMI = \frac{Band\ 8 - Band\ 11}{Band\ 8 + Band\ 11}$
Modified normalized difference water index (MNDWI)	$MNDWI = \frac{Band\ 3 - Band\ 11}{Band\ 3 + Band\ 11}$
Water ratio index (WRI)	$WRI = \frac{Band\ 3 + Band\ 4}{Band\ 8 + Band\ 11}$

The approach for examining all the possible combinations of spectral bands (Equation (1)) was adopted from the OBA-NDWI methods proposed in [19]. Table 6 shows the list of the band combinations that were considered. The means of the ten bands were also included to investigate whether any of these features and OBA-NDWI were useful for surface water detection.

$$NDSI = \frac{b_i - b_j}{b_i + b_j} \quad i = \{1, 2, \dots, n - 1\}, \quad j = \{i + 1, \dots, n\}. \tag{1}$$

Table 6. The normalized difference spectral indices (NDSI) generated from the seven bands, as well as two pan-sharpened band 11 features.

	B2	B3	B4	B6	B8	B11	B12
B2							
B3	$\frac{B2-B3}{B2+B3}$						
B4	$\frac{B2-B4}{B2+B4}$	$\frac{B3-B4}{B3+B4}$					
B6	$\frac{B2-B6}{B2+B6}$	$\frac{B3-B6}{B3+B6}$	$\frac{B4-B6}{B4+B6}$				
B8	$\frac{B2-B8}{B2+B8}$	$\frac{B3-B8}{B3+B8}$	$\frac{B4-B8}{B4+B8}$	$\frac{B6-B8}{B6+B8}$			
B11	$\frac{B2-B11}{B2+B11}$	$\frac{B3-B11}{B3+B11}$	$\frac{B4-B11}{B4+B11}$	$\frac{B6-B11}{B6+B11}$	$\frac{B8-B11}{B8+B11}$		
B12	$\frac{B2-B12}{B2+B12}$	$\frac{B3-B12}{B3+B12}$	$\frac{B4-B12}{B4+B12}$	$\frac{B6-B12}{B6+B12}$	$\frac{B8-B12}{B8+B12}$	$\frac{B11-B12}{B11+B12}$	
B11 _{ATWT}	$\frac{B2-B11_{ATWT}}{B2+B11_{ATWT}}$	$\frac{B3-B11_{ATWT}}{B3+B11_{ATWT}}$	$\frac{B4-B11_{ATWT}}{B4+B11_{ATWT}}$	$\frac{B6-B11_{ATWT}}{B6+B11_{ATWT}}$	$\frac{B8-B11_{ATWT}}{B8+B11_{ATWT}}$		$\frac{B2-B11_{ATWT}}{B2+B11_{ATWT}}$
B11 _{GS}	$\frac{B2-B11_{GS}}{B2+B11_{GS}}$	$\frac{B3-B11_{GS}}{B3+B11_{GS}}$	$\frac{B4-B11_{GS}}{B4+B11_{GS}}$	$\frac{B6-B11_{GS}}{B6+B11_{GS}}$	$\frac{B8-B11_{GS}}{B8+B11_{GS}}$		$\frac{B12-B11_{GS}}{B12+B11_{GS}}$

Note: The shaded part of the table was not considered.

Principal component analysis (PCA) was performed on ten S2 bands per image date and the first two components (PC1 and PC2) with the largest “percent of Eigenvalues” were retained [45]. Two types of textural measures, namely the grey level co-occurrence matrix (GLCM) and grey level difference vector (GLDV), were generated from each PC1. These texture measures were calculated based on equations as explained in [46]. These measures quantify differences in the grey levels within a local window [47]. In this study, the window size was set to (5 × 5) pixels, as suggested by Zhang et al. [48]. The GLDV texture measures employed were contrast, entropy, and mean, while correlation and homogeneity were selected from the GLCM analyses.

Nine popular speckle filters available in SNAP, namely boxcar, none, median (5 × 5), Lee-sigma, refined Lee, frost, gamma-MAP (maximum a posteriori), intensity driven adaptive neighbourhood

(IDAN), and Lee were applied to the HV and VV SAR polarizations [49]. In the interest of brevity, the reader is referred to [50–52] for overviews of these speckle filters.

2.4. Experimental Design

The thresholding results were compared to the classifications produced by the MLAs to get a sense of relative performance (i.e., the MLA results were used as benchmarks against which the autonomous rule-based approaches (e.g., thresholding) could be compared). Autonomous rule-based approaches classify images based on stipulated rules with little or minimum intervention [53,54]. The classification experiments were applied for each site separately and in combination (general model) to better understand how variations in waterbody types influence accuracies. Table 7 summarises the experiments, classification methods, and input features. The thresholding classified each feature individually, whereas MLAs considered them all in combination.

Table 7. Thresholding and machine learning experiments carried out in this study.

Experiment Set	Classification Method	Input Features	Number of Experiments
A	Thresholding	Each feature individually	$296 \times 9 = 2664$
B	k -NN	All features combined	$1 \times 9 = 9$
C	DT	All features combined	$1 \times 9 = 9$
D	RF	All features combined	$1 \times 9 = 9$
E	SVM	All features combined	$1 \times 9 = 9$
F	c -SVM	All features combined	$1 \times 9 = 9$

2.5. Image Thresholding

Threshold selection is a key step in using rule-based approaches for waterbody mapping [9]. Several researchers have noted the difficulty of selecting robust threshold values, as image variables (e.g., spectral indices and backscatter) are often dynamic [15,55–58]. Furthermore, threshold values vary both temporally and spatially among regions, depending on different image and water characteristics.

The use of a deterministic threshold, such as zero, and automatic thresholding techniques (e.g., zero in NDWI) can either overestimate or underestimate surface water areas [9,19]. Various automatic threshold selection methods have consequently been proposed in the literature, including histogram shape, measurement space entropy, spatial correlation, and local grey-level surface [11,57]. Although, threshold segmentation can distinguish water pixels, the methods have been known to yield unstable results in situations where the spectral characteristics between water and other dark objects, such as buildings and shadows, is similar [11].

In this study, waterbody masks were extracted from each of the 252 features (Table 4) by applying a threshold dynamically generated with the Otsu algorithm, which is based on histogram shape [22]. The algorithm is a widely used automatic thresholding method aimed at maximizing inter-class variance and minimising intra-class variance [9]. However, the method has been known to yield unstable results when a small area of water bodies and large non-water features exist [11].

The thresholding experiments per feature (Table 4) and per each site were automated in MATLAB software. Otsu automatically defines a threshold value t that divides the image into two classes. In this study, the two classes were set to water and non-water. The threshold value t separating these classes is determined by a set of equations as outlined in [9] as follows:

$$\delta^2 = P_{nw} \cdot (M_{nw} - M)^2 + P_w \cdot (M_w - M)^2, \quad (2)$$

$$M = P_{nw} \cdot M_{nw} + P_w \cdot M_w, \quad (3)$$

$$M = P_{nw} \cdot M_{nw} + P_w \cdot M_w, \quad (4)$$

$$P_{nw} + P_w = 1, \quad (5)$$

$$t^* = \underset{a \leq t \leq b}{\text{Arg Max}} \{P_{nw} \cdot (M_{nw} - M)^2 + P_w \cdot (M_w - M)^2\}, \quad (6)$$

where δ is the inter-class variance of the non-water class and the water class; P_{nw} and P_w are the probabilities of one pixel belonging to non-water and water, respectively; M_{nw} and M_w are the mean values of the non-water and water classes; and M is the mean value of the feature image.

2.6. Machine Learning

The Supervised Learning and Image Classification Environment (SLICE) software developed by the Centre for Geographical Analysis at Stellenbosch University [59] was used for the supervised machine learning classification. SLICE integrates five popular MLAs, namely DTs, k -NN, RF, constant optimisation parameter SVM (c -SVM), and SVM. These MLAs are well established in RS applications, due to their flexibility, simplicity and computational efficiency [59].

