Received: 5 May 2020 | Revised: 4 November 2020 | Accepted: 6 December 2020

DOI: 10.1111/wej.12681

FULL LENGTH ORIGINAL RESEARCH PAPER

Water and Environment Journal Promoting Sustainable Solutions CIWEM WILEY

Development of a method for flood detection based on Sentinel-1 images and classifier algorithms

Alireza Sharifi

Department of Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran

Correspondence
Department of Surveying Engineering, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran 16785-136, Iran.
Email: a_sharifi@sru.ac.ir

Funding information
Shahid Rajaee Teacher Training University, Grant/Award Number: 19059

Abstract

Floods are one of the most devastating natural disasters in the world, displacing millions of people each year and causing severe damage to people's lives and infrastructure. It is the most widespread hydrological hazard worldwide that affects water management, nature protection, economic activities, hydromorphological alterations on ecosystem services and human health. Real-time monitoring systems play a key role in flood risk reduction and disaster response decisions. Studies have shown that using earth observation data for flood monitoring and timely actions based on good quality information reduces damages. In this paper, after thresholding, a machine learning algorithm and an object-based classification method were used to classify the SAR data. Thresholding helps detect regions in the flooded areas. A comparison of the results showed that the machine learning algorithm obtained significant results. Because of the results obtained, the usefulness of Sentinel-1 images as a baseline data for the improvement of the methodological guide is appreciated and should be used as a new source to monitor the flood risks.

KEYWORDS
classifiers, flood detection, ICA algorithm, Sentinel-1

1 | INTRODUCTION

Floods are one of the most frequent, destructive and widespread natural hazards in the world. Flood monitoring is a useful tool for assessing damage and assisting crisis management organizations. It is a useful tool for assessing damage and assisting crisis management organisations. Dangerous water overflows in residential areas are a common occurrence in Iran after the earthquake and catastrophic floods have occurred annually in Iran (Sharifi et al., 2016). In the last decade, the occurrence of frequent floods in Iran has reached its historical peak and the amount of damage is increasing every year. Therefore, near-real-time flood monitoring is essential for crisis management and damage assessment.

Remote sensing is the best tool for flood mapping and damage assessment which is indicated by many researchers (Jacob et al., 2020; Kosari et al., 2020; Sharifi, 2020a). The capability of synthetic aperture radar (SAR) systems for large-scale flood monitoring has been demonstrated by numerous researchers. The ability to operate during daytime and night time and in all weather conditions is one of the key features that make SAR images important for flood mapping (Sharifi & Hosseingholizadeh, 2020). Also, the high sensitivity of the microwave bands (with a wavelength from 1 to 30 cm) to moisture makes it possible to differentiate SAR data between water and other land covers (Sunar et al., 2019). Operational flood mapping is performed with the aim to reduce the delay between accessing the satellite image and producing the map of the amount produced by it to the civil protection authorities for immediate relief measures. This goal is achieved with a fully automated flood detection and monitoring service. However, when the TerraSAR-X product is downloaded to the FTP server, the service starts and at the end of the mapping process, the amount of access for online visualisation is available through a web interface. It should be noted that the radar pulse from COSMO-SkyMed, which operates in the X band like TerraSAR-X, is reduced due to precipitation and

its relatively short wavelength. In addition, due to the time lag between the TerraSAR-X task assignment in an emergency and the actual image capture, the peak of the eruption may be missed as the satellite needs at least 2.5 days to access the site. In this case, the Sentinel-1 constellation acquisition mode will help a lot (Sharifi, 2020b). The process is generally completely unsupervised, but the recent threshold has yet to be determined experimentally. Sentinel-1 images are automatically loaded and pre-processed using the Sentinel Application Platform (SNAP) and then sorted in the same way. It was found that amongst the polarisations provided by the Sentinel-1 sensor, VV polar SAR images result in more accurate maps than cross-polar (VH) products. In this regard, some researchers have suggested that HH polarisation perceives the highest accuracy in terms of mapping (Twele et al., 2016). In order to assess the flood dynamics and flood duration effectively, it requires several snapshots recorded at various times by the satellite during the floods. The strength of the radar backscatter depends on multiple factors, notably surface roughness, dielectric properties and local topography about the radar look angle (Klemas, 2015). The smooth open water surface acts as a specular reflector of the radar pulse, which scatters the radar energy away from the sensor, resulting in the minimal signal returned to the satellite. As a result, the stagnant water pixels appear dark in radar data which is in contrast with non-water areas.

