



INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus

Géoscience

Sciences de la Planète


Hossein Bonakdari, Amir Hossein Zaji, Keyvan Soltani and Bahram Gharabaghi

Improving the accuracy of a remotely-sensed flood warning system using a multi-objective pre-processing method for signal defects detection and elimination

Volume 352, issue 1 (2020), p. 73-86.

<<https://doi.org/10.5802/crgeos.4>>

© Académie des sciences, Paris and the authors, 2020.
Some rights reserved.

 This article is licensed under the
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



*Les Comptes Rendus. Géoscience — Sciences de la Planète sont membres du
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org*



Hydrology, Environment — Original Article

Improving the accuracy of a remotely-sensed flood warning system using a multi-objective pre-processing method for signal defects detection and elimination

Hossein Bonakdari^{®*,a}, Amir Hossein Zaji^b, Keyvan Soltani^b
and Bahram Gharabaghi^{®c}

^a Department of Soils and Agri-Food Engineering, Université Laval, Québec, G1V 0A6, Canada

^b Department of Civil Engineering, Razi University, Kermanshah, Iran

^c School of Engineering, University of Guelph, Guelph, Ontario, N1G 2W1, Canada.

E-mails: hossein.bonakdari@fsaa.ulaval.ca (H. Bonakdari), amirzaji@gmail.com

(A. H. Zaji), keyvansoltanii@gmail.com (K. Soltani), bgharaba@uoguelph.ca

(B. Gharabaghi).

Abstract. One of the primary goals of watershed management is to proactively monitor and forecast flood water levels to provide early warning for timely evacuation plans and save lives. One of the most economical ways to accomplish this objective is to use remotely-sensed satellite signals. Previous studies have indicated that an Advanced Microwave Scanning Radiometer (AMSR) sensor can be used for river water level monitoring combined with a few in-situ hydrometric gauges for the ground-truth data collection. However, space-based signals are influenced by many error-inducing natural factors, such as dust and cloud cover. Hence, a hybrid method is proposed, which comprises of a multi-objective particle swarm optimization model, a decision tree classification algorithm, the Hotelling's T^2 outlier detection, and a regression model to identify and replace inaccurate space-based signals. This complex hybrid method will be referred to, in this study, with the acronym (OCOR). In the first phase of this hybrid method, the outlier signals are detected and eliminated from the dataset, and in the second phase, the eliminated signals along with signals lost due to satellite technical problems are estimated by ground-truth data calibration using in situ hydrometric stations. The two case studies of the White and Willamette Rivers demonstrate the performance of OCOR in practical situations.

Keywords. River discharge forecasting, Particle swarm optimization, Decision tree classification, Hotelling's T^2 outlier detection, Satellite, Passive microwave signals.

Manuscript received 21st July 2018, revised 2nd April 2019 and 22nd February 2020, accepted 3rd March 2020.

* Corresponding author.

1. Introduction

Despite the fact that climate change has increased the severity and frequency of floods and that floods have killed more people than all other natural disasters combined [De Groeve and Riva, 2009a], more studies are required to provide cost-effective early warning systems for timely evacuation plans in operational context [Zaji *et al.*, 2018]. Recent studies have successfully employed artificial intelligence models for solving nonlinear water resources problems [Akhbari *et al.*, 2017, Bonakdari *et al.*, 2020, Ebtehaj *et al.*, 2018, 2019, Espinoza-Villar *et al.*, 2018, Gholami *et al.*, 2019, Sattar *et al.*, 2019, Zaji and Bonakdari, 2019].

He *et al.* [2014] compared the performance of three of the most popular data-driven techniques, i.e., Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Support Vector Machines (SVM) in predicting river flow. The authors concluded that the SVM method performed better in simulating rivers in a semi-arid mountain region in China. Chen *et al.* [2015] applied a Hybrid Neural Network (HNN) based on fuzzy pattern recognition to simulate the Altamaha River. The authors compared three different optimization algorithms to determine the most appropriate HNN parameters and concluded that Differential Evolution (DE) and Ant Colony Optimization (ACO) outperformed the Artificial Bee Colony (ABC) algorithm.

Darras *et al.* [2017] employed feedforward and recurrent neural networks to forecast flash floods. The authors determined that the recurrent neural network performed better with longer lead time in flash flood prediction and the feedforward neural network performed better in simulating flash floods in short-range forecasting. Yaseen *et al.* [2017] combined the firefly algorithm with an adaptive neuro-fuzzy computing technique to propose a novel model for estimating of the stream flow. Young *et al.* [2017] introduced a hybrid model formed of a physically-based hydrologic modeling system and SVM and stated that this model showed good performance in simulating rainfall-runoff during typhoon events.