SVM is a classification technique based on a statistical learning theory and aims to determine the location of decision boundaries by maximizing the margin between classes [60]. In the case of two linearly separable classes, SVM selects, from among the infinite number of linear decision boundaries, the optimal separating hyperplane (OSH), which minimises the generalisation error. When the data are not linearly separable, SVM is extended by introducing slack variables and applying a kernel function to solve the optimisation problem [61]. The radial basis function (RBF) kernel usually trains much faster by mapping every point to a Gaussian function and was chosen for this study, as recommended by Jia et al. [62]. The c parameter in c -SVM helps to optimise SVM, since the value is tuned based on the input data. For large values of c , the optimisation will choose a smaller-margin hyperplane, whereas a very small value of c will cause the optimiser to look for a larger-margin OSH, even if that hyperplane misclassifies more points.

DT is a predictive, flexible, and comprehensive classification algorithm that labels an unknown class using a sequence of rules that leads to a classification decision [63]. A decision tree is composed of a root node, a set of interior nodes, and terminal nodes (termed leaf nodes). The root node and interior nodes are linked to decision stages, while the terminal nodes represent the final classification. The efficiency and performance of this algorithm are strongly affected by the set of rules inducting the path to be followed, starting from the root node and ending at one terminal node that represents the label for the object being classified. At each nonterminal node, a decision is made about the path to the next node [64].

RF is an ensemble MLA consisting of a combination of DT classifiers [65]. All trees are trained with the same features but on various training sets, which are generated randomly from the original training data. After training, each tree assigns a class label to the test data. Finally, the results of all decision trees are fused and the majority of votes determine the class label for each land cover [66]. Depth and minimum sample size are the two important tuning parameters in the RF algorithm. In this study, the maximum depth, the minimum number of samples, and pruning harshness was set to 50, one, and the minimum, respectively, as suggested by Garage [67].

The k -NN classifier is a distance rule-based technique which assigns an unknown sample to the class that occurs most frequently among its k nearest neighbours [68]. The basic functioning behind k -NN is that the group of k samples in the calibration dataset that are nearest (in feature space) to an unknown sample is used to infer (through a majority vote) its membership [69]. Therefore, k is the key tuning parameter in this classifier and largely determines the performance of the classifier. For this study k was set to 1, as proposed by [69].

2.7. Accuracy Assessment

A 3:2 sample split ratio was employed for classifier training and accuracy assessment, as suggested by Gilbertson, Kemp, and Van Niekerk [3]. The number of test samples needed for accuracy testing was

based on the multinomial distribution for a confidence interval of 95% for the accuracy assessment [70]. Testing samples per class was determined based on the percent coverage calculated from an initial unsupervised classification, as suggested by [71]. The percentage coverages were 24.8, 19.8, 30, 15, and 10.6 for water, trees & shrubs, bare & built, grass, and shadow, respectively (see Appendix A). The non-water classes were combined (reclassified) into one class, namely non-water, to assess the binary thresholding experiments. The same training (input) and testing (validation) datasets were used for all the classification experiments to ensure that differences in accuracy could be attributed to the nature of the class allocation processes.

A producer’s accuracy (PA), user’s accuracy (UA), overall accuracy (OA), and the kappa coefficient (K) were generated for each classification experiment. OA is easily interpreted as it represents the percentage of classified pixels in the image that have been correctly labelled, while K can be used to assess statistical differences between classifications [68]. The statistical significance of the accuracy differences among experiments was evaluated using non-parametric statistical tests, namely McNemar’s [72] and Friedman’s test, as implemented in the Statistical Package for Social Sciences (SPSS). Differences were considered as statistically significant at $p < 0.05$.

3. Results

3.1. Thresholding

Table 8 lists the results of the six best-performing Otsu-based thresholding experiments (named T1–T6 for easier notation). The Table 8 also defines what each T1 represents. Compared to the Sentinel-1 features, higher accuracies were achieved when thresholding was applied to the Sentinel-2 variables, with only one SAR-based experiment (T2) being among the six best results. When considering the combination of all the study sites, NDWI (T1), derived from the green and NIR Sentinel-2 bands, was the most successful in separating water from other land covers with an OA of 81.6% and K of 0.73. The second-best performing feature was the Sentinel-1 VH polarisation (T2), derived from the RL filter, with OA and K values of 77.7% and 0.67 respectively. According to McNemar’s test, the difference between T1 and T2 is statistically significant. The second-best performing Sentinel-2 feature (OA of 71.8%) was the MNDWI, derived from the green band, and the ATWT pan-sharpened SWIR Sentinel-2 Band 11 (T3). This result was significantly lower than both T1 and T2, but not significantly higher than when individual bands (T4 and T5) were used as input to the thresholding algorithm. The accuracy levels dropped off sharply in T6 when Gram–Schmidt pan-sharpening was used for MNDWI.

Table 8. Overall accuracies (OA), kappa coefficients (K), mean (\bar{x}), and standard deviation (δ) values for the six best performing thresholding features.

SITE	Thresholding												Average			
	T1		T2		T3		T4		T5		T6		OA		K	
	NDWI		VH Pol _{RL}		MNDWI _{ATWT}		Band 8		Band 11 _{ATWT}		MNDWI _{GS}					
	OA	K	OA	K	OA	K	OA	K	OA	K	OA	K	\bar{X}	δ	\bar{X}	δ
A	88.2	0.83	85.5	0.65	73.2	0.59	74	0.53	63.7	0.53	64.2	0.51	77.6	9.5	0.50	0.27
B	91.8	0.84	88.5	0.77	68.3	0.58	77.6	0.55	75.4	0.51	71.6	0.54	76.8	12.1	0.40	0.38
C	90.8	0.83	88.4	0.77	93.6	0.87	81.2	0.73	78	0.57	84.2	0.89	86.9	7.1	0.78	0.15
D	91.6	0.89	89.6	0.79	75.8	0.62	76.7	0.53	65.9	0.51	65.7	0.51	77.8	11.4	0.49	0.30
E	90.7	0.81	89.5	0.79	94.6	0.89	80.8	0.62	75.9	0.52	85.1	0.92	87.1	8.0	0.78	0.17
F	89.1	0.78	88.7	0.77	65.9	0.55	78.9	0.58	73.1	0.57	64.3	0.51	75.4	16.0	0.38	0.37
G	96.1	0.92	92.4	0.85	98.2	0.91	88.8	0.78	88.6	0.77	93.2	0.93	90.5	6.0	0.81	0.11
H	90.3	0.81	82.8	0.71	91.1	0.82	76.3	0.53	75.7	0.52	70.4	0.54	83.9	11.0	0.63	0.20
\bar{X}	90.7	0.82	86.3	0.70	85.2	0.67	81.3	0.61	77.8	0.55	75.2	0.73	82.8	5.8	0.7	0.13
SD	1.57	0.03	3.1	0.10	11.8	0.20	4.5	0.10	6.2	0.05	13.2	0.15	6.73	4.75	0.11	0.06
All sites	81.6	0.76	77.7	0.71	73.8	0.69	69.5	0.57	67.7	0.57	65.2	0.56	72.3	6.25	0.64	0.08

Notes: MNDWI_{GS} = the MNDWI produced from applying Gram–Schmidt pansharpening to band 11; MNDWI_{ATWT} = the MNDWI produced by applying à Trouw wavelet transform pan-sharpening to band 11; and VH Pol_{RL}= the VH polarisation produced from refined Lee speckle filtering.

Generally, thresholding was more successful when each site was classified individually (i.e., using a locally adapted threshold). For instance, the mean OA of the per-site NDWI (T1) classifications was 90.7%, which is significantly higher than the 81.6% OA achieved when all the sites were classified in combination. A similar pattern is observed for the other features (all differences between mean OAs per-site and OAs of all sites combined were statistically significant), although the variation among site-specific classifications varied considerably. Notably, the standard deviation (SD) of the NDWI (T1) classifications was 1.57%, while for $MNDWI_{CS}$ (T6) and $MNDWI_{ATWT}$ (T3) it was 13.2% and 11.8%, respectively, which brings the stability of the latter two features into question. The stability of the Sentinel-1 VH refined Lee (RL) speckle filter (T2) was better (SD of 3.1) than that of the two MNDWI-based features (T3 and T6), but still significantly lower than that of NDWI (T1). This suggests that no single threshold could accurately separate water and non-water land covers in all sites. This is supported by Figure 3a–c, which demonstrates the temporal variability of NDWI, MNDWI and VH/VV for the points taken at the same waterbody (Site G) on different dates. Furthermore, Figure 3d shows the spectral and spatial shift at each point, based on a Sentinel-2 image acquired on 22 November 2016, which suggests variability within the same waterbody.

