Surface Water Body Detection in Polarimetric SAR Data Using Contextual Complex Wishart Classification

E. Goumehei[1] , V. Tolpekin[2] , A. Stein[2], and W. Yan[1]

[1]Graduate School of Media and Governance, Keio University, Fujisawa, Japan, [2]Department of Earth Observation Science, Faculty of Geoinformation Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

Abstract Detection of surface water from satellite images is important for water management purposes like for mapping flood extents, inundation dynamics, and water resources distributions. In this research, we introduce a supervised contextual classification model to detect surface water bodies from polarimetric Synthetic Aperture Radar (SAR) data. A complex Wishart Markov Random Field (WMRF) combines Markov Random Fields with the complex Wishart distribution. It is applied on Single Look Complex Sentinel 1 data. Using Markov Random Fields, we utilize the geometry of surface water to remove speckle from SAR images. Results were compared with the Wishart Maximum Likelihood Classification (WMLC), the Gaussian Maximum Likelihood Classification, and a median filter followed by thresholding. Experiments demonstrate that the statistical representation of data using the Wishart distribution improves the *F*-score to 0.95 for WMRF, while it is 0.67, 0.88, and 0.91 for Gaussian Maximum Likelihood Classification, WMLC, and thresholding, respectively. The main improvement in the precision increases from 0.80 and 0.86 for WMLC and thresholding to 0.96 for WMRF. The WMRF model accurately distinguishes classes that have a similar backscatter, like water and bare soil. Hence, the high accuracy of the proposed WMRF model is a result of its robustness for water detection from Single Look Complex data. We conclude that the proposed model is a great improvement on existing methods for the detection of calm surface water bodies.

1. Introduction

Surface water body detection using satellite data has been addressed in many studies. The retrieved information has been utilized for water management tasks like indication the presence of water bodies and their extent, inundation dynamics, and flood extent. Yet there is at present insufficient knowledge of the spatial and temporal dynamics of available surface water (Alsdorf et al., 2007), since the variation of the spatial extent of inland water bodies both seasonally and interannually is strong (Papa et al., 2010). Also, over the past three decades several permanent water bodies have vanished or become seasonal, due to human and natural causes. Over 70% of global permanent water loss has occurred in five countries. Iran is among these five countries with 56% loss of permanent surface water between 1984 and 2015 (Pekel et al., 2016). Such losses raise major concerns of water security and sustainability. It is thus important to obtain accurate and updated information about the distribution of available surface water bodies.

Remote sensing technology provides advanced means for detecting, characterizing, and monitoring water bodies. It overcomes shortcomings of traditional ground-based surveys, such as being expensive, time-consuming, and influenced by other unknown factors in the field (Wang et al., 2011). Synthetic Aperture Radar (SAR) data have many advantages over optical images as they are independent of cloud cover; the sensors are able to operate day and night and are not subject to sun glint (Kutser et al., 2009). Applicability of SAR data for surface water detection has been demonstrated in the past (Henry et al., 2006; Hoque et al., 2011; Mertes, 2002; Tholey et al., 1997). Thresholding methods have been extensively used based upon the assumption of a strong contrast between the low backscatter of water and the higher backscatter of main land cover classes in the intensity images (Brisco et al., 2009; J. Li & Wang, 2015; White et al., 2014). Since backscatter values vary depending upon the incidence angle, image quality, and wind-induced surface water roughness, the threshold needs to be modified on a scene-by-scene basis and automating thresholding methods is still a challenge (Bolanos et al., 2016). Also, active contour methods have been applied for water mapping in SAR images (Hahmann & Wessel, 2010; Heremans et al., 2003; Horritt et al., 2001; Mason et al., 2007; Silveira & Heleno, 2009). Although these studies have improved thresholding methods

for water delineation, they involve postprocessing steps that rely upon availability of ancillary data to determine candidate pixels for water as well as on morphological operators (Bolanos et al., 2016). Recently, segmentation algorithms using auxiliary data have provided more successful results (Hahmann et al., 2010; Martinis et al., 2009). Their dependence upon the availability of high-resolution digital elevation models or a predetermined water mask limits their usefulness. In addition, they are not useful for mapping small ephemeral water bodies (Bolanos et al., 2016). Automatic processing chains also have recently been applied for inland water and flood mapping. Huang et al. (2017) and Twele et al. (2016) used threshold-based algorithms and Shuttle Radar Topography Mission (SRTM) data for water detection and compared automated classification results from using SRTM and Dynamic Surface Water Extent (DSWE). Bioresita et al. (2018) defined an automatic processing chain using the Finite Mixture Models to produce probability maps. Although encouraging results have been obtained, the preprocessing steps are time consuming and they rely as well on predetermined water mask data.

Using contextual information in classification of optical and SAR images leads to improvement in the accuracy and reliability of classification (Ardila et al., 2011; Hiremath et al., 2013). The potential of Markov Random Field (MRF) to effectively integrate contextual information associated with the image data during the analysis is desired, and research has been applied on MRF for SAR image classification (Fjortoft et al., 2003; Kenduiywo et al., 2014; Moser & Serpico, 2009; Reigber et al., 2010; Serpico & Moser, 2006). According to criteria of the maximum a posteriori probability (MAP), MRFs allow a global Bayesian optimization of the classification results. Also, a complex Wishart distribution in classification of polarimetric SAR images has been applied (Akbari et al., 2011; Doulgeris et al., 2008, 2011; Frery et al., 2007; Lee et al., 1994). Akbari et al. (2012) proposed an unsupervised contextual clustering model for classification of multilook SAR images with a κ-Wishart distribution for data statistics and the Pott model for the spatial context. Their research shows a clear improvement in using an appropriate statistical representation, but it does not use any backscattering signature of classes. Therefore, it affects detection of any class of interest, which in our case is water. A major and recurring problem is that this class could be mixed with other similar classes and that it could be completely missed by an unsupervised classifier.

The main objective of this research is to develop a contextual supervised algorithm based upon Bayesian statistics for mapping surface water bodies from Single Look Complex (SLC) images. We use complex values of single look polarimetric SAR data to avoid speckle filtering and information loss, whereas the high resolution of SLC images is preserved. We use a contextual model that effectively tackles speckle in SAR images. Using MRFs, we are utilizing the geometry of surface water to remove speckle in SAR images and improve classification. The proposed model is applied on Sentinel 1 data, and results are compared with the Gaussian Maximum Likelihood Classification (GMLC) Wishart Maximum Likelihood Classification (WMLC) models and thresholding. We finally investigate the effect of class definition by using different numbers of classes.

2. Methodology

2.1. WMRF

Bayesian statistics is widely used in remote sensing classification. Based upon backscatter values, we classify each pixel to the desired categories. The pixel values of a SAR image are denoted by $d = \{d_i, i = 1, ..., L\}$, where L represents the entire image. Each pixel takes a label from user-defined information classes defined as $w_i \in \{1, ..., m\}$, where m is the number of classes. The classification criterion is based upon the MAP probability criterion which maximizes the product of conditional and prior probability. The posterior probability can be expressed in terms of the posterior energy function (Li, 2009)

$$P(w_i|d_i) \propto e^{-U(w_i|d_i)}, \tag{1}$$

where

$$U(w_i|d_i) = U(w_i|w_{N_i}) + U(d_i|w_i). \tag{2}$$

Here, $U(w_i|w_{N_i})$ is the prior energy function and $U(d_i|w_i)$ is the conditional energy function for pixel i. Use of energy functions allows a more convenient way to express contextual relation compared to probability (Geman & Geman, 1984).