Alizadeh *et al.* [2017] combined the wavelet transform with ANN in order to predict rainfall-runoff in the Tolt River. The results showed that this model made appropriate forecasts for one and two months

ahead. Zeynoddin *et al.* [2018] proposed a novel hybrid model to forecast rainfall of a basin with a tropical climate in Malaysia. One of the primary goals of using satellite information is to monitor and forecast the floods in ungauged basins.

The problem with river flood forecasting is that all data-driven models only use in situ information. Hence, because of the direct relation of flood data to the in-situ gauges, regions with no measuring stations cannot provide any information about the river. Nonetheless, the majority of rivers suffer from a lack of in situ stations.

In addition, the number of existing in situ stations in some countries are limited [Calmant and Seyler, 2006, Khan *et al.*, 2012, Shiklomanov *et al.*, 2002, Sivalalan, 2003, Stokstad, 1999]. Hence, studying and forecasting floods is not possible on considerable regions of the earth. Therefore, Predicting in Ungauged Basins (PUB) has recently emerged as a topic of very high interest in flood forecasting [Salvia *et al.*, 2011]. Researchers in this field mainly use satellite information such as space-based precipitation measurements to simulate hydrological circumstances [Brakenridge *et al.*, 2007, Jiang *et al.*, 2014, Khan *et al.*, 2011, Su *et al.*, 2008, Temimi *et al.*, 2011, 2007].

An accurate means of simulating river flow applying satellite information is to utilize passive microwave sensors from the AMSR for Earth Observing System (AMSR-E) [Brakenridge *et al.*, 2007]. Studies in this field began at the Dartmouth Flood Observatory (DFO), followed by the Joint Research Centre (JRC). AMSR-E sensors observe all of the Earth's surface and detect wet river areas by calculating the differences in land and water brightness temperatures. Raw data is justified by the Global Flood Detection System (GFDS), which is accessible at <http://www.gdacs.org>.

To predict the rivers water level time series using satellite data, Zaji *et al.* [2019a] introduced a new approach of minimizing both horizontal and vertical errors. After that, Zaji *et al.* [2019b] developed a new Evolutionary Polynomial Regression-Time series predictor (EPR-T) and used it to predict the rivers discharge using satellite information. Bonakdari *et al.* [2019] developed two extensions of the Markov Chain (MC) method, namely Online-Markov Chain (O-MC) and Extreme Online-Markov Chain (EO-MC) to enhance the forecasting performance of rivers discharge using satellite information.

The signals received by satellites are affected by a range of weather conditions [Kugler and De Groeve, 2007]. For instance, severe weather alters the emission signals of both land and water. The impact of severe weather is more critical when the water-surrounded area is smaller, which is why space-based signal accuracy is poorer when river discharge is low [Khan et al., 2012]. So, in order to improve the performance and reliability of satellite data, Zaji et al. [2019a] proposed an approach based on trial and error, classification, and outlier detection methods. The authors concluded that the developed method could significantly increase the accuracy of satellite data.

The primary goal of the present study is to introduce a hybrid method that comprises of a multi-objective particle swarm Optimization model, decision tree Classification algorithm, Hotelling's T^2 Outlier detection, and a Regression model (OCOR) for evaluating raw satellite signals to obtain accurate continuous signals. With this method, outlier signals received through severe weather or other unknown environmental situations are initially detected and eliminated. Then the removed signals and those missed due to technical satellite problems are predicted using a regression method.

Previous studies have mainly focused on simulating river discharge using space-based signals [Frasson et al., 2017]. However, when the goal is to use historical information of satellite sensors to forecast the future river discharge, a gap-free continuous space-based dataset is needed [Weigend, 2018]. Thus, the second step entails evaluating a continuous dataset of space-based signals with no missed samples. In the training phase of the OCOR method, the in-situ measurements are needed in order to determine one of the objective functions of the multi-objective optimization algorithm and to develop the regression method to replace the detected noisy samples.

2. Space-based signals

AMSR-E and AMSR2 passive microwave sensors are employed in the present study to gather space-based signals [Brakenridge et al., 2007]. After that, the obtained signals are used to measure the river discharge indirectly. In order to obtain more stable observations, the descending orbit method was selected in this study [Kugler and De Groeve, 2007, Tekeli and

Fouli, 2017], and the Earth's surface was measured at least once a day.

Microwave measurements, which are very sensitive to water, are widely used in the soil moisture estimation field of science [Njoku et al., 2003, Schmugge, 1980, Srivastava, 2017, Theis et al., 1982, Ulaby et al., 1978, Wang et al., 1982]. By using the microwave signals, river discharge is estimated according to the differences between the thermal emission of dry and wet surfaces. The brightness temperature of the considered wet area is called Measurement (M), which is a proxy of river water and the received information is calibrated applying the surrounding dry area's brightness temperature (C) to obtain the final AMSR space-based Signal (S) as follows:

$$S = \frac{C}{M}. \quad (1)$$

Land brightness temperature is higher than wet area temperature [Kugler and De Groeve, 2007], such that S is low when river discharge is low and vice versa [Birkinshaw et al., 2010].