The accuracies among study sites varied substantially. Site G, which is slightly turbid and eutrophied, achieved the highest mean OA (90.5%) while the lowest accuracy was recorded at site F (mean OA 75.4%). The latter site is shallow with humic-rich water from a slow-moving channel flowing through forested plantations (Eucalyptus pine). MNDWI (T3) showed the highest accuracy for delineating sites C, E, and H. These sites represent clear and eutrophied water. Thresholding of NDWI (T1) produced the best results when humic water sites were classified (i.e., A, B, D, and F), while T2 (SAR backscatter) performed generally well ($> 82\%$) in all sites. This suggests that the SAR data were less affected by the optical variabilities among the waterbodies. A Friedman’s test showed that the difference between feature type and optical variability of water are statistically significant ($p = 0.002$).

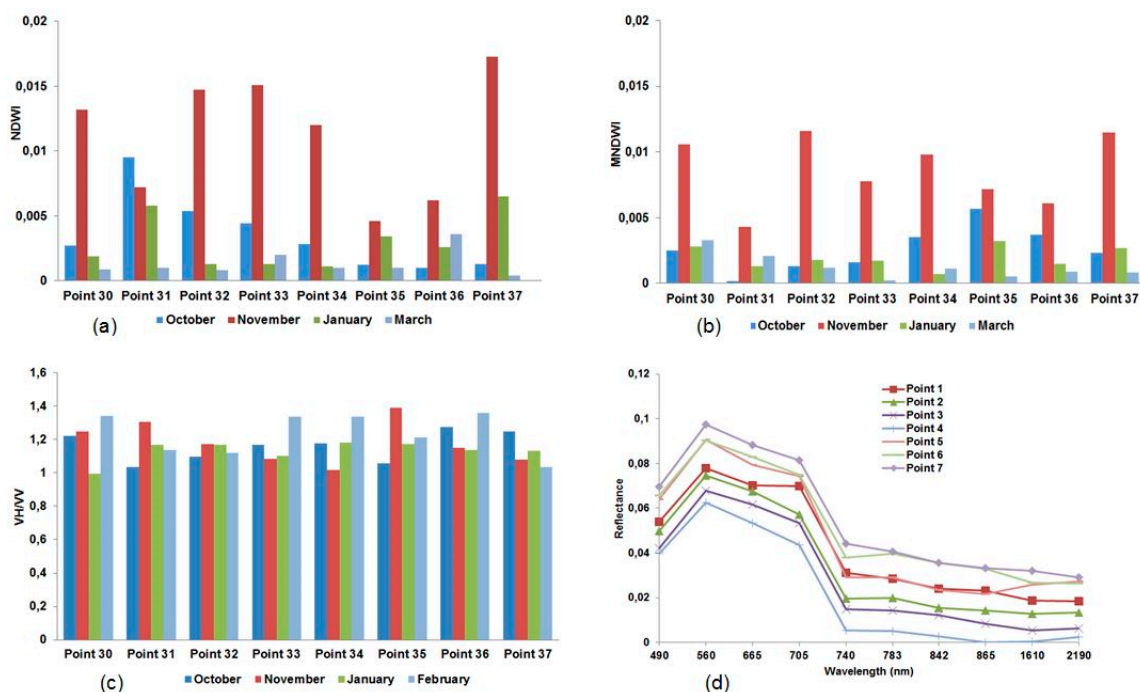


Figure 3. The selected GPS points collected on Site G showing the temporal and spatial variability in (a) NDWI, (b) MNDWI, (c) VH/VV on different dates, and (d) shows the spectral variability on a Sentinel-2 image acquired on 22 November 2016.

Unlike the other indices tested, NDWI was found to have the ability to spectrally differentiate surface water with different characteristics located among different land cover types, including shadows

or dark areas. For instance, Figure 4 shows that MNDWI incorrectly classified humic rich water as non-water and confused shadows with water (Figure 5). Details of confusion matrices, including commission and omission errors when applying NDWI, MNDWI, and VHPolarisation_{RL}, are shown in Tables A1–A3. The waterbodies were better captured by NDWI in all cases.

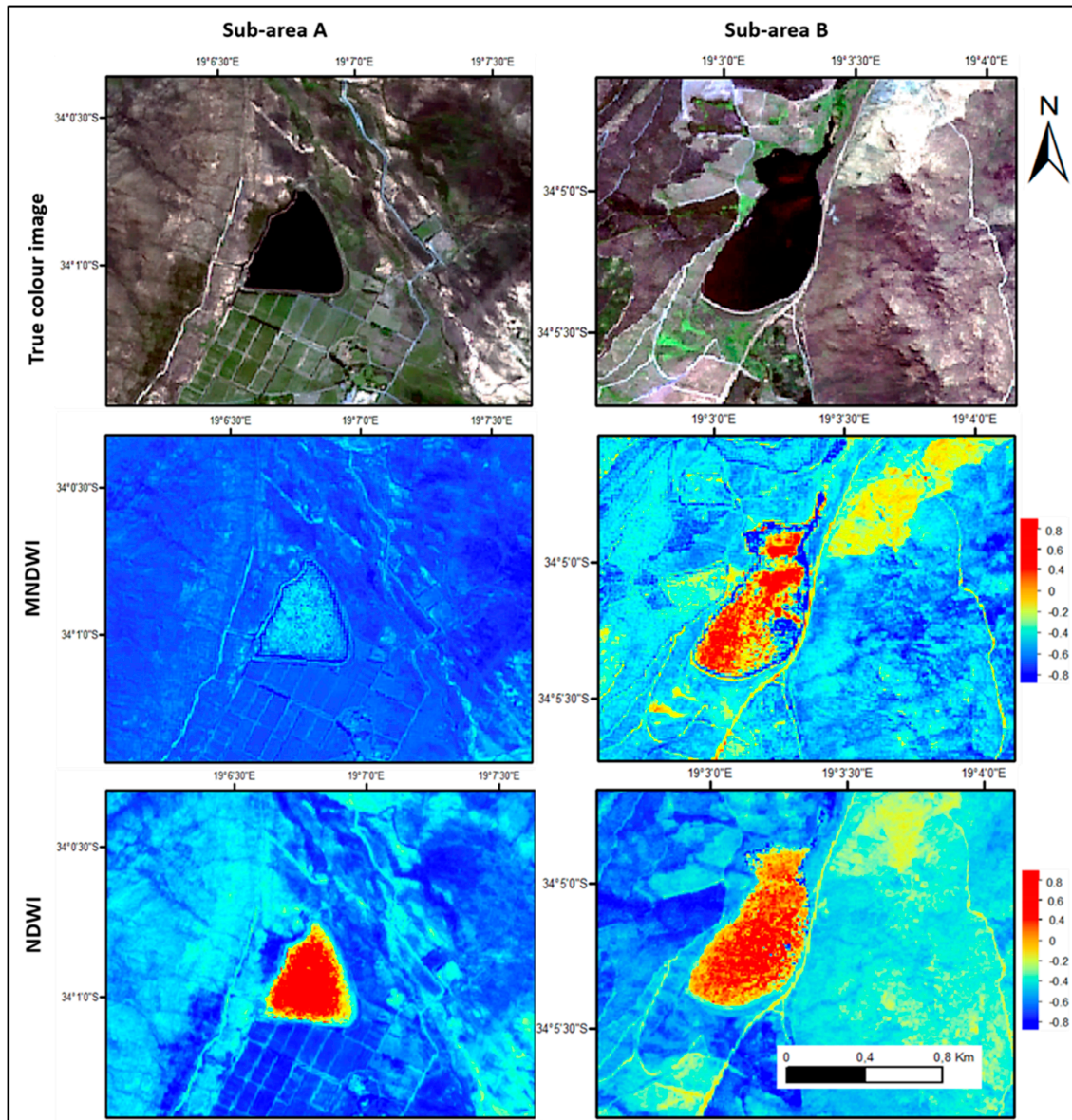


Figure 4. Detailed (large-scale) examples of the 10 m true colour maps of Sentinel-2 (4, 3, 2), MNDWI, and NDWI images. The first column represents site A and the second column is for site F.