The *prior energy*, $U(w_i|w_{N_i})$, is modeled as an MRF for each pixel i, adapted to a neighborhood system, N_i. For an image with L pixels, it is defined as follows:

$$U(w_i|w_{N_i}) = \sum_{i=1}^{L} \sum_{l \in N_i} \omega(s_{il}) \theta(w_i, w_l),$$ (3)

where s_{il} is the distance from pixel i to pixel l, in the neighborhood N_i of pixel i, and $\omega(s_{il})$ denotes the weight of contribution from $l \in N_i$ to the prior energy. The term $\theta(w_i, w_l)$ is defined as $\theta(a,b) = 1$ if $a \neq b$, and 0 otherwise. Weights s_{il} are defined as $\omega(s_{il}) \sim \frac{1}{s_{il}}$ and normalized, such that $\sum_{l \in N_i} \omega(s_{il}) = 1$. This term favors smooth class labels in the neighborhood and penalizes deviations for smooth classification.

The *conditional* energy, $U(d_i|w_i)$, is modeled by a complex Wishart density function (Goodman, 1985). Dual polarization sensors have a scattering matrix u with two elements that can be considered as a vector

$$\mathbf{u}_i = [S_{vh_i} \quad S_{vv_i}]^T.$$ (4)

where S_{vh_i} is the polarized backscattering element of a vertically transmitted and horizontally received signal for pixel i and S_{vv_i} is a vertically transmitted and vertically received signal. These elements are complex numbers since they carry both the magnitude and phase of the signal. Then the estimated single look covariance matrix equals

$$A_i = \sum_{i=1}^{n} \mathbf{u}_i \mathbf{u}_i^{\dagger},$$ (5)

where n is the number of pixels to estimate A_i. The distribution of A_i follows by a complex Wishart probability density function (Goodman, 1985)

$$P(A_i|w_i) = \frac{|A_i|^{n-q} \exp\left[-\text{Tr}\left((C_{w_i})^{-1} A_i\right)\right]}{K(n,q)|(C_{w_i})|^n},$$ (6)

where $\text{Tr}(B)$ is the trace of B, C_{w_i} is the complex covariance matrix given class w_i, and $K(n,q)$ is defined as

$$K(n,q) = \pi^{\frac{1}{2}q(q-1)} \Gamma(n)...\Gamma(n-q+1).$$ (7)

The parameter q is dimension of the matrix $\mathbf{u}_i$, and $\Gamma(.)$ is the Gamma function (Lee et al., 1994). Let d_i be flatten A_i, then $P(A_i|w_i) = P(d_i|w_i)$. Then, the corresponding conditional energy based on equation (6) is

$$U(d_i|w_i) = -\log P(d_i|w_i).$$ (8)

The posterior energy function, equation (2), based on equations (8) and (3) can be written as follows:

$$U(w|d) = \sum_{i=1}^{L} U(w_i|d_i) = \lambda \sum_{i=1}^{L} U(w_i|w_{N_i}) + (1-\lambda) \sum_{i=1}^{L} U(d_i|w_i).$$ (9)

Minimizing equation (8) with respect to w yields the MAP solution, where an additional parameter λ controls the relative contribution of the prior and conditional energy functions with $0 \leq \lambda < 1$. For $\lambda = 0$, the prior model is completely ignored and the classifier is not contextual. For $0 < \lambda < 1$ the MAP classifier is contextual, explicitly incorporating spatial contextual information by means of prior energy.

2.1.1. Energy Minimization

To maximize $P(w_i|d_i)$, the energy function equation (9) has to be minimized. In order to find the global minimum of the energy function, simulated annealing is employed (Geman & Geman, 1984; Metropolis et al., 1953)). This research applies the Metropolis-Hastings sampler (Geman & Geman, 1984). The algorithm starts at a high temperature $\tau = \tau_0$. The value of τ decreases using a cooling schedule. An iterative process follows until the system becomes is frozen ($\tau \to 0$). The temperature at the iteration k is changed such that

$$\tau_k = \sigma \times \tau_{k-1}, \tag{10}$$

for $\sigma \in (0; 1)$. Any τ_0 can be chosen for optimization, but its value can affect the solution. Optimal values of the annealing schedule (τ_0 and σ) depend upon complexity of the problem which in our study depends on class separability. For each iteration, the Metropolis-Hastings sampler updates all pixels and the number of successful updated pixels is counted which leads to a change of pixel value. A threshold of 0.1% of the total number of pixels (Tolpekin & Stein, 2009) is defined to stop the optimization process when for three consecutive iterations the counted updated pixels are below the threshold.

We compare our WMRF model with three current methods from the literature. One is the GMLC, which is based upon a Bayesian probabilistic framework, but without contribution of a prior energy function. For this classifier, the conditional energy, $U(d_i | w_i) = U(u_i | w_i)$, is modeled by a Gaussian density function

$$U(u_i | w_i) = \frac{1}{\pi^2 |C_{w_i}|} \exp\left(-u_i^* C_{w_i}^{-1} u_i\right) \tag{11}$$

that is, the distribution of u_i is assumed to follow a Gaussian distribution. Then, equation (9) can be written as follows:

$$U(w|d) = \sum_{i=1}^{L} U(w_i | d_i) = \sum_{i=1}^{L} U(d_i | w_i). \tag{12}$$

The second method for comparison is the WMLC which follows a similar framework as the WMRF model but only considers the conditional energy. For the WMLC model, the distribution of the single look covariance matrix is assumed to follow the complex Wishart distribution. Then the posterior energy function for all pixels in a SAR image is based on equation (8).

The third method for comparison is the thresholding method based upon the low backscattering of water with respect to land in SAR images. We applied a median filter followed by thresholding.

2.2. Accuracy Assessment

Evaluation of the results is done using the $m \times m$ confusion matrix as a common way to summarize performance of classification methods. A confusion matrix assesses the results by means of labeled pixels that relate classified data to reference data. Using the confusion matrix, two measures of precision and recall are used to evaluate the success rate of classifiers. Precision denotes the proportion of the number of true positive predictions divided by the total predicted positives, and recall is the number of true positive predictions divided by the total number of actual positives. We also use the F-score as a measure for accuracy assessment, defined as follows:

$$F = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \tag{13}$$

In addition, Cohen's κ (Hudson & Ramm, 1987) measures how closely the instance classified by the classifier matched the data labeled as reference data. To test the significance of results, we used the test statistic Z, (Congalton & Green, 2009) defined as follows:

$$Z = \frac{\kappa}{\sqrt{\text{var}(\kappa)}}. \tag{14}$$

It tests the null hypothesis $H_0 : Z = 0$ against the alternative hypothesis $H_1 : Z \neq 0$. Under the assumption of normality, we decide in favor of H_1 if $Z > 1.96$; otherwise, we decide in favor of H_0.

As our purpose is to detect water, the confusion matrix for more than two classes is formed by two elements: water and nonwater. More classes are defined during the classification to adequately model spectral variation of nonwater pixels and overcome the problem of misclassification due to the poor class separability of water and bare soil.