Several floods have been analysed using SAR images in previous studies. Kussul et al. (2011) used artificial neural networks and Self-Organising Kohonen Maps (SOMs), to SAR image segmentation and classification. The proposed method applied different flood events: Tisza River, Ukraine, and Hungary in 2001; Huaihe River, China in 2007; Mekong River, Thailand, and Laos in 2008; Koshi River, India, and Nepal in 2008; Norman River, Australia in 2009; Lake Liambezi, Namibia in 2009; Mekong River, Laos in 2009 (Kussul et al., 2011). A hybrid methodology, which combines radiometric thresholding and region growing proposed by Matgen et al. (2011) to evaluate the capability of SAR images for flood mapping in two cases: the July 2007 flood of the Severn River (UK) and the February 1997 flood of the Red River (US) (Matgen et al., 2011). During the summer 2007 floods of the River Severn (United Kingdom), the harmonic analysis of a multi-temporal time series of ENVISAT Advanced SAR (ASAR) scenes was used to characterise the seasonality in backscatter under non-flooded conditions (Schlaffer et al., 2015). Also, after the flood event in eastern India, time-series RADARSAT SAR images were used for flood water detection, monitoring of spatial extent and propagation of flood inundation (Rahman & Thakur, 2018). Uddin et al. (2019) aimed at developing an operational methodology for rapid flood inundation and potential flood-damaged area mapping to support a quick and effective event response in Bangladesh. The data and methodology of this study can be replicated every year to map flooding (Uddin et al., 2019). Morover, DeVries et al. (2020) presented an algorithm that exploits all available Sentinel-1 SAR images in combination with historical Landsat and other auxiliary data sources hosted on the Google Earth Engine to rapidly map surface inundation during flood events in Thessaly, Greece and Eastern Madagascar in January and March 2018 (DeVries et al., 2020).

This study will focus on the monitoring of the flood extent by proposing an automated classifier trained on a dataset retrieved from a pre-flood SAR image. The labelled training dataset is thereby extracted from the pre-flood SAR images in an automated fashion. The preprocessing of the dataset and the classification are invoked to map the extent of the inundation present on a post-flood SAR image. This application is intended mainly for emergency situations. As a consequence, it is of extreme importance to extract the flood map as quickly as possible and in an unsupervised way. The current paper is structured as follows. Section 2 introduces a machine learning method for the classification of SAR images for rapid flood monitoring was proposed. The advantages of the proposed method are significant compared to other image-processing techniques used to separate inundated from non-inundated areas in SAR images. The application of a machine learning algorithm is described in detail in Section 3 and finally, Section 5 draws the main conclusions.

2 | MATERIALS AND METHODS

2.1 | Study area

Khuzestan province with an area of 1763 km^2, which is in the southwest of the country and shares a border with Iraq and the Persian Gulf, is one of the main provinces of Iran (Figure 1a). Due to heavy rainfall in the Zagros Mountains, Dez River and Karkheh River overflowed and water accumulated in the Dez Dam and the Karkheh Dam so that Karkhe dam's reservoir was located in Khuzestan province, reached to 8400 m^3/sec. As dams reached their maximum capacity, the power ministry of Iran managed to release water. After the flood, six towns and about 70 villages along the river were affected. Also, rainfall reached 500 mm at some stations reached, which means more than 75% of annual rainfall occurred in 24 h (Sheikh et al., 2019). The depth of the water reached over 50 cm both in and outside of the city, the landslide caused rock falling from the mountain and block some of the roads.