According to Bjerklie et al. [2004] and Brakenridge et al. [2012], in most rivers, the discharge has a better correlation with flow width than with the flow velocity. So that, flow width, which is calculated by using the satellite information, could be considered as a robust proxy of river discharge.

3. Study area

In this work, two rivers are examined as case studies, namely the White River and the Willamette River. The properties of these two rivers are presented in Table 1. Both are located in the United States. The White and Willamette River sources are the Boston Mountains and the junction of the Middle Fork Willamette River with the Coast Fork Willamette River, respectively. Figure 1 presents a schematic overview of the in-situ stations and space-based measurement locations for the White and Willamette Rivers. In this figure, area M , area C , and the in-situ measurement station are denoted by blue, gray, and black, respectively.

The in-situ station information was collected from the United States Geological Survey (USGS) (<http://www.usgs.gov>), and the space-based information was obtained from GFDS (<http://www.gdacs.org>). The time periods examined for the White and Willamette Rivers are from 2009 to 2011 and from 2002 to 2016, respectively.

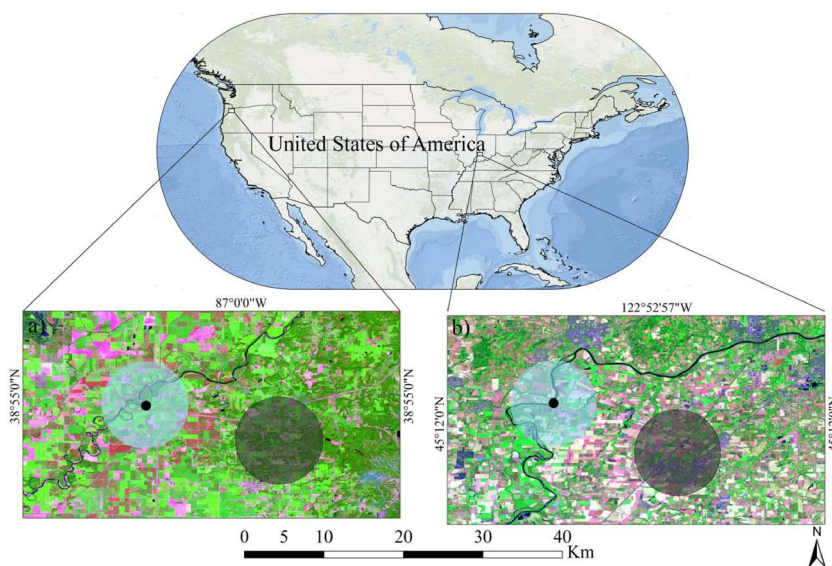


Figure 1. Schematic overview of the (a) White River and (b) Willamette River locations (Landsat satellite images in false color Composite (7,5,2)).

Table 1. White and Willamette River properties

River name	Site ID	Station ID	Space based site coordinates	In situ station coordinates	Mean discharge (m ³ /s)
			Lat./Long. (dd)	Lat./Long. (dd)	
White, Newberry	524	03360500	38.91/−87.07	38.92/−87.011	12137
Willamette	925	14191000	45.18/−123.01	44.94/−123.042	18928

4. Numerical models

The primary goal of the method proposed in current study is to re-evaluate the in situ and space-based datasets by detecting and replacing their errors. As mentioned in the literature, many studies have addressed calibrating space-based information within in situ measurements. However, the poor performance of satellite sensors, in some cases, can lead to inaccurate signals. In addition, satellite or in situ gauges are not always accessible. However, forecasting river discharge using space-based data necessitates a dataset with no gaps, whereby the technique should replace eliminated and missed data with appropriate values.

The OCOR method is introduced in this study to more efficiently (compred to exisiting methods) achieve these goals. The OCOR method involves an outlier detection approach combined with a classification algorithm to detect outliers and inaccurate

signals. The OCOR subsequently employs a regression model to evaluate the eliminated and missed information. In addition, the model is justified employing a multi-objective optimization algorithm.

Outlier detection is the science of detecting the abnormal samples by considering the amounts of other samples of the dataset [Mitra, 2009]. An outlier detection model tries to identify the Global and Contextual outliers of space-based signals. Global outliers occur when a sample deviates significantly from the whole dataset, and Contextual outliers occur when a sample deviates significantly from nearby samples in that season [Han et al., 2011].

Hotelling's T^2 outlier detection method [Hotelling, 1992] is used in the present study. Hotelling's T^2 has been applied as a quality checking algorithm in many studies [Mason and Young, 2002, Shabbak et al., 2011, Sullivan and Woodall, 1996]. In Hotelling's T^2 , the entire dataset is transferred to a

hypersphere with radius α by using rotation and normalization. Following the mentioned transformations, the samples that satisfy the condition $T^2 \leq \alpha^2$ are deemed normal, and the remaining samples are considered abnormal. Hence, α is a strict criterion of this method.