3. Experimental Results

3.1. Data Description

We used Sentinel-1 SAR image in Interferometric Wide swath mode to test the performance of the proposed method. Sentinel-1 is operating in C-band and collects images of a 250-km swath and a 5×20-m spatial resolution. The image acquisition date is 17 April 2017. We used a Level 1 SLC image with amplitude and phase of the SAR signal. It contains dual-polarization (VV+VH) data with both real and imaginary information. For preprocessing of the image, the Sentinel-1 toolbox (S1TBX) from SNAP software, version 5.0.8 of European Space Agency (ESA), was used (Veci et al., 2015). Interferometric Wide products have three subswaths, and each subswath image consists of a series of bursts, thus requiring debursting. Next, orbit file correction, radiometric calibration, and geometric terrain correction were applied. We used geometric terrain correction to reduce topographic effects. Range Doppler terrain correction (Small & Schubert, 2008) was applied, using SNAP software. The Digital Elevation Model for terrain correction was the SRTM elevation model of a 3-arcsecond resolution, whereas the nearest neighbor method was chosen as the resampling method. Complex values of the image data were preserved during the preprocessing steps. Figure 1 shows the intensity of the SLC image.

Two different reservoirs, located in the western part of Iran, were selected as the study area. These two reservoirs have different characteristic in terms of shape and size. Reservoir A is a reservoir with 10-km^3 capacity covered by $1,687 \times 649$ pixels, whereas Reservoir B has a triangular shape with a rocky island in the middle. It is smaller with capacity of 1 km^3 and is covered by 401×220 pixels. The two reservoirs are both in a mountainous environment that mainly consists of surface water bodies, bare soil, rocks, and agricultural lands.

Three different legends were considered: C2 (water, nonwater), C3 (water, bare soil, and others), and C4 (water, bare soil, rock, and others). Upon the classification results of C3 and C4, nonwater classes were recorded into a single class.

Training samples of Reservoir A contain 10 polygons for each class. The class water has 1,123 pixels, bare soil has 2,178 pixels, rocks has 1,876 pixels, and the class other includes 2,660 pixels. Training samples of Reservoir B consist of six polygons for each class. The number of training pixels for class water is 138, for bare soil it is 388, for rocks it is 245, and for the class other it is 173. Training samples for the two reservoirs were selected manually from optical images with a few days difference. Training samples were chosen based upon the literature, considering distribution, size, and balance aspects (Congalton & Green, 2009; Zhu et al., 2016).

The first reference set (rs1) contains 400 points for Reservoir A and 192 points for Reservoir B. These points were obtained by stratified random sampling using the WMRF classification results for C3. Among the 400 points, Reservoir A contains 136 points for water and 264 points for nonwater. Reservoir B has 74 points for water and 116 points for nonwater. These points are evenly distributed throughout the study area, and the points are labeled by an expert using visual interpretation of SAR and the Sentinel 2A optical image of the study area collected in 26 April 2017.To ensure that the rs1 data do not inform the calculation of accuracy metrics, a second reference set (rs2) was generated. This set has 400 stratified random points created using the same classified optical image of the area and contains 200 points for each class. We evaluated the WMRF classified images using the second reference dataset.

The algorithm was written in R (R Core Team, 2018), version 3.5.0. The Rcpp and rgdal packages have been used.

3.2. Classification

3.2.1. Simulated Annealing Parameter Optimization

Parameters for energy minimization part of classification were set equal to $\tau_0 = 4.0$ and $\sigma = 0.s9$. A neighboring system N_i was chosen as the eight nearest pixels. The covariance matrix in equation (4) was estimated by 5×5 neighboring pixels; hence, $n = 25$. In order to optimize energy minimization, other values of τ_0 and σ parameters of the algorithm have been tested for Reservoir A. The value $\lambda = 0.5$, corresponding to the highest accuracy of the model, was selected. We also varied τ_0 between 0 and 10 (Figure 2a). Each experiment was repeated 10 times to investigate variation of the results. Also, 100 repetitions were considered for $\tau_0 = 1.0$ and $\tau_0 = 4.0$ to check if the variance is estimated adequately with only 10 samples. The standard deviation (SD) of
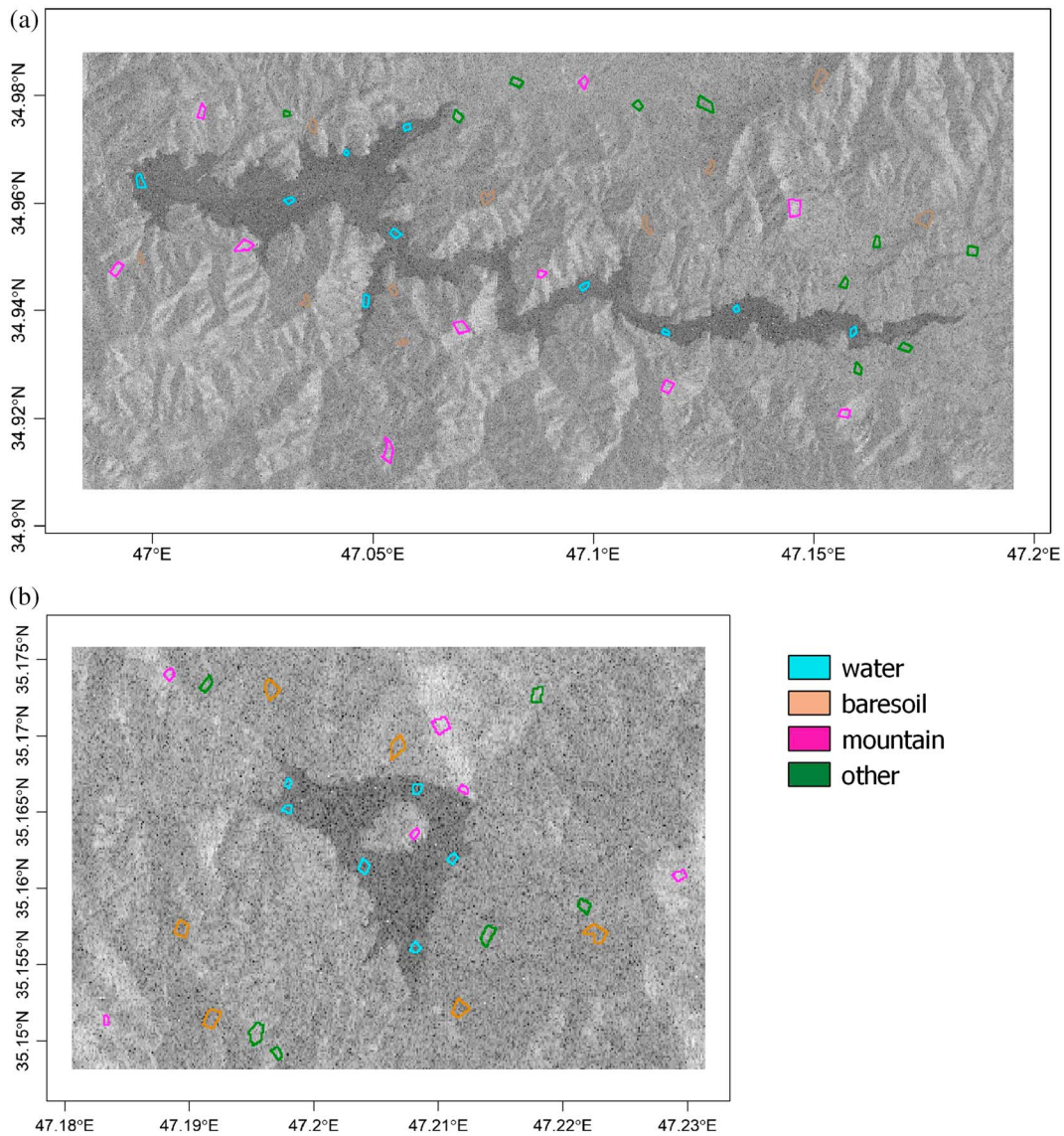
Figure 1. (top) Reservoir A and (bottom) Reservoir B. Polygons represent training samples.
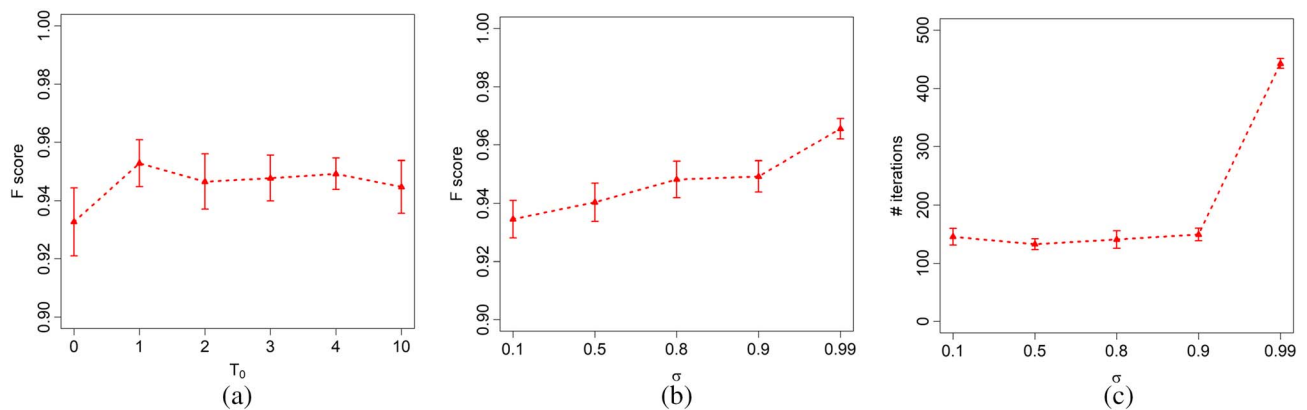
Figure 2. Parameter optimization results. (a) *F*-scores for different values of initial temperature (τ_0) with error bars indicating standard deviation for 10 repetitions, (b) *F*-scores for the selected value of $\tau_0 = 0.3$ to optimize the σ value with error bars indicating standard deviation for 10 repetitions, and (c) number of iterations for energy minimization for different σ values.