Shadows and water are spectrally similar and were consequently difficult to discriminate, as depicted by large errors of omission and commission in the shadow class with all the MNDWI, NDWI and VH Polarisation_{RL}. For example, for MNDWI, a higher commission error in the shadow class was detected (47%) (mainly due to misclassification of water), which is also reflected in the high omission error (16.5%). Furthermore, this is supported by the visualisation of false positives for MNDWI, especially in mountainous terrain (Figure 5).

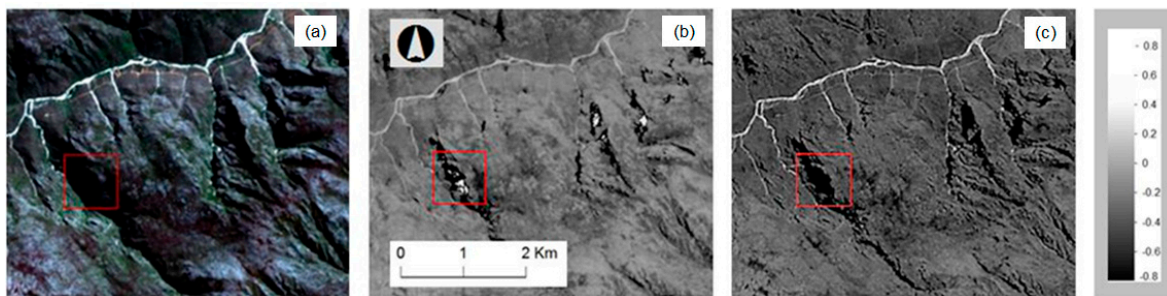


Figure 5. Visual comparison of Sentinel-2 (a) true colour image (4, 3, 2), (b) MNDWI, and (c) NDWI on mountain slopes showing the misrepresentation of shadows by MNDWI.

Figure 6 provides a qualitative comparison of T1 and T2 in test site F generated from images captured on 31 January 2017. In general, it seems that T1 classified water with greater accuracy than T2; however, T1 (marked with green squares) and T2 (marked with red squares) omitted water in some areas (Figure 6). To reduce these errors and in the interest of finding a solution to classify water automatically and accurately, an additional experiment (called “T1+T2”) was carried out in which T1 and T2 were unioned (i.e., using the Boolean operator OR). Visual inspection of Figure 6 suggests that T1+T2 resulted in a better accuracy of surface water mapping compared to either T1 or T2. The accuracy of T1+T2 was significantly (8%) higher than that of T1, achieving an OA of 89.3%.

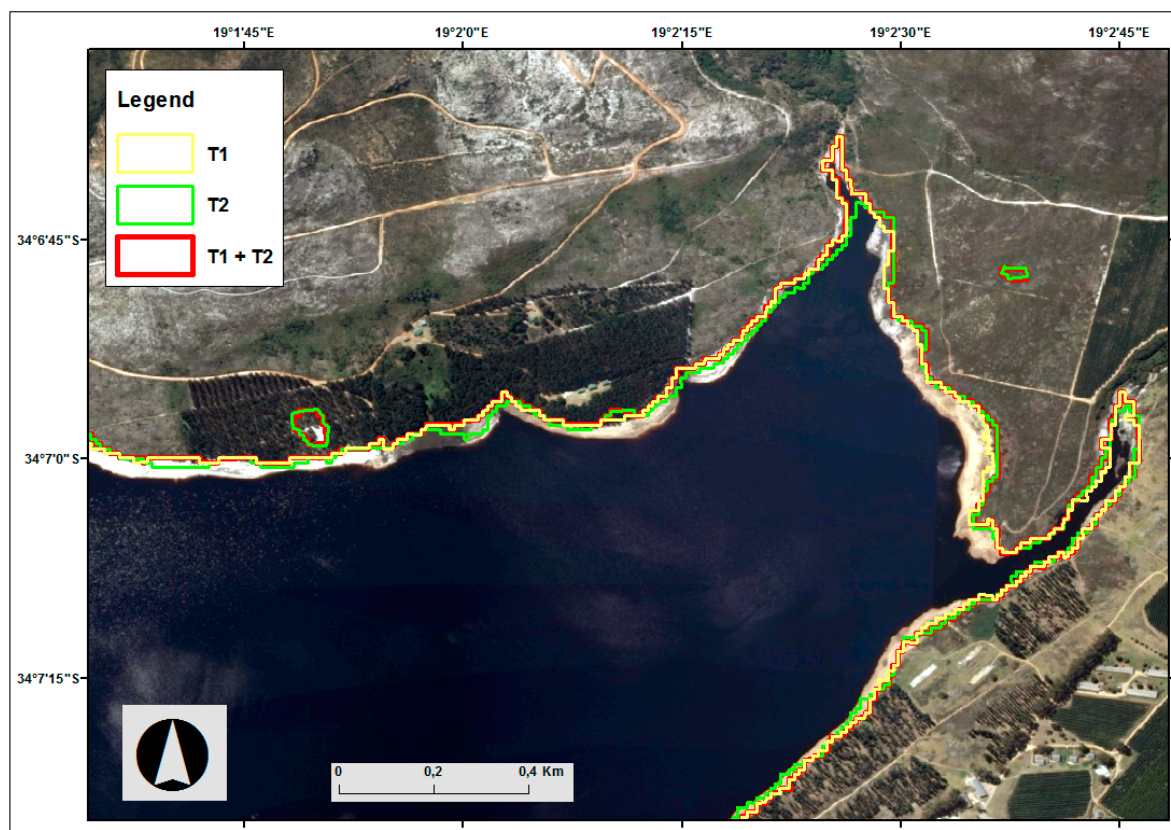


Figure 6. Visualisation of water masks derived from T1 (NDWI), T2 (SAR VH polarisation), and T1+T2 (fusion of T1 and T2). The background image is an aerial photograph taken in November 2014 when water levels were very low.

3.2. Benchmarking Thresholding to Machine Learning

Table 9 summarises the machine learning classification results. Generally, all the classifiers performed well at classifying slightly turbid water (site G). SVM significantly outperformed the other classifiers when the classifications were carried out per individual site, with site G recording the highest mean OA of 95.9%. This result is significantly higher ($p = 0.03$) than the second-best classifier c -SVM (mean OA = 93.3%). On average, DT was the worst performing classifier (mean OA of 88.4%) when the classifications were carried out per site, except for site G (94.6%), where it outperformed RF (92.5%) and k -NN (93.8%). With a SD of 3.7, DT was also the least stable of the five classifiers. The c -SVM was the second-best performing classifier, but it did not perform well at classifying sites D and E (relative to k -NN and RF).

SVM consistently outperformed the other classifiers, with an OA and K values of 91.7% and 0.82, respectively, when all sites were combined. This was significantly higher ($p = 0.03$) than the second-best performing classifier c -SVM, which achieved an OA of 89.6%. DT delivered the poorest overall classification results (OA = 78.7%), followed by RF (79.5%), and k -NN (80.7%). The accuracies of all classifiers dropped significantly when all the sites were classified in combination (i.e., when the complexity of the target classes increased), with RF and k -NN being the most affected (reduction in mean OA of more than 10%).

Table 9. Overall accuracies (OA), kappa coefficients (K), mean (\bar{x}), and standard deviation (δ) values for the MLAs.