In the present study, supervised outlier detection is utilized, which, together with Hotelling's T^2 outlier detection, necessitates a classification algorithm [Han et al., 2011]. In supervised outlier detection, the classification algorithm labels the samples as experts, after samples which samples that enlarge the classes are detected as outliers and eliminated from the dataset.

The OCOR method employs the decision tree algorithm, as it is one of the simplest and fastest classification algorithms [Quinlan, 1986, 1987, Quinlan and Rivest, 1989]. This algorithm has a tree-like shape and comprises three parts. The first part entails internal nodes that make decisions on the attributes; the second part contains branches that represent the decisions made in the internal nodes; and the third part, or the final level of tree nodes, includes leaf nodes that present the class labels. The decision tree employed was introduced by Breiman et al. [1984] as the CART algorithm for measuring training dataset impurity as follows:

$$GI(\text{dataset}) = 1 - \sum_{i=1}^m p_i^2 \quad \text{where } p_i = \frac{|C_{i,\text{dataset}}|}{|\text{dataset}|} \quad (2)$$

where C_i represents the class label p_i is the probability that the sample considered is in class C_i , and m is the number of classes considered. GI leads to a binary decision in each internal node. The precision of the decision tree classification algorithm is justified using the Minimum Parent Size (MPS) parameter. MPS determines the minimum number of samples allowed in each node and serves as a termination criterion. Lower MPS leads to a larger, more accurate decision tree and vice versa.

Parameters α (strict Hotelling's T^2 criterion), m (number of classes), and MPS (decision tree termination criterion) should be determined to attain an optimum model. The optimum position is a model with the lowest number of outlier samples and the highest coefficient of determination (R^2) between the satellite signals and in situ discharge measurements. Hence, two goals must be attained, and a multi-objective optimization algorithm is needed.

In the present study, Multi-Objective Particle Swarm Optimization (MOPSO) [Coello and Lechuga, 2002] is used. MOPSO is a combination of Particle Swarm Optimization (PSO) [Kennedy et al., 2001, Shaghaghi et al., 2017, Zaji et al., 2015] and the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) [Deb et al., 2000]. The goal of MOPSO is to determine the most suitable α , m , and MPS values to attain a model that detects the lowest number of outliers with the highest R^2 .

The final step of the OCOR method is the regression procedure. Here, the satellite-based samples that were eliminated in previous steps are replaced with accurate values. This procedure involves in situ discharge measurements as input samples to the regression model and satellite-based signals as the output samples. Following the elimination, the regression model is trained using the remaining samples, after which it models the eliminated samples.

The OCOR method tries to detect and eliminate as few inaccurate and noisy samples as possible by using a combination of outlier detection, classification, and multi-objective optimization algorithms and after that, replace the eliminated samples using a regression method.

Applying this model before using satellite rivers information is necessary for any further application of them. However, this version of the proposed method has some limitations. Firstly, in this method, the in-situ discharge measurements are used in order to find the highest R^2 , which is one of the objectives of the optimization algorithm and also for replacing the eliminated samples, which is the last phase of OCOR method.

Therefore, this method is not applicable to ungauged basins. Secondly, this model has only been tested on the AMSR satellite sensors' information. The footprint size of AMSR is really big (about $8 \times 12 \text{ km}^2$). The information obtained by this sensor are more suitable for large rivers. Thirdly, in order to use the proposed preprocessing method on other satellite sensors, some justifications should be done to this model. For instance, despite having a high resolution of MODerate resolution Imaging Spectroradiometer (MODIS), the gathered information is more sensitive to the cloudy weather condition, and those images should be eliminated from the dataset separately.

5. Results and discussion

As mentioned earlier, one of the optimal alternatives to river discharge measurement is satellite information. For the satellite signals to be usable, they must undergo a preprocessing phase. The OCOR method entails two phases: preprocessing and calibration. Satellite signals have two major drawbacks. The first is inaccuracy due to poor environmental conditions, and the second is missed signals on account of inaccessibility to the satellite.

Figure 2 presents the in-situ discharge measurements and satellite signals for the two rivers considered. This figure indicates a good correlation between the satellite signals and in situ discharge measurements. However, it is evident that many time periods have no information and there are missed samples. For the White River, besides missed satellite signals, there is a period of time for which in situ measurements were lost.

According to the study area section, in the present study, three years of daily information was used for the White River, and 15 years of daily information was used for the Willamette River. It should be noted that selecting the most appropriated time period for each case study should be done according to the complexity of the considered problem. For instance, according to Figure 2, for the Willamette River, the satellite signals were missed for a long period of time (the black areas). In this case, longer periods of in-situ measurements and satellite signals are needed to evaluate a reliable model.