the F-score for $\tau_0 = 1.0$ and $\tau_0 = 4.0$ reduced with an amount of only 0.006 and 0.0003 from 10 to 100 repetitions, respectively. Therefore, we decided that 10 repetitions provided an adequate representation of the variation in the results.

The suggested range for initial temperature found in the literature is between 3 and 4. This corresponds with our optimized value, so we selected $\tau_0 = 4.0$ to optimize the value of σ. Figure 2b shows that the F-score is increasing with increasing σ values and that at the same time SD is decreasing. The results show that the optimal value equals $\sigma = 0.99$, corresponding with the highest F-score and the lowest SD values. In order to investigate the efficiency of the optimized value in terms of time consumption for energy minimization, the number of iterations for each value of σ is plotted in Figure 2c. From this figure we observe a slight growing in the number of iterations by increasing the value of σ. All σ values require approximately $k = 150$ iterations except for $\sigma = 0.99$, where more than $k = 400$ iterations were needed. As it turned out, there is trade-off between k and speed of the model. Hence, we suggest to select $\sigma = 0.9$ with admissible F-score and SD values, specifically for large data classification.

3.2.2. WMRF Model

To evaluate the performance of the WMRF model for surface water body detection, the WMRF was applied on the SLC data for different values of parameter λ. To deal with similar scaling variables, normalization has been applied on the conditional energy values, as we are dealing with complex values which cause unbalanced ranges of prior and conditional energy. Figure 3 shows average value of F-score, precision, and recall for 10 runs and their variance for $0 \leq \lambda \leq 1$. The model classified the data into two classes (C2 legend) as the maximum observed F-score = 0.73 for Reservoir A and 0.81 for Reservoir B. Accuracy increases smoothly with increasing λ. It reaches its maximum F-score for $\lambda = 0.9$, whereas for $\lambda = 1$ it drops to 0.41 and 0.44 for study areas A and B, respectively. The maximum observed precision equals 0.58 for Reservoir A and 0.70 for Reservoir B. Recall results show the highest performance for all values of λ, whereas recall = 1 for Reservoir A and recall = 0.97 for Reservoir B. These results show that the model can successfully classify water pixels as class water but for many pixels misclassifies nonwater pixels as water. The following section aims to overcome this problem by defining new classes.

3.2.3. Number of Classes

We now evaluate the choice for legends C2, C3, and C4. Results for C3 and C4 show a higher sensitivity to changing λ than C2 (Figure 4). We observe a remarkable improvement in F-score and precision results from C2 to C3 and C4, for both reservoirs. The highest F-score equals 0.95 for C3 and C4 and 0.72 for C2, for Reservoir A (Figure 4a). Reservoir B also experiences an improved F-score from 0.81 for C2 to 0.95 for C3 and C4 (Figure 4d). From Figures 4b and 4e, we note that the improvement of the results for C3 and C4 is even stronger in terms of precision as compared to C2. For Reservoir A, the precision increases to 0.94 for C3 and 0.93 for C4 from 0.58 for C2 (Figure 4b). A similar increase is observed for Reservoir B: The precision increases from 0.70 for C2 to 0.97 for C3 and C4 (Figure 4d).

Including a class bare soil in the legend remarkably improves the precision of water classification and only slightly decreases the recall of the classification. Recall of C3 and C4 decreases 3% to 4% with respect to C2 for Reservoir A (Figure 4c), because the class separability of water and bare soil is poor and these classes have overlapping distributions. For C2, all low backscatter pixels are assigned to class water, so recall is high and all water pixels are classified correctly. For C3 and C4, pixels in the overlapping area, which have low backscatter, can be labeled as either water or bare soil.

To ensure that the reference data, rs1, do not inform the calculation of accuracy metrics, we evaluated WMRF classified images using the second reference data set, rs2, and compared those with our results using pairwise the test statistics Z, for testing the significance of the difference between two independent error matrices (Congalton & Green, 2009). The test statistics is $Z = 0.16$ and shows that the results are not significantly different.

