Postprint of: Berezowski T., Gierszewska M., Bieliński T., Flood Classification in a Natural Wetland for Early Spring Conditions Using Various Polarimetric SAR Methods, 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 6853-6856, DOI: 10.1109/IGARSS47720.2021.9554713

FLOOD CLASSIFICATION IN A NATURAL WETLAND FOR EARLY SPRING CONDITIONS USING VARIOUS POLARIMETRIC SAR METHODS

Tomasz Berezowski, Monika Gierszewska, Tomasz Bieliński

Faculty of Electronics, Telecommunication and Informatics, Gdansk University of Technology.
Gabriela Narutowicza 11/12, 80233 Gdansk, Poland,
e-mail: tomberez@eti.pg.edu.pl

ABSTRACT

One of the major limitations of remote sensing flood detection is the presence of vegetation. Our study focuses on a flood classification using Radarsat-2 Quad-Pol data in a natural floodplain during leafless, dry vegetation (early spring) state. We conducted a supervised classification of a data set composed of nine polarimetric decompositions and Shannon entropy followed by the predictors' importance estimation to reveal which decomposed component had the strongest effect on classification models. Also, we tested two variants of polarimetric speckle filtering to see if this step influences the results. The classification accuracy was 0.78 and 0.83 for the boxcar and improved Lee sigma filter respectively. The Cloude - Pottier decomposition produced the highest accuracy (0.67) in a single-decomposition scenario, but the volume component of Pauli decomposition was the most important for classification in a multi-decomposition scenario.

Index Terms— SAR, polarimetry, decomposition, speckle filter, vegetation, flooding

1. INTRODUCTION

Synthetic aperture radar (SAR) data is frequently used to map flooding either for crisis management or for monitoring. Currently, many methods, such as thresholding [1, 2] or classification [3] are being used for this purpose. The majority of the already developed methods use the SAR data transformed to backscattering coefficient (σ^0) [dB]. While an important feature of contemporary SAR sensors is the ability to acquire polarimetric data, which, after decomposition, can provide more useful input variables than σ^0.

The polarimetric data is related to the physical or statistical properties of the object. These properties can be derived using the decomposition of the polarimetric matrix. The decomposed components provide information about how different processes or object types contributed to the backscattering process. This can be useful for recognizing complex scattering such as from emergent vegetation over water. However, some shortcomings can be identified that can influence the identification of the minor components. For example, backscatter from the minor component (e.g. from water) can be too low to be depicted by decomposition, or two types of an object (e.g. flat bare soil and water) can produce very similar components.

Nonetheless, several studies used polarimetric decompositions to classify flooding or wetland habitats, e.g. [4, 5, 6, 7]. Yet, thir results are difficult to compare because they use different sensors, different time-series lengths, were conducted in different study areas, and different preprocessing was conducted. Certainly, we lack a consensus about which polarimetric decompositions are the best for flood classification in vegetated areas. Moreover, most of the studies analyze flooding during the full vegetation development stage whereas spring flooding with dry or leafless vegetation is often overlooked in research.

In this study, we aim to rank the decompositions for their utility in flood mapping in various dry and leafless vegetation areas. To achieve this we conduct several polarimetric decompositions for a heterogeneous temperate zone natural wetland and classify them using the random forest supervised classification followed by the importance estimation for the components of the decomposition. Further, we test the effect of using different speckle filtering methods on the classification results.

2. METHODS

2.1. Study area and field sampling

The study is conducted in the lower Biebrza River floodplain located in northeastern Poland (Figure 1). The floodplain area is approximately 210 km^2 and the dominant land-cover is low and high wetland vegetation, such as carr, reed, scrub, or meadow. The flooding occurs each year in early spring and is composed of river water and by other sources, such as groundwater, snowmelt, and rainfall [8]. Effectively, deep water is located mostly in the fluvial terrace ˜up to 1.5 km around the channel, and much shallower water is present in

This research was financed by the National Science Centre, Poland, grant: 2017/26/D/ST10/00665.