2.2 | Data set

The Sentinel-1 SAR data were obtained from the study area. Acquired images are related to before and after the flood, which can be seen on March 23, 2016 (before the flood event) and on April 16, 2016 (after the flood event) can be seen in Figure 1b. The reduction of speckle noise has been performed by a fast independent component analysis (ICA) algorithm (Sharifi et al., 2015). Although, an accurate geometric calibration has been performed since the Commissioning Phases on operationally generated Level-1 SAR products (Portela et al., 2019).

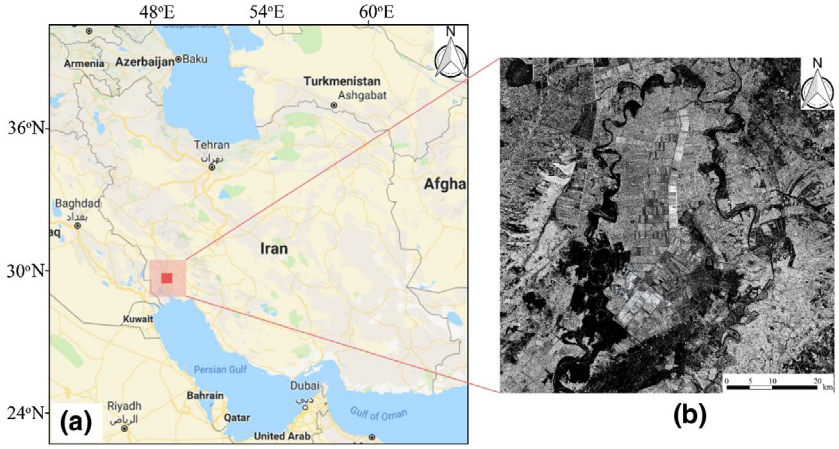

FIGURE 1 (a) Location of the study area. (b) Sentinel-1 image of the study area which acquired on 16 April 2016 after the flood event

2.3 | Methodology

After preprocessing, masks of shadow and layover were extracted automatically from the SAR images with the help of the free SRTM DEM. Pixels will be systematically disregarded during the training and classified as unflooded after that. The shadow and layover pixels detected on the pair of SAR images showing the studied area appear in a homogeneous black colour in the mountains located to the right of the river. The SAR images with masked shadow and layover were eventually converted to dB and projected to the ground-range with a terrain- correction (TC) using the same DEM. The subsetting operation was left until the very end of the preprocessing to get perfectly rectangular images after the terrain-correction (Twele et al., 2016).

The basis of the proposed method is based on thresholding. So, the gamma PDF is used to select two thresholds in the SAR image. Goodman (1975) indicated that similar regions (i.e. flooded area) have the gamma PDF in SAR intensity images (Goodman, 1975). Also, use the gamma PDF for thresholding of SAR images has been proven in some studies (Schlaffer et al., 2015; Wan et al., 2019). The gamma PDF has two parameters including shape (k) and scale (θ) which is defined by Equation (1):

$$f(x|k,\theta) = \frac{x^{(k-1)}e^{(-x/\theta)}}{\theta^k \Gamma(k)} \quad (1)$$

where Γ is the gamma distribution and x is the intensity value.

After the thresholding step, a machine learning classification (MLC) based on the kernel is used to classify remotely sensed images. It uses a Bayesian formulation of a linear model with an appropriate prior which leads to model sparseness (Sharifi, 2020c). Sparsity means few non-zero coefficients define the model which provides an accurate classification using the limited training data (Pal and Foody, 2012). Gaussian prior and Gamma hyper prior are used to define parameters of weight and variance, respectively. It is desired to predict the posterior probability of class membership even the input x. The linear model has been generalised by applying the logistic sigmoid function as Equation (2) and writing the likelihood as Equation (3).