SITE	Classifier										Overall Average			
	SVM		c -SVM		k -NN		RF		DT		OA		K	
	OA	K	OA	K	OA	K	OA	K	OA	K	\bar{X}	δ	\bar{X}	δ
A	95.8	0.91	92.7	0.90	87.2	0.85	88.7	0.86	81.7	0.79	89.2	4.73	0.86	0.05
B	97.4	0.93	94.2	0.91	89.0	0.87	89.8	0.86	85.0	0.82	91	4.07	0.88	0.05
C	96.8	0.94	94.4	0.91	90.5	0.89	89.3	0.86	86.2	0.83	91.2	3.39	0.88	0.04
D	95	0.94	90.7	0.92	91.8	0.90	91.3	0.89	90.2	0.89	93.9	1.89	0.91	0.02
E	96.3	0.88	93.3	0.95	93.6	0.90	91.4	0.89	91.2	0.89	93	1.52	0.90	0.03
F	94.5	0.90	92.2	0.90	89.7	0.87	88.7	0.86	86.2	0.83	90	2.1	0.87	0.03
G	98.2	0.96	95.7	0.95	93.8	0.89	92.5	0.89	94.6	0.95	94	2.67	0.93	0.03
H	94.6	0.91	94.3	0.93	91.5	0.89	92.7	0.90	91.3	0.90	93	1.08	0.91	0.02
\bar{X}	95.9	0.93	93.3	0.92	90.8	0.9	90.5	0.88	88.4	0.89	91.8	2.89	0.90	0.03
SD	2.17	0.03	1.1	0.02	2.26	0.02	2.6	0.04	3.7	0.03	2.37	0.93	0.03	0.01
All	91.7	0.82	89.6	0.81	80.7	0.78	79.5	0.77	78.7	0.76	81.2	2.32	0.79	0.03

The OAs of the MLAs and best thresholding classifications are graphically compared in Figure 7. SVM and c -SVM performed the best, regardless of the characteristics of the waterbody. T1 performed better than the worst performing machine learning classifier (k -NN) at sites A, B, and C, which are characterised by moderately eutrophied water. At site D (humic water), T1 achieved a 1.5% higher OA than c -SVM. Although T3 was the worst performing classification when all sites were combined, it performed on par with the machine learning classifiers at sites C, E, G, and H. For instance, at site E its accuracy was significantly (1.3%) higher than what was obtained with c -SVM.

Although SVM was superior, the fusion of the T1 and T2 rulesets improved the threshold-based classification outcome to achieve competitive results. T1+T2 achieved a higher accuracy than k -NN at all individual sites and, at site D, it outperformed c -SVM by about 2.7%. At site E (eutrophied waterbody), T1+T2 attained the highest accuracy, whilst at sites C and G, its accuracy was almost on par with that of c -SVM. It is important to note that the fusion of T1 and T2 did not improve the OAs at sites D and H by much. Figure 7 shows that all the classifiers struggled (OAs below 95%) at sites D, F, and H. These sites are characterised by humic rich water and are located in mountainous terrain.

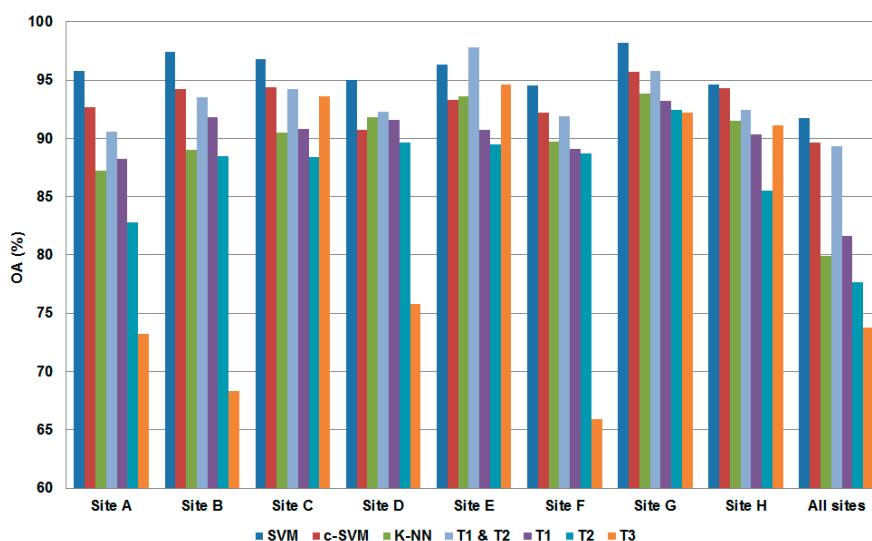


Figure 7. Comparison between thresholding and MLAs for all sites

4. Discussion

The results show that the characteristics of the water, type of classifier, and input feature dataset had a significant impact on the accuracies of the surface water classifications. With the multispectral data, the selection of the spectral index had a significant impact on accuracies. MNDWI's lower OA compared to NDWI was mainly due to the under classification of humic rich water (Figure 4) and over classification of shadows (Figure 5).

NDWI was able to highlight dark, turbid, and eutrophied water more effectively than MNDWI. This finding contrasts with those of Xu [41] and Zhai et al. [73], who noted that MNDWI provided better discriminatory power than NDWI for shadowed and dark areas in close spectral proximity to water. Zhai, Wu, Qin, and Du [73] found that MNDWI performed substantially better than NDWI in mapping waterbodies that have similar spectral profiles to shadows, while Xu [41] showed that MNDWI performed significantly better than NDWI for extracting turbid water, which has a high spectral resemblance to some non-water classes. It should be noted, however, that these studies used spectral bands from Landsat 7 and Landsat 8, which differ from Sentinel-2 bands used in this study. However, our observations support those of Rokni et al. [74] and Zhou et al. [75], who found NDWI to be superior to other indices in delineating shallow and turbid lakes respectively. A likely explanation for NDWI performing better than MNDWI in our study was the study region. Although MNDWI is known to be more effective than NDWI in suppressing built-up features [9,11], it performed poorly in our study region, which is located in a rural setting. Nevertheless, the different OAs of NDWI and MNDWI suggest that the NIR and SWIR bands were more sensitive to the variations in physical and chemical properties of water than the green band.

It was observed that the SAR VH polarisation classified water more accurately than the VV polarisation did, irrespective of the targeted water characteristics. Classification errors at site H were mainly due to windy conditions at the time of acquisition, which created waves on the water surface and resulted in high backscatter signals. The VV polarization produced higher backscatter values over water surfaces than the VH polarization, which suggests that the former configuration is more sensitive to variations between water and non-water features. A bigger difference between the backscatter responses of land and water features was noted in the VH polarization than in the VV polarization. This corresponds well with Clement et al. [76] who also noted that VH outperformed VV polarization for turbid water mapping. Our study observed that the refined Lee speckle filter can suppress the speckle effect and maintain details of the water boundary [14], which is important for the identification of water pixels at the water/soil interface.

In this study, the semi-automated MLAs were used for benchmarking the autonomous thresholding results. All multispectral and SAR features were included in the MLAs to produce a best-case scenario. Although inequality within the waterbodies (e.g., depth, colour, and sediment variations) has been shown to affect classification results when using remotely sensed data [75], this study has proved SVM to be less sensitive to intra-class variations compared to other classifiers. Moreover, SVM was credited with its ability to effectively separate classes that are spectrally similar (e.g., humic rich water and shadows). This was likely a major contributing factor to its outstanding performance in this study.

Challenges relating to different applications and data used were encountered when attempts were made to directly compare the findings of this study with those of previous studies. The majority of the published studies that focus on the use of MLAs for the supervised classification of RS data have been done for vegetation and crop type classification using Landsat data. However, the outcomes of this study are closely related to those of Sarp and Ozcelik [77], who revealed that machine learning algorithms marginally outperform thresholding.