### 3.2.4. Comparison With GMLC, WMLC, and Thresholding

As the last experiment to assess the performance of the WMRF model, we compared its results with those from the GMLC, the WMLC models, and thresholding. Table 1 summarizes the results for both reservoirs. Results for the WMRF model provide the best classification as indicated by the highest observed values for all accuracy measurements for C2, C3, and C4. Differences among the models for C2 are similar for Reservoirs A and B, while F-score and κ for WMRF are 4% and 6% higher than the other two models. Recall for all models in both study areas is larger than 0.95, where performance of
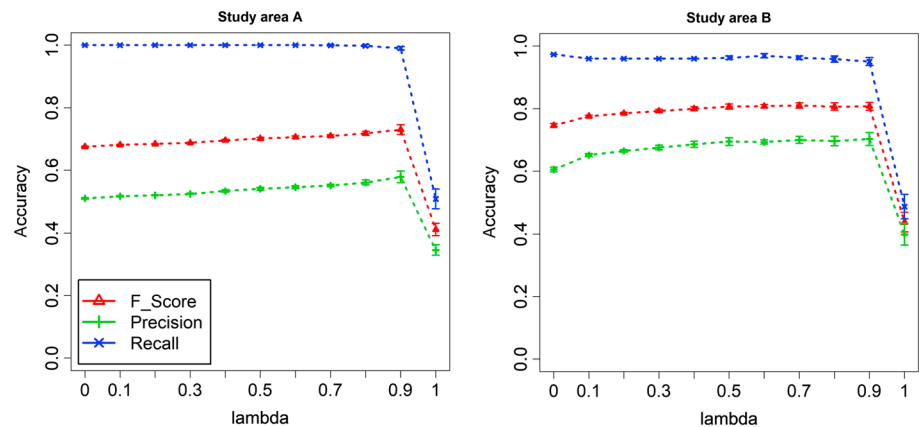
Figure 3. Classification results of Wishart Markov Random Field model for different values of λ, for two classes (C2) of water and nonwater. The lines are added to assist interpretation.

precision is not as high as recall. The WMRF model gets the highest precision of 0.55 and 0.70 for Reservoirs A and B, respectively.

The strength of the WMRF model is more obvious for C3 and C4. The *F*-score for the WMRF model raises to 0.95 from 0.88 for WMLC and 0.67 for GMLC. This increase is also evident for precision that increases from 0.81 and 0.62 for WMLC and GMLC to 0.94 for WMRF. This improvement of precision shows that the WMRF model benefits from contextual information for classifying nonwater pixels properly. It accurately turns individual predicted water pixels to nonwater label, using labels of neighboring pixels. Therefore, misclassified nonwater pixels will be correctly labeled (Figure 5).
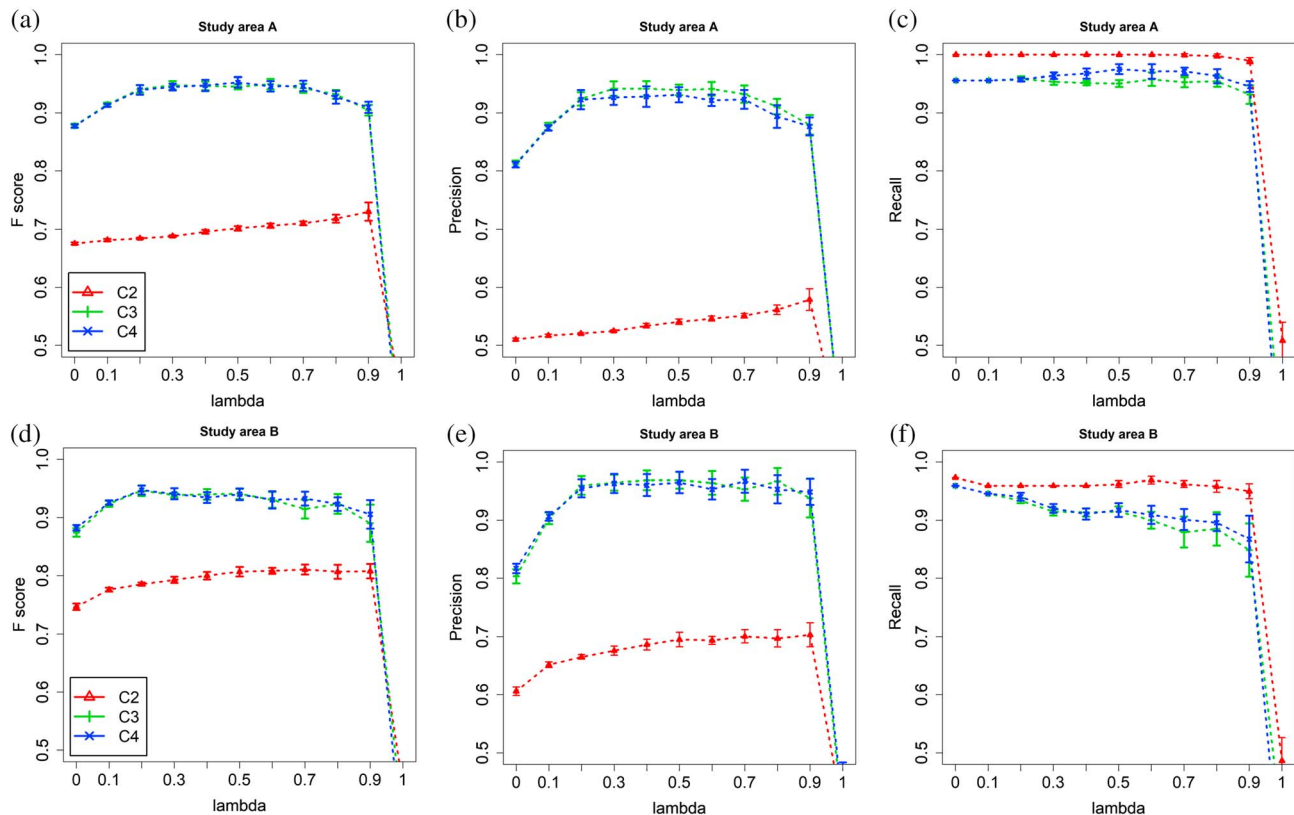
Figure 4. Wishart Markov Random Field classification results of different λ values for C2, C3, and C4: (a) *F*-score results, (b) precision, and (c) recall for Reservoir A and (d) *F*-score, (e) precision, and (f) recall for Reservoir B.

Table 1
Summary of Results for Three Different Classification Models and Three Legends, C2, C3, and C4 Defined in Section 3.1

	Reservoir A					Reservoir B				
	Precision	Recall	*F*-score	κ	Z	Precision	Recall	*F*-score	κ	Z
C2										
GMLC	0.52	0.97	0.68	0.42	8.78	0.66	0.95	0.78	0.59	10.40
WMLC	0.51	**1**	0.67	0.41	8.58	0.61	**0.97**	0.75	0.52	8.56
WMRF	**0.55**	0.98	**0.71**	**0.45**	11.14	**0.7**	0.96	**0.81**	**0.66**	12.35
C3										
GMLC	0.62	0.74	0.67	0.48	10.58	0.75	0.8	0.77	0.62	10.71
WMLC	0.81	**0.95**	0.88	0.81	26.57	0.8	**0.96**	0.87	0.78	17.08
WMRF	**0.94**	**0.95**	**0.95**	0.92	43.39	**0.96**	0.93	**0.95**	0.91	29.81
C4										
GMLC	0.62	0.74	0.67	0.48	10.58	0.75	0.8	0.77	0.62	10.71
WMLC	0.81	0.95	0.88	0.81	26.57	0.82	**0.96**	0.88	0.8	17.08
WMRF	**0.93**	**0.98**	**0.95**	0.92	45.58	**0.95**	0.94	**0.95**	0.91	29.97

Note. WMRF result is for optimum value of λ. Z is the test statistic for testing the significance of results. The bold numbers are presenting the highest value among the three compared models for each legend. GMLC = Gaussian Maximum Likelihood Classification; WMLC = Wishart Maximum Likelihood Classification; WMRF = Wishart Markov Random Field.