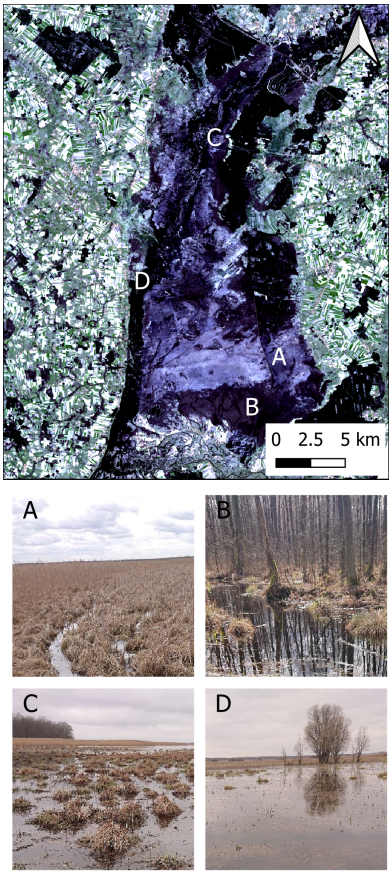

Fig. 1. Study area illustrated by the true-color composition of the Landsat-8 image acquired 2nd of April 2019 (three days after the SAR image). The Biebrza floodplain comprises the center part of the image with grey (A, C) and black colors (B). The deep open water zone (D) corresponds to the Biebrza fluvial terrace (NE-S).

the remote areas of the floodplain. Usually, in the areas of shallow water, vegetation emerging from water or vegetation obscuring the water is present. Field sampling of the flooding extent was conducted on the 27-29th of March 2019 in several transects in the floodplain. In total 170 points were collected and 340 new points were added in post-processing based on analysis of the 1x1 m digital elevation model. During the field sampling, the annual vegetation was mostly in the dry stage and the deciduous trees and shrubs were leafless (Figure 1).

2.2. Remote sensing data and processing

A Radarsat-2 single look complex image in the fine quad polarization mode (FQ3) was acquired for the study area on 30th of March 2019. In the FQ3 mode, the incidence angle is approx. 21° and the pixel size is approx. 14 m in the range direction.

The image was processed using the python snappy package, which is an extension of the ESA SNAP 8.0 remote sensing processing software. The processing was conducted in the following steps: radiometric calibration, polarimetric speckle filtering, polarimetric decomposition, and range-Doppler terrain correction. Two variants of speckle filtering were conducted: boxcar and improved Lee sigma in 9x9 window size. Each filtering product was decomposed using the following decompositions: Sinclair, Pauli, Freeman-Durden, Generalized Freeman-Durden, Yamaguchi, van Zyl, Cloude-Pottier, Cloude, and Touzi and for each filtering product, Shannon entropy was calculated, divided into two contributions: intensity and polarimetry.

2.3. Flood classification and importance estimation

The decomposition products in the two variants of speckle filtering and a terrain slope raster were sampled in the locations of field sampling and labeled as open water, flooded vegetation $<$ 10 cm high, flooded vegetation $>$ 10 cm high, not flooded vegetation, and soil. This data set was split into training and validation subsets (50/50%). The training subset was used as input to the random forest machine learning classification model in three scenarios: (1) each decomposition product was used separately as a model input; (2) all decomposition products with the same speckle filtering method were used as a model input; (3) all decomposition products with both speckle filtering methods were used together as the model input. Each model was further subjected to the predictor importance estimation, which ranks the predictor by the decrease in node impurities from splitting on the predictors expressed as the Gini index. The Gini index quantified the inequality in the data, the higher the value the more the predictor is important for the model. This approach of predictor importance estimation is relative and cannot be compared between the different models. To compare the different model variants we used the accuracy measure, which shows what was the percentage of correctly classified pixels with reference to all pixels in the validation subset.

3. RESULTS AND DISCUSSION

The classification accuracy was 0.40 to 0.67 for the classification of each decomposition separately (scenario 1), with the minimum for the Shannon entropy intensity parameter (boxcar filter) and maximum for the Cloude - Pottier (boxcar filter) and Pauli (improved Lee sigma filter) decompositions (Figure 2). For the classification of all decompositions with the same speckle filtering the accuracy was 0.78 for the boxcar filter and 0.83 for the improved Lee sigma filter (scenario 2). For the classification of all decompositions with both filtering methods (scenario 3), the accuracy was 0.84. The highest accuracy was obtained for the scenario which used all available data, but the accuracy for scenario 2, where data were filtered by using the improved Lee sigma filter, was lower only

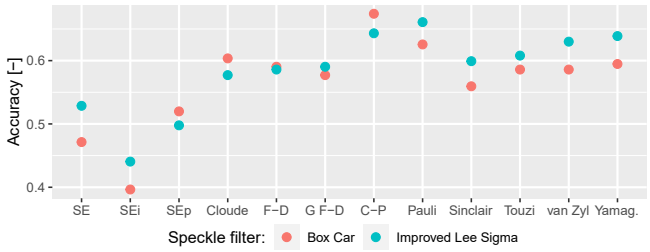

Fig. 2. Classification accuracy for scenario 1. Some decomposition names are abbreviated at the bottom axis: SE is Shannon entropy, SEi is Shannon entropy intensity, SEp is Shannon entropy polarimetry, F-D is Freeman-Durden, G F-D is generalized Freeman-Durden, C-P is Cloude-Pottier, and Yamag. is Yamaguchi. Each point represents the accuracy of different classification models with the same training and validation data.

about 1%. The scenario using only one decomposition product achieved considerably worse accuracy.