Although MLAs (specifically SVM) outperformed the thresholding methods in individual sites and when the sites were combined (i.e., when complexity increased), the main drawback of supervised MLAs is their dependence on training data. The application of supervised approaches is limited to regions for which representative samples of labelled data are available. Once training samples are established, they can be reused and applied to images with different dates and even of different areas. However, the accuracy of the resulting classifications is usually negatively affected [24,78,79], mainly due to temporal and regional variations. Waterbodies are highly dynamic as they continuously fill up and empty, which makes the reuse of training sets very challenging and limits the operational implementation of supervised techniques for monitoring changes in surface water reservoirs.

Despite the relatively lower recorded accuracies of thresholding (compared to those of MLAs), it seems to be a viable solution for operational implementations. In contrast to supervised approaches that require training data and rule-set (expert system) approaches that make use of a set of static thresholds, thresholding generates dynamic rules (appropriate thresholds) that do not require human interaction or training data. However, our results show that the use of a single feature (rule) for thresholding produced relatively poor and unstable results. Combining the outputs of different thresholding results produced much better and more robust results. For instance, we combined the two best thresholding outputs (NDWI and VH Polarisation_{RL}) and found that the combination (using Boolean OR) of these SAR and multispectral features significantly improved the accuracy and stability of the surface water classifications. More work is needed to investigate the efficacy of other combinations of thresholding outputs. Furthermore, the differences in results between thresholding and MLAs can be related to many other issues, such as pre-processing (atmospheric effects), the thresholding algorithm used, and illumination geometry. Previous works have shown that larger variances between the water features and non-water features typically minimize the accuracy of water body mapping, especially when a small area of water bodies and large non-water features exist [80,81]. Future studies are recommended to quantitatively consider how the variations in depth and concentrations of sediments and chlorophyll can affect classification accuracies.

5. Conclusion

Accurate temporal and spatial changes for small waterbodies are critical for water security, drought monitoring, and crop irrigation decision-making. Remote sensing offers a reliable, cost-effective, and potentially autonomous alternative for surface water mapping of large and inaccessible areas. The recently launched Sentinel-1 and Sentinel-2 satellites provide fine spatial and temporal resolution remote sensing data, which makes it ideal for monitoring waterbodies at regional and even global scales.

In this study, we proposed an approach that combines automatic thresholding of near-concurrent NDWI (generated from Sentinel-1) and VH backscatter polarisations (generated from Sentinel-1) for mapping waterbodies (mainly reservoirs and dams) with diverse spectral and spatial characteristics. Waterbodies of different sizes and varying levels of turbidity, sedimentation, and eutrophication were

targeted. The resulting maps were compared to the classification performances of five machine-learning algorithms (MLAs), namely decision tree (DT), k-nearest neighbour (k-NN), random forests (RF), and two implementations of the support vector machine (SVM). The results showed that the physical and chemical properties of water significantly affected classification accuracies. The performance of the best machine learning classifier (SVM) and thresholding (NDWI) dropped by more than 10% when the complexity of the task was increased (i.e., when the classifiers were applied to all sites in combination). However, the union of the two best thresholding results (NDWI and VH_{RL}) was relatively accurate and stable, likely because it takes advantage of both SAR and multispectral data. Although several heterogeneous sites were used to evaluate the results, more work is needed to test whether the dynamic NDWI– $VH_{Pol_{RL}}$ rule-set will be as effective in other areas, on other water types, during different seasons, and under contrasting conditions. Other indices, such as the automated water extraction index (AWEI) and tasselled cap wetness transformation, should also be evaluated when the coefficients for Sentinel-2 bands are made available. In addition, when it comes to PCA, it might be useful to see the correlation of the various component images with the water locations; as one of the components might be a good water index.

In summary, the techniques and datasets evaluated in this study show much promise for the accurate classification of optically complex waterbodies. Moreover, the relatively accurate and stable classifications achieved when the multispectral and SAR data were fused and automatically thresholded are very encouraging and may provide a viable solution for the operational monitoring of surface waterbodies in the Winelands district of South Africa. The implementation of this technique will provide invaluable information for water management and water security.

Author Contributions: All authors contributed extensively to the work presented in this paper. T.B, A.V.N, and M.M contributed to the concept design and research development. Satellite data was processed by T.B. Analysis was carried out by all authors with significant contribution from T.B and A.V.N. T.B prepared the manuscript. All authors read and approved the manuscript. The coding was prepared by S.M.

Funding: This research was carried out under the framework of ESA's (European Space Agency) ALCANTARA Initiative (4000112465/ 14/F/MOS 14-P11), and was facilitated by Delft University of Technology, The Netherlands, and Stellenbosch University, Graduate School, South Africa. Data was provided by ESA under the project ID C1F.3105.

Acknowledgments: The authors thank: The Western Cape Department of Water and Sanitation for providing access to dams; the Centre for Geographical Analysis at Stellenbosch University for the use of the classification and accuracy assessment software, SLICE, developed by Gerhard Myburgh; TU Delft University for the visitation and help with coding received from Silvia Alfieri and www.linguafix.net for language editing. The authors would like to thank the anonymous reviewers for their helpful comments and constructive suggestions.

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

Appendix A

Table A1. NDWI confusion matrix.

	Water	Trees & Shrubs	Bare & Built	Grass	Shadow	TOTALS	PA*%	EO+%
Water	3607	69	88	54	183	4160	86.7	13.3
Trees & shrubs	67	2645	123	326	68	3329	79.5	20.5
Bare & built	104	87	4162	189	248	4990	83.4	16.6
Grass	27	458	101	1794	66	2496	71.9	28.1
Shadow	119	38	62	21	1425	1665	85.6	14.4
TOTALS	3924	3297	4536	2384	1990	16,640		
CA‡%	91.9	80.2	91.7	75.3	71.6			
EC§%	8.1	19.8	8.3	24.7	28.4			
Overall accuracy	81.6							
Overall kappa	0.76							

*PA = Producer's accuracy; †EO = Errors of omission; ‡CA = Consumer's accuracy; §EC = Errors of commission.

Table A2. VH polarisation confusion matrix.

	Water	Trees & Shrubs	Bare & Built	Grass	Shadow	TOTALS	PA*%	EO ⁺ %
Water	3597	83	79	89	262	4110	86.5	13.5
Trees & shrubs	88	2679	60	218	284	3329	80.5	19.5
Bare & built	179	193	3805	381	145	4703	76.3	23.7
Grass	41	303	158	1920	49	2471	76.9	23.1
Shadow	168	104	162	63	1168	1665	70.1	29.9
TOTALS	4073	3362	4264	2671	1908	16,640		
CA [‡] %	88.3	79.7	89.2	71.9	61.2			
EC [§] %	11.7	20.3	10.8	28.1	38.8			
Overall accuracy	77.7							
Overall kappa	0.71							

*PA = Producer's accuracy; †EO = Errors of omission; ‡CA = Consumer's accuracy; EC = Errors of commission.

Table A3. MNDWI confusion matrix.

	Water	Trees & Shrubs	Bare & Built	Grass	Shadow	TOTALS	PA*%	EO ⁺ %
Water	3271	85	189	56	469	4160	83.5	16.5
Trees & shrubs	69	2879	97	128	56	3329	78.2	21.8
Bare & built	90	126	4271	147	186	4990	73.8	26.2
Grass	49	401	93	1625	48	2496	74.1	25.9
Shadow	463	49	97	89	857	1665	70.2	29.8
TOTALS	3942	3540	4747	2045	1616			
CA [‡] %	83	81.3	89	79.5	53			
EC [§] %	17	18.7	11	20.5	47			
Overall accuracy	73.8							
Overall kappa	0.69							

*PA = Producer's accuracy; †EO = Errors of omission; ‡CA = Consumer's accuracy; EC = Errors of commission.