$$\sigma(\varphi(x)) = \frac{1}{1+e^{-\varphi(x)}} \quad (2)$$

$$p((W\phi(x)+\varepsilon)\,|\,W,S) = \prod_{i=1}^{n}\prod_{j=1}^{m}\mathbb{N}((W\phi(x)+\varepsilon)\,|\,W.\phi(x),S) \quad (3)$$

where W is the weight matrix, φ is the basis function, ϕ is the kernel function, S is the score function, ε is the penalty, N is the sparsity, design matrix, and σ is the logistic sigmoid function. Also, m and n are used to product of numbers, and i and j are used for counters.

However, it cannot integrate out the weights to obtain the marginal likelihood analytically and so utilize an iterative procedure, and the most probable weights has been found (the location of the posterior mode). This is equivalent to a standard optimisation of a regularised logistic model, and the efficient iteratively-reweighted least-squares algorithm to find the maximum was used. This procedure is repeated until some suitable convergence criteria are satisfied. Note that in the Bayesian treatment of multilayer neural networks, the Gaussian approximation is considered a weakness of the method if the posterior mode is unrepresentative of the overall probability mass (Sharifi & Hosseingholizadeh, 2020).

In the third step, an object-based classification (OBC) has been implemented, which is based on the image segmentation that divides an image into its constituent regions. Image segmentation is often a time-consuming process that has low efficiency in a near real-time application such as flood monitoring (Wan et al., 2019). The eCognition 9.5.1 was used for the segmentation of the Sentinel-1 image. The main factor in this algorithm is the scale parameter in which the suitable value was calculated with different trials. The smaller values provide finer results, and numerous small objects can be provided, but the generation of large parcels needs larger values. Therefore, the scale parameter was set to 4 after 11 trials. Two other parameters, including shape and compression, were set to 0.1 and 0.8, respectively. As a result, 8312 objects were found in the Sentinel-1 image.

3 | RESULTS AND DISCUSSION

Before performing any operations, pre-processing must be applied to the SAR image. To detect the flooded area in the SAR image, a suitable speckle reduction method is required that the details preserved and the speckle noise is eliminated as much as possible. So, a fast ICA algorithm was implemented which is not only efficient in the speckle reduction and keeping the details, but also it is very fast (Sharifi et al., 2015). The equivalent number of looks (ENL) was measured to assess the performance of the fast ICA method which was calculated using Equation (4) and its value equals 1.70.

$$\mathrm{ENL} = \frac{\mathrm{Mean}(x)}{\mathrm{Variance}(x)} \tag{4}$$

The gamma PDF is used to select thresholds that the lower threshold is the minimum number of the histogram which has no overlap with non-flooded areas. It can be used to extract the flooded areas, while restricting over-detection. Also, the maximum value of pixels that satisfy the gamma PDF ($f(x) < 10^{-4}$) is selected as the higher threshold which can extract the non-flooded areas, while restricting under-detection. To train the classification algorithms and also accuracy assessment, a ground truth map was produced using the digitisation method from the SAR image. The accuracy of the two classification methods was compared to evaluate the performance of the proposed method.

The learning dataset serves to train the supervised classifier to recognise water and land in the post-flood SAR image using machine learning algorithms such as vector-based machine methods. It is not necessary in this case that the training and test datasets are normally distributed with zero mean and unit variance since there is only one feature (the backscatter in dB). Eventually, the trained classifier was employed to segment the post-flood SAR image into water and non-water (land) classes. In summary, the classifier is trained on water and land pixels from the pre-flood SAR image and used to categorise the post-flood SAR image pixels into the same two classes. The classification is thereby based on the assumption that the flood has the same low backscatter signature on the SAR image as permanent water bodies and rivers. As shown in Figure 2, the flooded area are extracted using two classification methods. Also, the results of the accuracy assessment for both machine learning and object-based classification are mentioned in Table 1. As can be seen, the results indicate that both methods worked well and could extract the flooded areas with adequate accuracy. However, in both methods, narrow streams are not well-extracted, but the broad flooded areas are well-classified. The over detected area (ODA) value is higher by the MLC method equals 417, which is due to the application of thresholding only at the pixel level and, as a result, sensitive to the speckle noise. The object-based classification method has a lower ODA value (ODA = 298) because it implements thresholding at the object level which results in less sensitivity to the speckle noise. Most of the over-extracted areas are related to regions that their backscatters are similar to water,