The speckle filtering method had an impact on the accuracy results. Scenario 2 shows that overall the model using the improved Lee sigma filter was more accurate than the boxcar filter. In models using only one decomposition, the higher accuracy was also achieved when the improved Lee sigma filter was used, except for the Cloude, Cloude – Pottier, polarimetry Shannon entropy (Figure 2).

The predictors' importance varied between the decompositions and filtering methods, however, it has a similar pattern for all three modeling scenarios. Therefore we focus here on the analysis of scenario 3, which used all decompositions and all filtering methods in one classification model (Figure 3). The predictor importance results in scenario 3 are partially in contradiction to the accuracy results in scenario 1. The Yamaguchi and van Zyl-based classification model had high accuracy (Figure 2) but the components of this decompositions were one of the least important predictors of all in the scenario 3 model (Figure 3). On the other hand, components of the Pauli decompositions, which had high accuracies in scenario 1, also had the highest importance in scenario 3. The comparison of scenario 1-3 accuracy and scenario 3 importance may be interpreted that even if a single decomposition can be used to classify flooded vegetation relatively well there is a synergy between the individual components of different decomposition that can produce a more accurate model.

The classification map for scenario 3 shows continuous flooding for the entire fluvial terrace with some gaps resulting either from not flooded spots (such as local elevation increase) or classification error (Figure 4). We identified that most of the systematic errors were present in the reed belt surrounding the fluvial terrace, which was under-sampled during the field measurements. Also, some of the agricultural fields and smaller river valleys in the upland surrounding the Biebrza floodplain were labeled as open water or flooded vegetation. To improve the accuracy more ground truth data is required for the model.

4. SUMMARY

Each analyzed scenario was able to produce a good model of flood classification in spring (dry, leafless vegetation) with a certain combination of speckle filtering and polarimetric decompositions methods. Our results indicate that the Pauli or Cloude – Pottier decomposition alone was able to produce the best model. Yet, scenario 2, which used multiple decompositions with the same speckle filtering method, showed considerable improvement in the accuracy in reference to the single decomposition approach scenario 1. Moreover, the predictors' importance analysis showed that when multiple decompositions are used in the same model the volume component of the Pauli decomposition is the most important for the classification model. Our results do not show unambiguously which speckle filtering method is better for the classification results, yet, scenario 2 indicates that improved Lee sigma may be in favor of the boxcar. However, to achieve more clear results, both in terms of finding the best polarimetric decomposition and speckle filtering method, more ground truth data is required along with more detailed statistical analysis.

5. REFERENCES

[1] Stephanie Long, Temilola E Fatoyinbo, and Frederick Policelli, "Flood extent mapping for Namibia using change detection and thresholding with SAR," *Environmental Research Letters*, vol. 9, no. 3, pp. 035002, mar 2014.

[2] Tomasz Berezowski, Tomasz Bielinski, and Jakub Osowicki, "Flooding extent mapping for synthetic aperture radar time series using river gauge observations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2626–2638, 2020.

[3] Dora Roque, Nuno Afonso, Ana M Fonseca, and Sandra Heleno, "OBIA flood delimitation assisted by threshold determination with principal component analysis," *Photogrammetric Engineering & Remote Sensing*, vol. 80, no. 6, pp. 551–557, 2014.

[4] R. Touzi, A. Deschamps, and G. Rother, "Phase of target scattering for wetland characterization using polarimetric c-band SAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 9, pp. 3241–3261, 2009.

[5] Luiz Felipe de Almeida Furtado, Thiago Sanna Freire Silva, and Evlyn Márcia Leão de Moraes Novo, "Dual-season and full-polarimetric C band SAR assessment for vegetation mapping in the Amazon várzea wetlands," *Remote Sensing of Environment*, vol. 174, pp. 212–222, 2016.