References

1. Araujo, J.A.; Abiodun, B.J.; Crespo, O. Impacts of drought on grape yields in western cape, South Africa. *Theor. Appl. Climatol.* **2016**, *123*, 117–130. [CrossRef]
2. Botai, C.; Botai, J.; de Wit, J.; Ncongwane, K.; Adeola, A. Drought characteristics over the western cape province, South Africa. *Water* **2017**, *9*, 876. [CrossRef]
3. Gilbertson, J.K.; Kemp, J.; Van Niekerk, A. Effect of pan-sharpening multi-temporal landsat 8 imagery for crop type differentiation using different classification techniques. *Comput. Electron. Agric.* **2017**, *134*, 151–159. [CrossRef]
4. DAFF. Abstract of Western Cape Province Agricultural Statistics. 2018. Available online: <http://www.daff.gov.za/Daffweb3/Portals/0/Statistics%20and%20Economic%20Analysis/Statistical%20Information/Abstract%202018.pdf> (accessed on 23 April 2018).
5. Hoffman, M.T.; Carrick, P.; Gillson, L.; West, A. Drought, climate change and vegetation response in the succulent karoo, South Africa. *S. Afr. J. Sci.* **2009**, *105*, 54–60. [CrossRef]
6. Engelbrecht, C.J.; Landman, W.A.; Engelbrecht, F.A.; Malherbe, J. A synoptic decomposition of rainfall over the cape south coast of South Africa. *Clim. Dyn.* **2015**, *44*, 2589–2607. [CrossRef]
7. Evans, J. Western cape dam levels drop even more. *Mail & Guardian*, 5 March 2018.
8. Bangira, T.; Maathuis, B.H.; Dube, T.; Gara, T.W. Investigating flash floods potential areas using ascat and trmm satellites in the western cape province, South Africa. *Geocarto Int.* **2015**, *30*, 737–754. [CrossRef]
9. Du, Y.; Zhang, Y.; Ling, F.; Wang, Q.; Li, W.; Li, X. Water bodies' mapping from sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sens.* **2016**, *8*, 354. [CrossRef]
10. Hanqiu, X. A study on information extraction of waterbody with the modified normalized difference water index (mndwi). *J. Remote Sens.* **2005**, *5*, 589–595.
11. Yang, X.; Zhao, S.; Qin, X.; Zhao, N.; Liang, L. Mapping of urban surface water bodies from sentinel-2 msi imagery at 10 m resolution via ndwi-based image sharpening. *Remote Sens.* **2017**, *9*, 596. [CrossRef]
12. Bangira, T.; Alfieri, S.; Menenti, M.; van Niekerk, A.; Vekerdy, Z. A spectral unmixing method with ensemble estimation of endmembers: Application to flood mapping in the caprivi floodplain. *Remote Sens.* **2017**, *9*, 1013. [CrossRef]

13. Schlaffer, S.; Chini, M.; Dettmering, D.; Wagner, W. Mapping wetlands in zambia using seasonal backscatter signatures derived from envisat asar time series. *Remote Sens.* **2016**, *8*, 402. [[CrossRef](#)]
14. Pham-Duc, B.; Prigent, C.; Aires, F. Surface water monitoring within cambodia and the vietnamese mekong delta over a year, with sentinel-1 sar observations. *Water* **2017**, *9*, 366. [[CrossRef](#)]
15. Chini, M.; Hostache, R.; Giustarini, L.; Matgen, P. A hierarchical split-based approach for parametric thresholding of sar images: Flood inundation as a test case. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6975–6988. [[CrossRef](#)]
16. Zhang, X.-K.; Zhang, X.; Lan, Q.; Baig, M.H.A. Automated Detection of Coastline Using Landsat tm Based on Water Index and Edge Detection Methods. Paper Presented at Second International Workshop on Earth Observation and Remote Sensing Applications, Shanghai, China, 8–11 June 2012; pp. 153–156.
17. Pierdicca, N.; Pulvirenti, L.; Chini, M.; Guerriero, L.; Candela, L. Observing floods from space: Experience gained from cosmo-skymed observations. *Acta Astronaut.* **2013**, *84*, 122–133. [[CrossRef](#)]
18. Foody, G.M.; Muslim, A.M.; Atkinson, P.M. Super-resolution mapping of the waterline from remotely sensed data. *Int. J. Remote Sens.* **2005**, *26*, 5381–5392. [[CrossRef](#)]
19. Niroumand-Jadidi, M.; Vitti, A. Reconstruction of river boundaries at sub-pixel resolution: Estimation and spatial allocation of water fractions. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 383. [[CrossRef](#)]
20. Li, L.; Chen, Y.; Xu, T.; Liu, R.; Shi, K.; Huang, C. Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm. *Remote Sens. Environ.* **2015**, *164*, 142–154. [[CrossRef](#)]
21. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated water extraction index (awei): A new technique for surface water mapping using landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35. [[CrossRef](#)]
22. Otsu, N. A threshold selection method from gray-level histograms. *Automatica* **1975**, *11*, 23–27. [[CrossRef](#)]
23. Martinis, S.; Twele, A.; Voigt, S. Unsupervised extraction of flood-induced backscatter changes in sar data using markov image modeling on irregular graphs. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 251–263. [[CrossRef](#)]
24. Hasmadi, M.; Pakhriazad, H.; Shahrin, M. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geogr. Malays. J. Soc. Space* **2009**, *5*, 1–10.
25. Xie, H.; Luo, X.; Xu, X.; Pan, H.; Tong, X. Evaluation of landsat 8 oli imagery for unsupervised inland water extraction. *Int. J. Remote Sens.* **2016**, *37*, 1826–1844. [[CrossRef](#)]
26. Pradhan, B.; Tehrany, M.S.; Jebur, M.N. A new semiautomated detection mapping of flood extent from terrasar-x satellite image using rule-based classification and taguchi optimization techniques. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4331–4342. [[CrossRef](#)]
27. Feng, Q.; Gong, J.; Liu, J.; Li, Y. Flood mapping based on multiple endmember spectral mixture analysis and random forest classifier—The case of yuyao, China. *Remote Sens.* **2015**, *7*, 12539. [[CrossRef](#)]
28. Verpoorter, C.; Kutser, T.; Tranvik, L. Automated mapping of water bodies using landsat multispectral data. *Limnol. Oceanogr. Methods* **2012**, *10*, 1037–1050. [[CrossRef](#)]
29. Gilbertson, J.K.; Van Niekerk, A. Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. *Comput. Electron. Agric.* **2017**, *142*, 50–58. [[CrossRef](#)]
30. Harding, W.R.; Quick, A.J.R. Management options for shallow hypertrophic lakes, with particular refernce to zeekovlei, western cape. *S. Afr. J. Aquat. Sci.* **1992**, *18*, 3–19.
31. Dalvie, M.A.; Cairncross, E.; Solomon, A.; London, L. Contamination of rural surface and ground water by endosulfan in farming areas of the Western Cape, South Africa. *Environ. Health* **2003**, *2*, 1. [[CrossRef](#)]
32. Hong, S.; Jang, H.; Kim, N.; Sohn, H.-G. Water area extraction using radarsat sar imagery combined with landsat imagery and terrain information. *Sensors* **2015**, *15*, 6652–6667. [[CrossRef](#)]
33. Frazier, P.S.; Page, K.J. Waterbody detection and delineation with landsat tm data. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 1461–1468.
34. Yang, X.; Chen, L. *Evaluation of Automated Urban Surface Water Extraction from Sentinel-2a Imagery Using Different Water Indices*; SPIE: Bellingham, DC, USA, 2017; p. 11.
35. Donlon, C.; Berruti, B.; Buongiorno, A.; Ferreira, M.H.; Féménias, P.; Frerick, J.; Goryl, P.; Klein, U.; Laur, H.; Mavrocordatos, C.; et al. The global monitoring for environment and security (gmes) sentinel-3 mission. *Remote Sens. Environ.* **2012**, *120*, 37–57. [[CrossRef](#)]
36. Matthews, M.W. Eutrophication and cyanobacterial blooms in South African inland waters: 10years of meris observations. *Remote Sens. Environ.* **2014**, *155*, 161–177. [[CrossRef](#)]