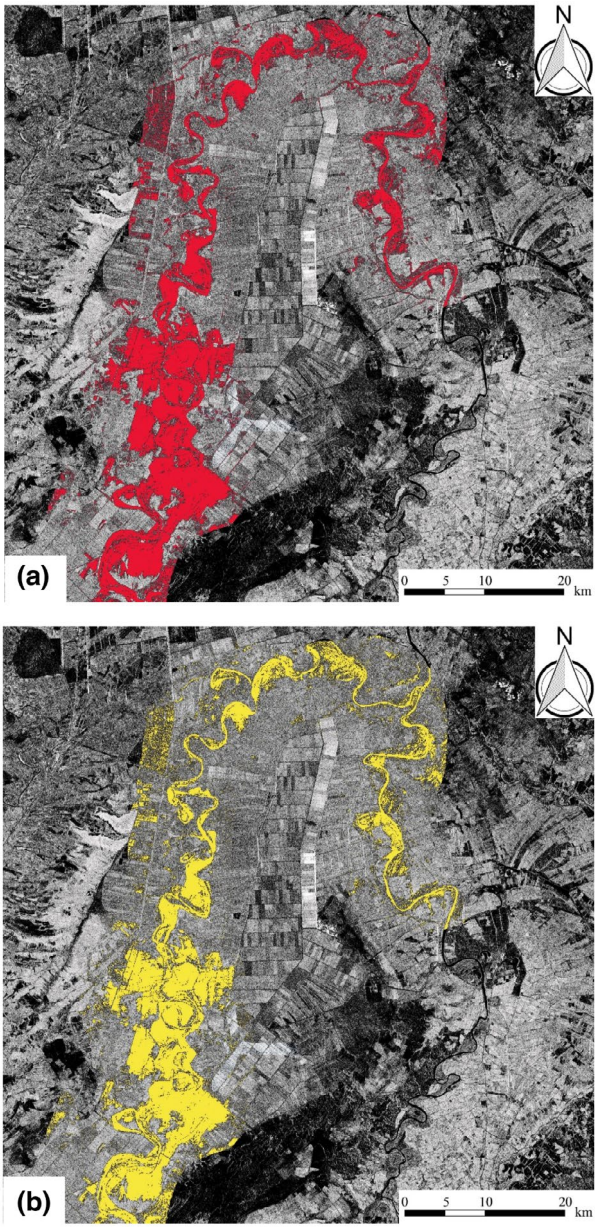

FIGURE 2 The results of classification for the machine learning (a) and the object-based (b) classification methods

TABLE 1 Results of validation for classification methods

Method	MLC	OBC
Overall accuracy	417	298
Kappa	12	44
Over-detection areas	0.77	0.68
Uver-detection areas	0.75	0.63

such as flat areas, roads, and shadow of objects. Therefore, in flood mapping, if a pixel-based algorithm was used in regions with a similar backscatter to water, the results will be over-detected. The second threshold is related to objects with greater backscatters, so both methods have less sensitivity to the speckle noise for

detection. Of course, its value for the MLC method is less than the object-based method (12 < 44). Also, two methods obtain high OA, which indicates these methods could correctly extract the flooded areas and have a high accuracy relative to the reference map (0.68 in comparison to 0.77).