Performance of the WMRF model for both reservoirs is more robust than the other two models (Figure 6). The WMRF model acquires similar results for both study areas, specifically for C3 and C4, while results are varying for the GMLC model. The test statistic for testing the significance of κ represents the trustworthiness of the results. The high Z value of 43.39 and 29.81 for κ shows significance at the 95% confidence level, indicating that the results for WMRF model are substantially better than random.
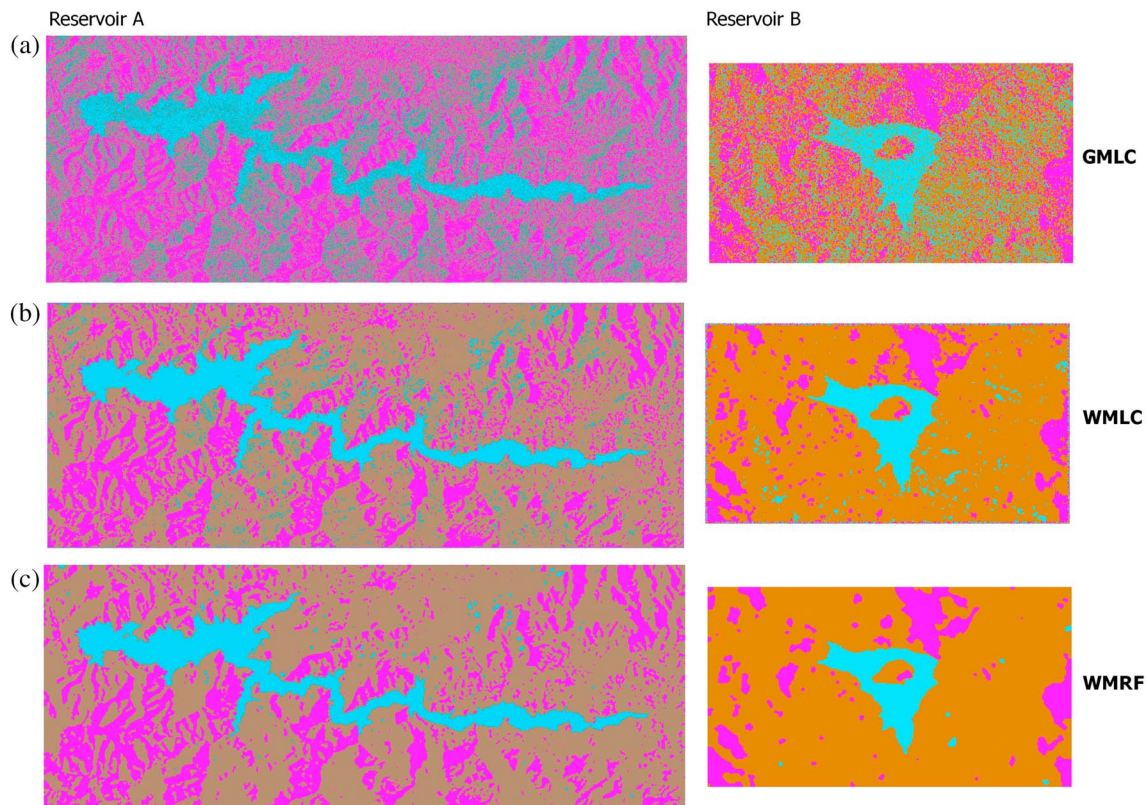
Figure 5. Classification results for (a) GMLC, (b) WMLC, and (c) WMRF models. GMLC = Gaussian Maximum Likelihood Classification; WMLC = Wishart Maximum Likelihood Classification; WMRF = Wishart Markov Random Field.
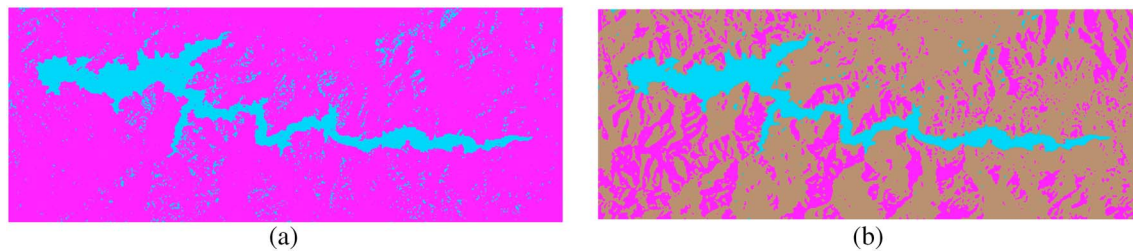
Figure 6. The comparison of water body detection (a) 5 × 5 median filter following thresholding and (b) Wishart Markov Random Field model.

Figure 5 shows the results for the WMRF, WMLC, and GMLC models. The WMRF model improves the classification results. Surface water in both WMRF and WMLC model is homogenous, and it can be clearly interpreted as a water object. Nonwater pixels that are misclassified as water are obvious in the GMLC and WMLC models classification, whereas the WMRF model achieves smoother results.

A final issue in the evaluation of the different models concerns the efficiency of the algorithms. Computation time for one run of the algorithm is on average 19.388 s for 149 iterations, that is, 0.130 s per iteration, so the model runs on average 1.187×10^{-7} s per pixel per iteration. The algorithm has run on an Ultrabook with processor of Intel® Core™ i7-6560U CPU @ 2.20 GHz. The compared models, GMLC and WMLC, have computation times equal to 9.41×10^{-7} s per pixel and 11.23×10^{-7} s per pixel, respectively.

For the third comparison we selected a median filter with size of 5 × 5 window, comparable with the neighborhood system of our WMRF model. We used 80 randomly chosen reference points to estimate the optimal threshold value. These points were removed from the reference set. Next, we compared the results of the thresholding with the results of the WMRF model using a pairwise test statistic, Z. We obtained $Z =$ 3.21 showing a significant difference in the results of two different methods. When a smaller 3 × 3 median filter is used, the pairwise test is even higher, $Z = 5.48$. In addition to the statistical analysis, also, a visual interpretation of the results showed the superiority of the proposed model. Figure 6 indicates that the thresholding method has a higher commission error (lower precision) in comparison to the thematic map of the proposed method. Detailed results are provided in Table 2.

4. Discussion

This study proposed a supervised contextual algorithm for SLC polarimetric SAR imagery. The main difficulty in mapping water in SAR images is differentiating water from other classes with similar backscattering properties.

When applying a contextual classifier for the detection of a specific class, one can decide on the number of background classes to be defined. One possibility is to define all spectral classes present in the image. This would lead to higher accuracy, although at the expense of additional training efforts. For a spectral

Table 2
The Accuracy Assessment of Thresholding and WMRF

	3 × 3 median filter		5 × 5 median filter		WMRF	
	Water	Nonwater	Water	Nonwater	Water	Nonwater
Water	239	34	285	45	130	6
Nonwater	47	262	9	243	8	256
kappa		0.72		0.81		0.92
precision		0.87		0.86		0.94
recall		0.83		0.97		0.96
F		0.85		0.91		0.95
Z		5.48		3.21		

Note. WMRF = Wishart Markov Random Field.

noncontextual classifier one can introduce an unclassified category. For a contextual classifier, however, this is inappropriate because contextual relationships between all class categories need to be explicitly defined. Our results show that a legend with a single background class, C2, is insufficient, because existence of a class with similar backscatter decreases the accuracy. Therefore, addition of a background class, bare soil (C3), spectrally similar to target class substantially improves the classification accuracy. Further splitting of background class, C4, however, did not lead to more improvement.