37. Matthews, M.W.; Bernard, S. Eutrophication and cyanobacteria in South Africa's standing water bodies: A view from space. *S. Afr. J. Sci.* **2015**, *111*, 1–8. [[CrossRef](#)]
38. Louis, J.; Debaecker, V.; Pflug, B.; Main-Knorn, M.; Bieniarz, J.; Mueller-Wilm, U.; Cadau, E.; Gascon, F. Sentinel-2 sen2cor: L2a processor for users. Paper Presented at Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; pp. 1–8.
39. 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](#)]
40. Wilson, E.H.; Sader, S.A. Detection of forest harvest type using multiple dates of landsat tm imagery. *Remote Sens. Environ.* **2002**, *80*, 385–396. [[CrossRef](#)]
41. 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](#)]
42. Shen, L.; Li, C. Waterbody extraction from landsat etm+ imagery using adaboost algorithm. Paper presented at 18th International Conference on Geoinformatics, Beijing, China, 18–20 June 2010; pp. 1–4.
43. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpener. U.S. Patent 6 011 875, 4 January 2000.
44. Shensa, M.J. The discrete wavelet transform: Wedding the a trous and mallat algorithms. *IEEE Trans. Signal Process.* **1992**, *40*, 2464–2482. [[CrossRef](#)]
45. Kalantari, Z.; Nickman, A.; Lyon, S.W.; Olofsson, B.; Folkesson, L. A method for mapping flood hazard along roads. *J. Environ. Manag.* **2014**, *133*, 69–77. [[CrossRef](#)]
46. Hall-Beyer, M. Practical guidelines for choosing glcm textures to use in landscape classification tasks over a range of moderate spatial scales. *Int. J. Remote Sens.* **2017**, *38*, 1312–1338. [[CrossRef](#)]
47. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [[CrossRef](#)]
48. Zhang, X.; Cui, J.; Wang, W.; Lin, C. A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. *Sensors* **2017**, *17*, 1474. [[CrossRef](#)]
49. Salehi, M.; Mohammadzadeh, A.; Maghsoudi, Y. Adaptive speckle filtering for time series of polarimetric sar images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 2841–2848. [[CrossRef](#)]
50. Lee, J.-S.; Jurkevich, L.; Dewaele, P.; Wambacq, P.; Oosterlinck, A. Speckle filtering of synthetic aperture radar images: A review. *Remote Sens. Rev.* **1994**, *8*, 313–340. [[CrossRef](#)]
51. Lee, J.; Ainsworth, T.L.; Wang, Y. A review of polarimetric sar speckle filtering. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 5303–5306.
52. Argenti, F.; Lapini, A.; Bianchi, T.; Alparone, L. A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–35. [[CrossRef](#)]
53. 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. [[CrossRef](#)]
54. Skidmore, A.K.; Watford, F.; Luckananurug, P.; Ryan, P. An operational gis expert system for mapping forest soils. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 501–511.
55. Mueller, N.; Lewis, A.; Roberts, D.; Ring, S.; Melrose, R.; Sixsmith, J.; Lymburner, L.; McIntyre, A.; Tan, P.; Curnow, S. Water observations from space: Mapping surface water from 25 years of landsat imagery across australia. *Remote Sens. Environ.* **2016**, *174*, 341–352. [[CrossRef](#)]
56. Liu, Z.; Yao, Z.; Wang, R. Assessing methods of identifying open waterbodies using landsat 8 oli imagery. *Environ. Earth Sci.* **2016**, *75*, 873. [[CrossRef](#)]
57. Al-Bayati, M.; El-Zaart, A. Automatic thresholding techniques for sar images. In Proceedings of the International Conference of Soft Computing, Dubai, United Arab Emirates, 2–3 November 2013; pp. 18–19.
58. Liu, Y. Why ndwi threshold varies in delineating waterbody from multitemporal images? In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 4375–4378.
59. Myburgh, G.; Van Niekerk, A. Impact of training set size on object-based land cover classification: A comparison of three classifiers. *Int. J. Appl. Geospatial Res.* **2014**, *5*, 49–67. [[CrossRef](#)]
60. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Cham, Switzerland, 2013.

61. Steinwart, I.; Christmann, A. *Support Vector Machines*; Springer Science & Business Media: Cham, Switzerland, 2008.
62. Jia, K.; Wu, B.; Li, Q. Crop classification using hj satellite multispectral data in the north china plain. *APPRES* **2013**, *7*, 073576. [[CrossRef](#)]
63. Sun, D.; Yu, Y.; Goldberg, M.D. Deriving water fraction and flood maps from modis images using a decision tree approach. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2011**, *4*, 814–825. [[CrossRef](#)]
64. Mather, P.; Tso, B. *Classification Methods for Remotely Sensed Data*; CRC Press: Boca Raton, FL, USA, 2016.
65. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [[CrossRef](#)]
66. Amani, M.; Salehi, B.; Mahdavi, S.; Granger, J.; Brisco, B. Wetland classification in newfoundland and labrador using multi-source sar and optical data integration. *GISci. Remote Sens.* **2017**, *54*, 779–796. [[CrossRef](#)]
67. Garage, W. Opencv 2.0 and 2.2 Open Source Computer Vision Library. 2011. Available online: <http://opencv.willowgarage.com/wiki/> (accessed on 15 January 2011).
68. Campbell, J.B.; Wynne, R.H. *Introduction to Remote Sensing*, 5th ed.; Guilford Press: New York, NY, USA, 2011.
69. Qian, Y.; Zhou, W.; Yan, J.; Li, W.; Han, L. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sens.* **2015**, *7*, 153–168. [[CrossRef](#)]
70. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2008.
71. Ballanti, L.; Blesius, L.; Hines, E.; Kruse, B. Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sens.* **2016**, *8*, 445. [[CrossRef](#)]
72. Adedokun, O.A.; Burgess, W.D. Analysis of paired dichotomous data: A gentle introduction to the mcneemar test in spss. *J. MultiDiscip. Eval.* **2011**, *8*, 125–131.
73. Zhai, K.; Wu, X.; Qin, Y.; Du, P. Comparison of surface water extraction performances of different classic water indices using oli and tm imageries in different situations. *Geo-Spat. Inf. Sci.* **2015**, *18*, 32–42. [[CrossRef](#)]
74. Rokni, K.; Ahmad, A.; Selamat, A.; Hazini, S. Water feature extraction and change detection using multitemporal landsat imagery. *Remote Sens.* **2014**, *6*, 4173–4189. [[CrossRef](#)]
75. Zhou, Y.; Dong, J.; Xiao, X.; Xiao, T.; Yang, Z.; Zhao, G.; Zou, Z.; Qin, Y. Open surface water mapping algorithms: A comparison of water-related spectral indices and sensors. *Water* **2017**, *9*, 256. [[CrossRef](#)]
76. Clement, M.A.; Kilsby, C.G.; Moore, P. Multi-temporal synthetic aperture radar flood mapping using change detection. *J. Flood Risk Manag.* **2018**, *11*, 152–168. [[CrossRef](#)]
77. Sarp, G.; Ozcelik, M. Waterbody extraction and change detection using time series: A case study of lake burdur, turkey. *J. Taibah Univ. Sci.* **2017**, *11*, 381–391. [[CrossRef](#)]
78. Verhulp, J.; Van Niekerk, A. Transferability of decision trees for land cover classification in a heterogeneous area. *S. Afr. J. Geomat.* **2017**, *6*, 30–46. [[CrossRef](#)]
79. Ireland, G.; Volpi, M.; Petropoulos, G. Examining the capability of supervised machine learning classifiers in extracting flooded areas from landsat tm imagery: A case study from a mediterranean flood. *Remote Sens.* **2015**, *7*, 3372–3399. [[CrossRef](#)]
80. Li, W.; Du, Z.; Ling, F.; Zhou, D.; Wang, H.; Gui, Y.; Sun, B.; Zhang, X. A comparison of land surface water mapping using the normalized difference water index from tm, etm+ and ali. *Remote Sens.* **2013**, *5*, 5530–5549. [[CrossRef](#)]
81. Du, Z.; Li, W.; Zhou, D.; Tian, L.; Ling, F.; Wang, H.; Gui, Y.; Sun, B. Analysis of landsat-8 oli imagery for land surface water mapping. *Remote Sens. Lett.* **2014**, *5*, 672–681. [[CrossRef](#)]