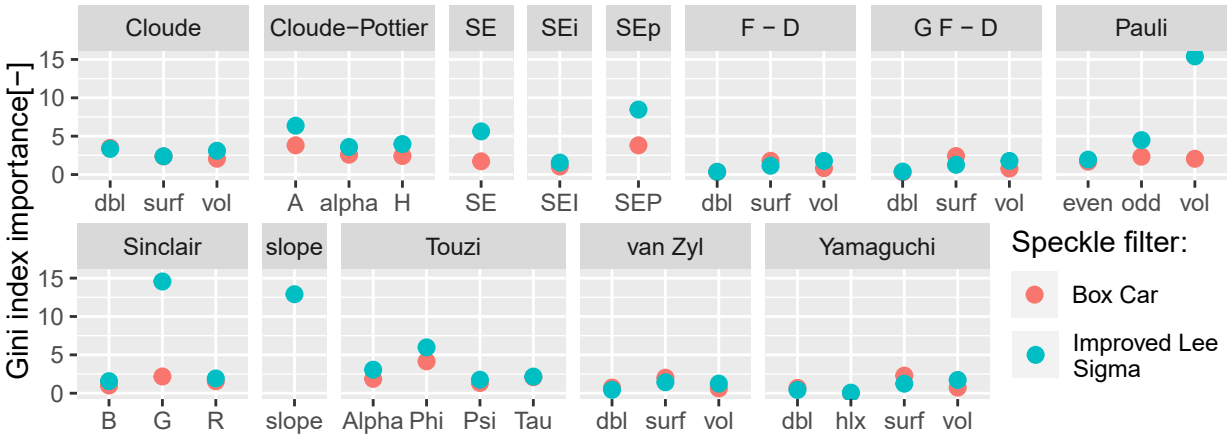

Fig. 3. Predictors importance for the classification model using all decompositions and both speckle filtering methods. Some decomposition names are abbreviated at the bottom axis: SE is Shannon entropy, SEi is Shannon entropy intensity, SEp is Shannon entropy polarimetry, F-D is Freeman-Durden, G F-D is generalized Freeman-Durden.

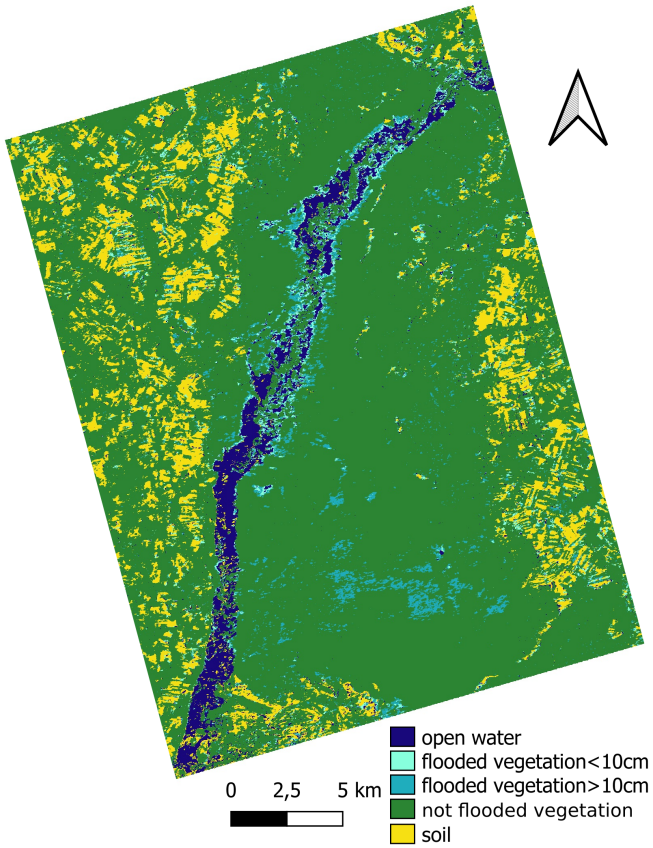

Fig. 4. Classification result for scenario 3, in which each decomposition product and each speckle filtering method were used. The map extent is the same as in Figure1.

[6] Lori White, Brian Brisco, Mohammed Dabboor, Andreas Schmitt, and Andrew Pratt, "A collection of SAR methodologies for monitoring wetlands," *Remote Sensing*, vol. 7, no. 6, pp. 7615–7645, jun 2015.

[7] Simon Plank, Martin Jüssi, Sandro Martinis, and André Twele, "Mapping of flooded vegetation by means of polarimetric Sentinel-1 and ALOS-2/PALSAR-2 imagery," *International Journal of Remote Sensing*, vol. 38, no. 13, pp. 3831–3850, mar 2017.

[8] T. Berezowski, D. Partington, J. Chormański, and O. Batelaan, "Spatiotemporal dynamics of the active perirheic zone in a natural wetland floodplain," *Water Resources Research*, vol. 55, no. 11, pp. 9544–9562, nov 2019.