Our main focus was using SLC SAR data. In polarimetric SAR classification, most research, for example, Akbari et al. (2012) and Wu et al. (2008), used multilook data. Multilooking reduces speckle by averaging adjacent pixels at the expense of losing spatial resolution. Therefore, multilooked SAR data have a coarser resolution as compared to SLC data. The aim of using SCL data is to preserve finer resolution which is of interest for smaller surface water bodies. In this study, we use a contextual model that is effective to tackle speckle in SAR images. Using MRF, we build on the geometry of the surface water bodies to remove speckle from SAR images and improve classification. Common speckle filtering algorithms smooth away high-frequency information (Mather & Tso, 2009). Using MRF, class information of neighboring pixels are taken to account to suppression speckle.

One of the main advantages of using SLC data is the preservation of the spatial resolution of the sensor. This is most critical in small objects or objects with an irregular boundary, as is the case for reservoirs in mountainous areas considered in this paper. The proposed model is a supervised classifier so its application is not restricted to water body detection. The WMRF model assumes that the classes can be modeled with the complex Wishart distribution. As long as this assumption can be justified and the covariance matrices of the classes are different, the proposed model can distinguish different classes from SLC images. The reason is that the covariance matrix of a class includes information on the Radar Cross Section and polarimetric scattering mechanism of the object.

This research focused on homogenous surface water body detection. Although the study of surface water in arid and semiarid climate is of importance of this research, the integration of water and vegetation should be further investigated. Our study area is an arid and semiarid area without vegetation and reeds in water. Presence of vegetation is a problem, for example, in wetlands. The case study of this research is a calm water body with weak wind conditions. In case of wind or turbulent water, a choice of proper satellite data with appropriate wavelength is crucial. The Sentinel 1 SAR instrument operates in C-band corresponding to a radar wavelength of about 5.6 cm. This means that Sentinel 1 ignore water waves smaller than this wavelength, which is applicable for mapping calm water. For turbulent water or wind-generated waves higher than 5.6 cm, other satellite data with longer wavelengths are required. In addition, at substantial wave height in relation to radar wavelength, relative orientation and incidence angle with respect to wave should be considered.

Several studies, like Martinis et al. (2009), Hahmann et al. (2010), and Westerhoff et al. (2013), have benefited from using auxiliary data, such as digital elevation models. Although the use of auxiliary data improves the results, it may not be available everywhere. One benefit of our model is that it achieves high accuracy without use of any auxiliary data except for training data set. Therefore, our model is generally applicable in semiarid environments, even without auxiliary data.

During optimization, there is a clear trade-off between accuracy and computation time. Taking the advantage of high accuracy requires more computational time, which is reflected in the number of iterations. For large data sets, a higher value of σ increases the computational time, whereas a smaller data set can benefit from a higher accuracy and a lower number of iterations. The value of σ, however, can be controlled and be selected based on user preference.

5. Conclusions

This study showed how the complex Wishart distribution was satisfactorily incorporated into MRFs. The experimental results show a classification accuracy of 95% for two lakes in Iran. The high recall values for all experiments illustrate the strength of the model for correctly classifying water pixels. Using two classes only, we noted that pixels with low backscatter like bare soil are incorrectly labeled as water. Such a misclassification was prevented by splitting class water and bare soil which significantly improves

accuracy. We concluded that in the case of calm surface water bodies, the WMRF can perform robustly. The strength of the proposed model is a high classification accuracy of SLC polarimetric SAR data. It is reliable for differentiating classes with similar backscatter.

References

Akbari, V., Doulgeris, A. P., Moser, G., Eltoft, T., Anfinsen, S. N., & Serpico, S. B. (2012). A textural model for unsupervised segmentation of multipolarization synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(4), 2442–2453. https://doi.org/10.1109/TGRS.2012.2211367

Akbari, V., Moser, G., Doulgeris, A. P., Anfinsen, S. N., Eltoft, T., & Serpico, S. B. (2011). A K-Wishart Markov random field model for clustering of polarimetric SAR imagery. In: Proceedings of 2011 IEEE International Geoscience and Remote Sensing Symposium, (9037), 1357–1360. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6049317

Alsdorf, D. E., Rodriguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, *45*, RG2002. https://doi.org/10.1029/2006RG000197

Ardila, J. P., Tolpekin, V. A., Bijker, W., & Stein, A. (2011). Markov-random-field-based super-resolution mapping for identification of urban trees in VHR images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(6), 762–775. https://doi.org/10.1016/j.isprsjprs.2011.08.002

Bioresita, F., Puissant, A., Stumpf, A., & Malet, J. P. (2018). A method for automatic and rapid mapping of water surfaces from Sentinel-1 imagery. *Remote Sensing*, *10*(2). https://doi.org/10.3390/rs10020217

Bolanos, S., Stiff, D., Brisco, B., & Pietroniro, A. (2016). Operational surface water detection and monitoring using Radarsat 2. *Remote Sensing*, *8*(4). https://doi.org/10.3390/rs8040285

Brisco, B., Short, N., Van Der Sanden, J., Landry, R., & Raymond, D. (2009). A semi-automated tool for surface water mapping with RADARSAT-1. *Canadian Journal of Remote Sensing*, *35*(4), 336–344. https://doi.org/10.5589/m09-025

Congalton, G., & Green, K. (2009). Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. The Photogrammetric Record (2nd ed., Vol. *25*). CRC Press.

Doulgeris, A. P., Anfinsen, S. N., & Eltoft, T. (2008). Classification with a non-Gaussian model for PoISAR data. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(10), 2999–3009. https://doi.org/10.1109/TGRS.2008.923025

Doulgeris, A. P., Anfinsen, S. N., & Eltoft, T. (2011). Automated non-gaussian clustering of polarimetric synthetic aperture radar images. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(10 PART 1), 3665–3676. https://doi.org/10.1109/TGRS.2011.2140120

Fjortoft, R., Delignon, Y., Pieczynski, W., Sigelle, M., & Tupin, F. (2003). Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields. *Transactions on Geoscience and Remote Sensing*, *41*(3), 675–686. https://doi.org/10.1109/TGRS.2003.809940

Frery, A. C., Correia, A. H., & Freitas, C. D. C. (2007). Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(10), 3098–3109. https://doi.org/10.1109/TGRS.2007.903828

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Goodman, J. W. (1985). *Statistical Optics*. New York: John Wiley & Sons, Inc. Retrieved from https://cdn.preterhuman.net/texts/science_and_technology/physics/Optics/Statistical Optics - Goodman.pdf

Hahmann, T., Twele, A., Martinis, S., & Buchroithner, M. (2010). In M. Konecny, S. Zlatanova, & T. L. Bandrova (Eds.), *Strategies for the automatic extraction of water bodies from TerraSAR-X/TanDEM-X data BT - Geographic information and cartography for risk and crisis management: Towards better solutions*, (pp. 129–141). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-03442-8_9

Hahmann, T., & Wessel, B. (2010). Surface water body detection in high-resolution TerraSAR-X data using active contour models. Synthetic Aperture Radar (EUSAR), ... , 897–900. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5758876

Henry, J.-B., Chastanet, P., Fellah, K., & Desnos, Y.-L. (2006). Envisat multi-polarized ASAR data for flood mapping. *International Journal of Remote Sensing*, *27*(10), 1921–1929. https://doi.org/10.1080/01431160500486724

Heremans, R., Wiilekens, A., Borghys, D., Verbeeck, B., Valckenborgh, J., Acheroy, M., & Perneel, C. (2003). Automatic detection of flooded areas on ENVISAT/ASAR images using an object-oriented classification technique and an active contour algorithm. *RAST 2003 - Proceedings of International Conference on Recent Advances in Space Technologies*, 311–316. https://doi.org/10.1109/RAST.2003.1303926