Nevertheless, an MLC method had higher Kappa (k = 0.75) that showed better results. The results showed that our proposed method is more accurate than the object-based classification method, which even has less sensitivity to the speckle noise and it was more useful for flood mapping. The map of flooded areas was examined more thoroughly to study the proposed method more accurately. The MLC method could extract most of the flooded areas with a few under detected area (UDA) and also keep the shape of flooded areas compared to the reference map. In general, the results of the MLC algorithm are entirely satisfactory. Each operation performed during the flood mapping process was profiled individually, by measuring the time it takes to run. The maximum computation time for the preprocessing of the Sentinel-1 images was 8 min and 15 seconds.

4 | CONCLUSION

In this work, a fully automated processing chain has been proposed for inundation mapping and monitoring based on systematically acquired Sentinel-1 data. The processing chain is comprised of modules for preprocessing, an initial classification using automatic thresholding, a machine learning and object-based classification methods, and accuracy assessment. The performance and thematic accuracy of the processor have been assessed for validation sites covering a flood situation in Northern Iran. In comparison to flood mapping approaches, the proposed service has the potential to further enhance the utility of Earth Observation data for emergency management through the provision of decision-critical flood information in near-real-time. A major advantage compared to the Sentinel-1 flood service presented in an earlier study of the authors is the systematic acquisition strategy of the Sentinel-1 mission, which allows the utilisation of the flood service for continuous monitoring purposes without the need for a time-consuming manual tasking of new acquisitions in case of evolving flood situations. In future studies, high-resolution SAR images (i.e. Cosmo Sky-med) can be explored using the proposed approach.

ACKNOWLEDGEMENT

This work was supported by the Shahid Rajaee Teacher Training University under contract number 19059.

ORCID

Alireza Sharifi https://orcid.org/0000-0001-7110-7516

REFERENCES

DeVries, B., Huang, C., Armston, J., Huang, W., Jones, J.W. & Lang, M.W. (2020) Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine. *Remote Sensing of Environment*, 240, 111664. https://doi.org/10.1016/j.rse.2020.111664.

Goodman, J.W. (1975) Statistical properties of laser speckle patterns. *Laser Speckle and Related Phenomena*. 9, 57. https://doi.org/10.1007/bfb0111436.

Jacob, X.K., Bisht, D.S., Chatterjee, C. & Raghuwanshi, N.S. (2020) Hydrodynamic modeling for flood hazard assessment in a data scarce region: a case study of Bharathapuzha river basin. *Environmental Modeling and Assessment*, 25(1), 97–114. https://doi.org/10.1007/s10666-019-09664-y.

Klemas, V. (2015) Remote sensing of floods and flood-prone areas: an overview. *Journal of Coastal Research*, 314, 1005–1013. https://doi.org/10.2112/jcoastres-d-14-00160.1.

Kosari, A., Sharifi, A., Ahmadi, A. & Khoshsima, M. (2020) Remote sensing satellite's attitude control system: rapid performance sizing for passive scan imaging mode. *Aircraft Engineering and Aerospace Technology*, 92(7), 1073–1083. https://doi.org/10.1108/AEAT-02-2020-0030.

Kussul, N., Shelestov, A. & Skakun, S. (2011) Flood Monitoring from SAR Data. In: Kogan, F., Powell, A. & Fedorov, O. (Eds), *Use of satellite and in-situ data to improve sustainability*. Nato science for peace and security series C: environmental security. Dordrecht: Springer, pp. 19–29. https://doi.org/10.1007/978-90-481-9618-0_3.

Matgen, P., Hostache, R., Schumann, G., Pfister, L., Hoffmann, L. & Savenije, H.H.G (2011) Towards an automated SAR-based flood monitoring system: Lessons learned from two case studies. *Physics and Chemistry of the Earth*, 36(7-8), 241–252. https://doi.org/10.1016/j.pce.2010.12.009.

Pal, M. & Foody, G.M. (2012) Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(5), 1344–1355. https://doi.org/10.1109/JSTARS.2012.2215310.