On the other hand, selecting a longer period of time leads to having more computational time as well, in this case, by using a MacBook Pro with 3.1 GHz Intel Core i7 CPU and 16 GB Memory, the running time of Willamette River model takes about 10 h. However, with the same computer, the running time of the White River model takes about 2 h.

One of the most critical objectives of studies on river discharge is forecasting. Forecasting river discharge via any method requires a continuous dataset with no missed samples. The goal of the preprocessing phase in the OCOR method is to identify inaccurate satellite signals and eliminate them from the dataset. The purpose of the calibration phase in the OCOR method is to re-evaluate the excluded outliers and missed signals in order to obtain a continuous and accurate dataset for use as time

series in river discharge forecasting using satellite information.

An optimum outlier detection model will attain maximum R^2 with the least number of eliminated outliers. Hence, there are two aims, and this problem cannot be solved using simple optimization algorithms. In the present study, MOPSO was utilized to adjust three parameters, i.e., MPS , α , and m , to attain the best model. In the present model, MPS varied from 1 to 500, while α and m varied from 1 to 10. Each case study underwent 100 iterations and the population number considered was 300. With multi-objective optimization methods, the best model is selected by using the concept of domination. According to the definition, $j = [Goal_{1j}, \dots, Goal_{nj}]$ dominates $i = [Goal_{1i}, \dots, Goal_{ni}]$ if all parameters of j are less than or equal to the parameters of i and none of the parameters of j are less than those of i .

In the MOPSO procedure, the non-dominated samples are saved in a repository. The present model repository can save 100 non-dominated samples. Upon reaching this number of samples, the best model is selected manually. Figure 3 illustrates the repository samples for the White and Willamette Rivers as red circles. In this figure, the best models are denoted by blue diamonds. These models evidently represent a good balance between high R^2 and a low number of eliminated samples.

The most suitable model of the White River had MPS of 3, α of 1.49, and m of 9, while the best model of the Willamette River had MPS of 4, α of 2.18, and m of 10. All samples for these models and the eliminated samples are shown in Figure 4, where the classes are shown in different colors. Moreover, Figure 4 indicates that these datasets were not classified accurately. However, the goal of the present study was not accurate classification, but rather to determine the best outlier samples that increase the R^2 of the satellite signals and in situ discharge measurements.

Thus far, a combination of multi-objective optimization, classification, and outlier selection methods were applied to detect and eliminate inaccurate signals from the dataset. The dataset additionally had gaps due to technical satellite problems. In order to fill these gaps, the second OCOR phase done was calibration. In this phase, the missed and outlier signals were evaluated using in situ discharge measurements, which served as nonlinear regression model inputs while the satellite signals were the targets.

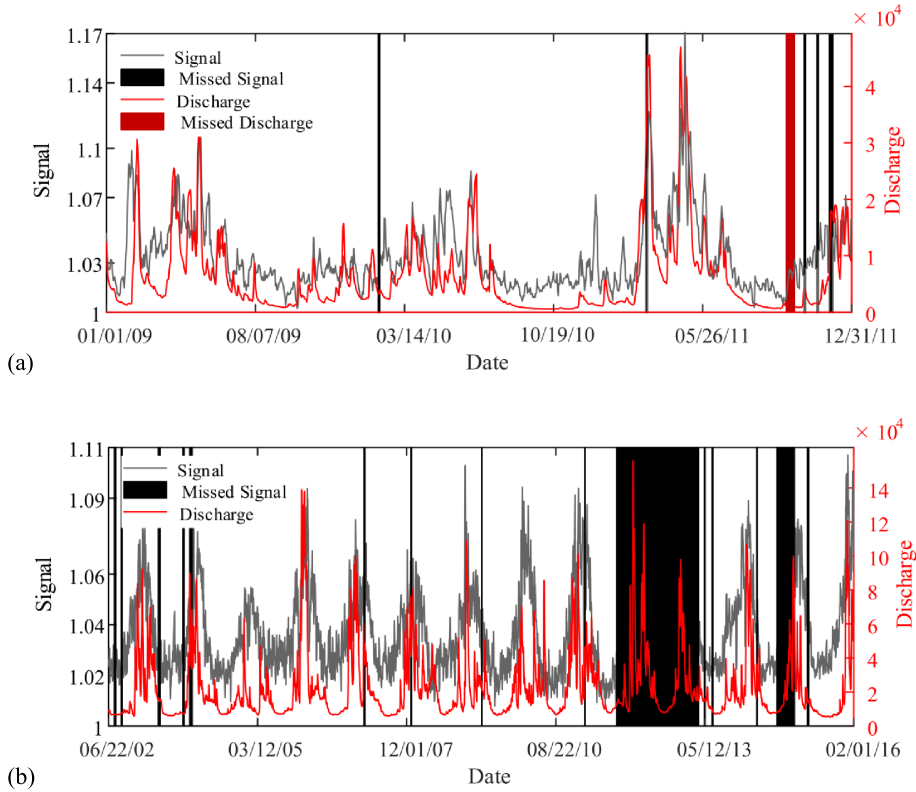


Figure 2. Satellite signals and in situ discharge measurements for the (a) White River and (b) Willamette River.