Hiremath, S., Tolpekin, V. A., Van Der Heijden, G., & Stein, A. (2013). Segmentation of Rumex obtusifolius using Gaussian Markov random fields. *Machine Vision and Applications*, *24*(4), 845–854. https://doi.org/10.1007/s00138-012-0470-0

Hoque, R., Nakayama, D., Matsuyama, H., & Matsumoto, J. (2011). Flood monitoring, mapping and assessing capabilities using RADARSAT remote sensing, GIS and ground data for Bangladesh. *Natural Hazards*, *57*(2), 525–548. https://doi.org/10.1007/s11069-010-9638-y

Horritt, M. S., Mason, D. C., & Luckman, A. J. (2001). Flood boundary delineation from Synthetic Aperture Radar imagery using a statistical active contour model. *International Journal of Remote Sensing*, *22*(13), 2489–2507. https://doi.org/10.1080/01431160116902

Huang, W., DeVries, B., Huang, C., Jones, J., Lang, M., & Creed, I. (2017). Automated extraction of inland surface water extent from Sentinel-1 data. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 2259–2262). https://doi.org/10.1109/IGARSS.2017.8127439

Hudson, W. D., & Ramm, C. W. (1987). Correct formulation of the kappa-coefficient of agreement. *Photogrammetric Engineering and Remote Sensing*, *53*, 421. Retrieved from http://sfx.metabib.ch/sfx_uzh?sid=google&auinit=WD&aulast=Hudson&atitle=Correct formulation of the kappa-coefficient of agreement&title=Photogrammetric engineering and remote sensing&volume=53&issue=4&date=1987&spage=421&issn=0099-1112

Kenduiywo, B. K., Tolpekin, V. A., & Stein, A. (2014). Detection of built-up area in optical and synthetic aperture radar images using conditional random fields. *Journal of Applied Remote Sensing*, *8*, 83619–83672. Retrieved from https://doi.org/10.1117/1.JRS.8.083672

Kutser, T., Vahtmäe, E., & Praks, J. (2009). A sun glint correction method for hyperspectral imagery containing areas with non-negligible water leaving NIR signal. *Remote Sensing of Environment*, *113*(10), 2267–2274. https://doi.org/10.1016/j.rse.2009.06.016

Lee, J. S., Miller, A. R., & Mango, S. A. (1994). Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *32*(5), 1017–1028. https://doi.org/10.1109/36.312890

Li, J., & Wang, S. (2015). An automatic method for mapping inland surface waterbodies with Radarsat-2 imagery. *International Journal of Remote Sensing*, *36*(5), 1367–1384. https://doi.org/10.1080/01431161.2015.1009653

Li, S. Z. (2009). *Markov random field modeling in image analysis*, (3rd ed.). Incorporated: Springer Publishing Company.

Martinis, S., Twele, A., & Voigt, S. (2009). Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data. *Natural Hazards and Earth System Sciences*, *9*(2), 303–314. https://doi.org/10.5194/nhess-9-303-2009

Mason, D. C., Horritt, M. S., Dall'Amico, J. T., Scott, T. R., & Bates, P. D. (2007). Improving river flood extent delineation from synthetic aperture radar using airborne laser altimetry. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(12), 3932–3943. https://doi.org/10.1109/TGRS.2007.901032

Mather, P., & Tso, B. (2009). *Classification Methods for Remotely Sensed Data*. https://doi.org/10.1201/9781420090741

Mertes, L. A. K. (2002). Remote sensing of riverine landscapes. *Freshwater Biology*, *47*(4), 799–816. https://doi.org/10.1046/j.1365-2427.2002.00909.x

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114

Moser, G., & Serpico, S. B. (2009). Unsupervised change detection from multichannel SAR data by Markovian data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), 2114–2128. https://doi.org/10.1109/TGRS.2009.2012407

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B., & Matthews, E. (2010). Interannual variability of surface water extent at the global scale, 1993-2004. *Journal of Geophysical Research*, *115*, D12111. https://doi.org/10.1029/2009JD012674

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

R Core Team (2018). R: A Language and environment for statistical computin. Retrieved from https://www.r-project.org/

Reigber, A., Jäger, M., Neumann, M., & Ferro-Famil, L. (2010). Classifying polarimetric {SAR} data by combining expectation methods with spatial context. *International Journal of Remote Sensing*, *31*(3), 727–744. https://doi.org/10.1080/01431160902897809

Serpico, S. B., & Moser, G. (2006). Weight parameter optimization by the Ho-Kashyap algorithm in MRF models for supervised image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(12), 3695–3705. https://doi.org/10.1109/TGRS.2006.881118

Silveira, M., & Heleno, S. (2009). Separation between water and land in sar images using region-based level sets. *IEEE Geoscience and Remote Sensing Letters*, *6*(3), 471–475. https://doi.org/10.1109/LGRS.2009.2017283

Small, D., & Schubert, A. (2008). Guide to ASARGeocoding. *Remote Sensing Laboratores, University of Zurich*, (1), 36.

Tholey, N., Clandillon, S., & Fraipont, P. D. E. (1997). The contribution of spaceborne SAR and optical data in monitoring flood events: Examples in northern and southern France. *Hydrological Processes*, *11*(10), 1409–1413. https://doi.org/10.1002/(SICI)1099-1085(199708)11:10<1409::AID-HYP531>3.0.CO;2-V

Tolpekin, V. A., & Stein, A. (2009). Quantification of the effects of land-cover-class spectral separability on the accuracy of markov-random-field-based superresolution mapping. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(9), 3283–3297. https://doi.org/10.1109/TGRS.2009.2019126

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: A fully automated processing chain. *International Journal of Remote Sensing*, *37*(13), 2990–3004. https://doi.org/10.1080/01431161.2016.1192304

Veci, L., Lu, J., Prats-iraola, P., Scheiber, R., & Collard, F. (2015). The Sentinel-1 Toolbox. Retrieved from https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1

Wang, Y., Ruan, R., She, Y., & Yan, M. (2011). Extraction of water information based on RADARSAT SAR and Landsat ETM+. *Procedia Environmental Sciences*, *10*(PART C), 2301–2306. https://doi.org/10.1016/j.proenv.2011.09.359

Westerhoff, R. S., Kleuskens, M. P. H., Winsemius, H. C., Huizinga, H. J., Brakenridge, G. R., & Bishop, C. (2013). Automated global water mapping based on wide-swath orbital synthetic-aperture radar. *Hydrology and Earth System Sciences*, *17*(2), 651–663. https://doi.org/10.5194/hess-17-651-2013

White, L., Brisco, B., Pregitzer, M., Tedford, B., & Boychuk, L. (2014). RADARSAT-2 beam mode selection for surface water and flooded vegetation mapping. *Canadian Journal of Remote Sensing*, *40*(2), 135–151. https://doi.org/10.1080/07038992.2014.943393

Wu, Y., Ji, K., Yu, W., & Su, Y. (2008). Region-based classification of polarimetric SAR images using wishart MRF. *IEEE Geoscience and Remote Sensing Letters*, *5*(4), 668–672. https://doi.org/10.1109/LGRS.2008.2002263

Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., et al. (2016). ISPRS Journal of photogrammetry and remote sensing optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS Journal of Photogrammetry and Remote Sensing*, *122*, 206–221. https://doi.org/10.1016/j.isprsjprs.2016.11.004