Portela, F.C.S., Macieira, B.P.B., Zanetti, L.V., Gama, V.N., Silva, D.M., Milanez, C.R.D. & Cuzzuol, G.R.F. (2019) How does Cariniana estrellensis respond to different irradiance levels? *Journal of Forestry Research*, 30(1), 31–44. https://doi.org/10.1007/s11676-017-0578-1.

Rahman, M.R. & Thakur, P.K. (2018) Detecting, mapping and analysing of flood water propagation using synthetic aperture radar (SAR) satellite data and GIS: A case study from the Kendrapara District of Orissa State of India. *Egyptian Journal of Remote Sensing and Space Science*, 21, S37–S41. https://doi.org/10.1016/j.ejrs.2017.10.002.

Schlaffer, S., Matgen, P., Hollaus, M. & Wagner, W. (2015) Flood detection from multi-temporal SAR data using harmonic analysis and change detection. *International Journal of Applied Earth Observation and Geoinformation*, 38, 15–24. https://doi.org/10.1016/j.jag.2014.12.001.

Sharifi, A. (2020a) Flood mapping using relevance vector machine and SAR data: a case study from Aqqala, Iran. *Journal of the Indian Society of Remote Sensing*, 48(9), 1289–1296. https://doi.org/10.1007/s12524-020-01155-y.

Sharifi, A. (2020b) Remotely sensed vegetation indices for crop nutrition mapping. *Journal of the Science of Food and Agriculture*, 100(14), 5191–5196. https://doi.org/10.1002/jsfa.10568.

Sharifi, A. (2020c) Yield prediction with machine learning algorithms and satellite images. *Journal of the Science of Food and Agriculture*. https://doi.org/10.1002/jsfa.10696.

Sharifi, A., Amini, J. & Pourshakouri, F. (2016) Development of an allometric model to estimate above-ground biomass of forests using MLPNN algorithm, case study: Hyrcanian forests of Iran. *Caspian Journal of Environmental Sciences*, 14(2), 125–137.

Sharifi, A., Amini, J., Sri Sumantyo, J.T. & Tateishi, R. (2015) Speckle reduction of PolSAR images in forest regions using fast ICA algorithm. *Journal of the Indian Society of Remote Sensing*, 43(2), 339–346. https://doi.org/10.1007/s12524-014-0423-3.

Sharifi, A. & Hosseingholizadeh, M. (2020) Application of sentinel-1 data to estimate height and biomass of rice crop in Astaneh-ye Ashrafiyeh, Iran. *Journal of the Indian Society of Remote Sensing*, 48(1), 11–19. https://doi.org/10.1007/s12524-019-01057-8.

Sheikh, V., Kornejady, A. & Ownegh, M. (2019) 'Application of the coupled TOPSIS–Mahalanobis distance for multi-hazard-based management of the target districts of the Golestan Province, Iran. *Natural Hazards*, 96(3), 1335–1365. https://doi.org/10.1007/s11069-019-03617-0.

Sunar, A.F., Yagmur, N. & Dervisoglu, A. (2019) Flood analysis with remote sensing data – a case study: Maritsa River, Edirne. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. https://doi.org/10.5194/isprs-archives-XLII-3-W8-497-2019.

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.

Uddin, K., Matin, M.A. & Meyer, F.J. (2019) Operational flood mapping using multi-temporal Sentinel-1 SAR images: a case study from Bangladesh. *Remote Sensing*, 11(13), 1581. https://doi.org/10.3390/rs11131581.

Wan, L., Liu, M., Wang, F., Zhang, T. & You, H.j (2019) Automatic extraction of flood inundation areas from SAR images: a case study of Jilin, China during the 2017 flood disaster. *International Journal of Remote Sensing*, 40(13), 5050–5077. https://doi.org/10.1080/01431161.2019.1577999.

How to cite this article: Sharifi A. Development of a method for flood detection based on Sentinel-1 images and classifier algorithms. *Water and Environment Journal*. 2020;00:1–6. https://doi.org/10.1111/wej.12681