Following regression model training, the missed and eliminated signals were modeled. The calibration regression equations for the White and Willamette Rivers are given by (3) and (4), respectively, where S represents satellite signal values, and Q denotes in situ discharge measurements.

$$S = 4.63e^{-16} \times Q^3 - 6.16e^{-11} \times Q^2 + 3.79e^{-6} \times Q + 1.014 \quad (3)$$

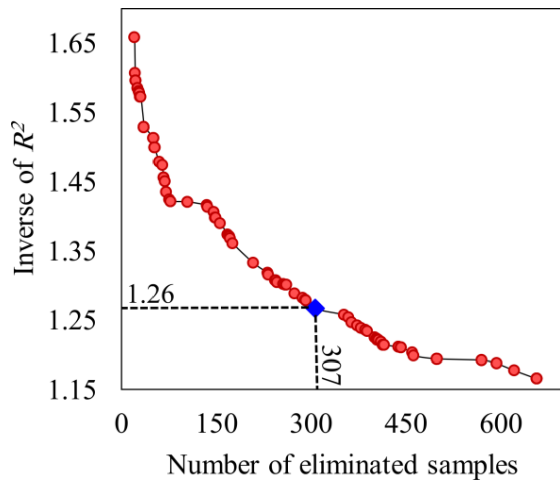
$$S = 6.52e^{-18} \times Q^3 - 6.86e^{-12} \times Q^2 + 1.15e^{-6} \times Q + 1.017 \quad (4)$$

Values missed because of technical satellite problems are not specified. In situ gauges are also sometimes inaccessible, and the in-situ discharge measurement dataset has lost samples too. In such cases, a regression model can be employed to model in situ discharge measurements using satellite signals.

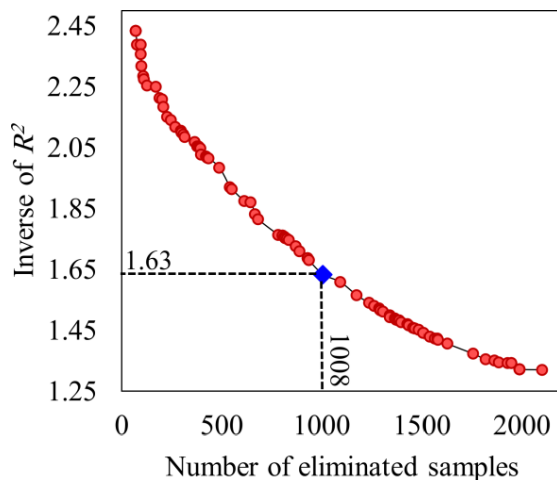
The signals evaluated for the two case studies are shown in Figure 5. Here, the red points are outlier signals eliminated from the dataset in the OCOR prepro-

cessing phase and replaced in the calibration phase using (3) and (4). The blue points represent satellite signals missed because of technical problems. From this figure, it is evident that the Willamette River has lost signals for a long period of time.

These lost samples were re-evaluated by using the mentioned regression equations. The pink points in the figure for the White River dataset represent sample inputs replacing in situ measurement values missed due to technical gauge problems. For the White and Willamette Rivers, the R^2 amounts between the initial satellite signals and in-situ discharges are 0.57 and 0.39, respectively, and the R^2 amounts between the developed satellite signals and in-situ discharges are 0.86 and 0.7, respectively. So, using the OCOR method increases the performance of the satellite signals significantly. In addition, in the initial dataset, there are some gaps in the satellite signals and in-situ measurements, which are replaced in the final datasets.



(a)

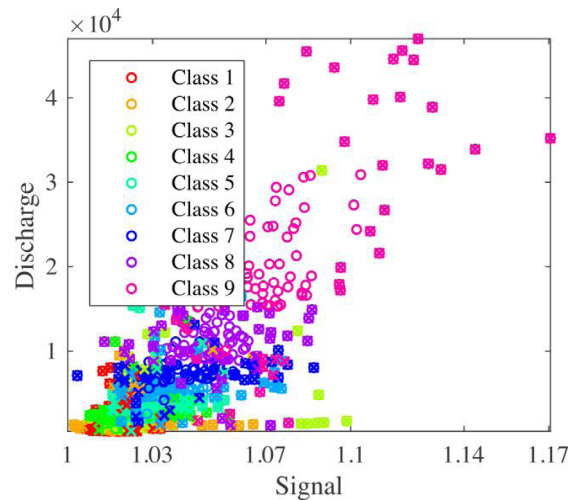


(b)

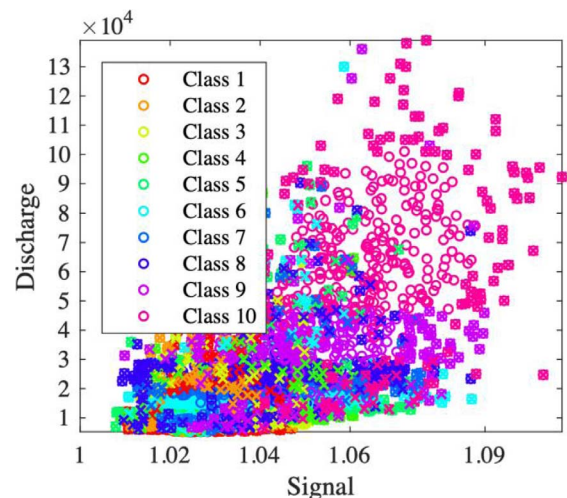
Figure 3. Repository samples for the (a) White River and (b) Willamette River.

After calibration, the OCOR method has attained the objectives. First, inaccurate satellite signals were detected and eliminated. Second, the missed and excluded signals were re-evaluated using regression models. According to Figure 6, subsequent to these two phases, the satellite signals exhibited an excellent correlation with the in-situ discharge measurements; hence, this dataset can be used as a time series to forecast future river discharge.

Finally, the OCOR method is used to distinguish between the normal water level of rivers and the flood water levels. In this theory, which was used



(a)



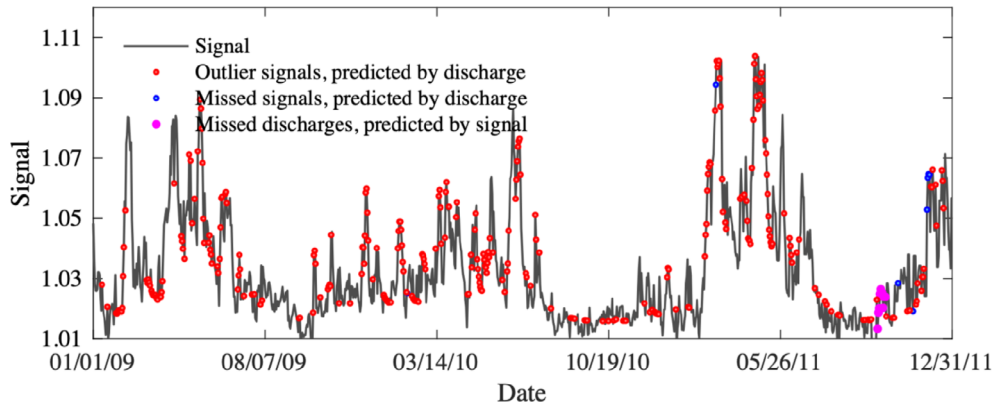
(b)

Figure 4. Classification and outlier samples of the (a) White River and (b) Willamette River.

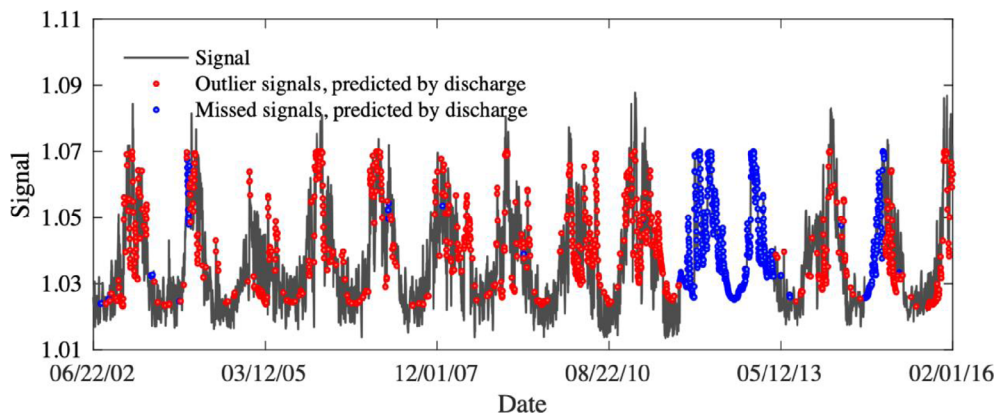
in the field of satellite-based flood detection by De Groeve and Riva [2009b], the Flood Magnitude (FM) is defined using the following equation:

$$FM = \frac{S - \bar{S}}{\sigma(S)} \quad (5)$$

where S is the satellite signals as presented in (1), and $\sigma(S)$ is the standard deviation of the S dataset, which



(a)



(b)

Figure 5. The final dataset of satellite-based signals for the (a) White River and (b) Willamette River.

is defined as follow:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (S_i - \bar{S})^2} \quad (6)$$

where N is the number of S dataset's samples and \bar{S} represents the average value of S . A normal distribution, when the average value of the dataset is zero, and its standard deviation is one the distribution is defined as follows:

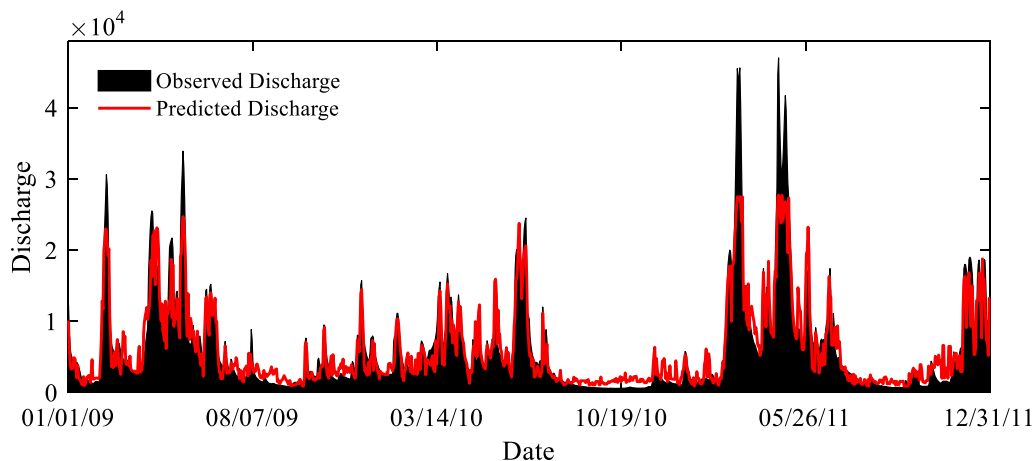
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (7)$$

where x represents a sample of the S dataset. The histograms of the FM datasets for the Willamette and

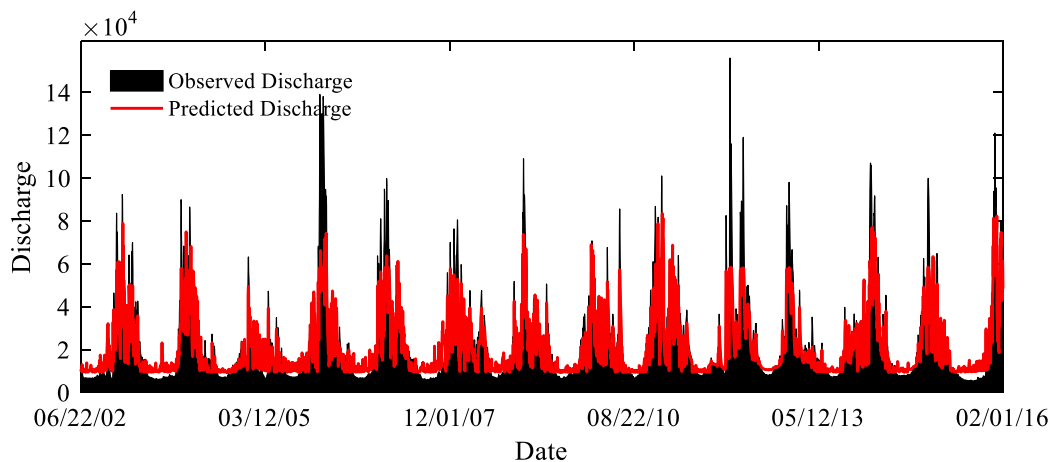
White Rivers are presented in Figure 7. In this figure, the normal distribution was obtained using (7) is shown by a red line and the real distributions of the FM datasets are shown by blue bins. De Groeve and Riva [2009b] and the GFDS organization proposed that considering FM samples with a standard deviation between two and four ($2 < \sigma < 4$) as small and regular floods and FM samples with a standard deviation higher than 4 ($\sigma \geq 4$) as extreme floods.

6. Conclusion

One of the most worthwhile aims of flood disaster management is to provide real-time flood water level forecasting for early warning evacuation



(a)



(b)

Figure 6. Simulations of in situ discharge from the (a) White River and (b) Willamette River using satellite signals.

plans to save lives. Various studies have been recently conducted in this field of science. Satellite-based signals are becoming more readily available and cost-effective method to monitor a vast network of rivers during major storm events. In this study, a novel method is proposed to improve the accuracy of satellite-based water level forecasts, comprising a multi-objective particle swarm Optimization model, decision tree Classification algorithm, Hotelling's T^2 Outlier detection, and a Regression model (OCOR).

The new OCOR method can effectively identify inaccurate satellite signals and replace them with reasonable values.

Two case studies were selected to showcase the novel application of the new OCOR method, including the White River and the Willamette River, located in the United States. In total, 307 White River samples and 1008 Willamette River samples were identified as outliers or missing data due to satellite technical problems. Upon eliminating these samples, the

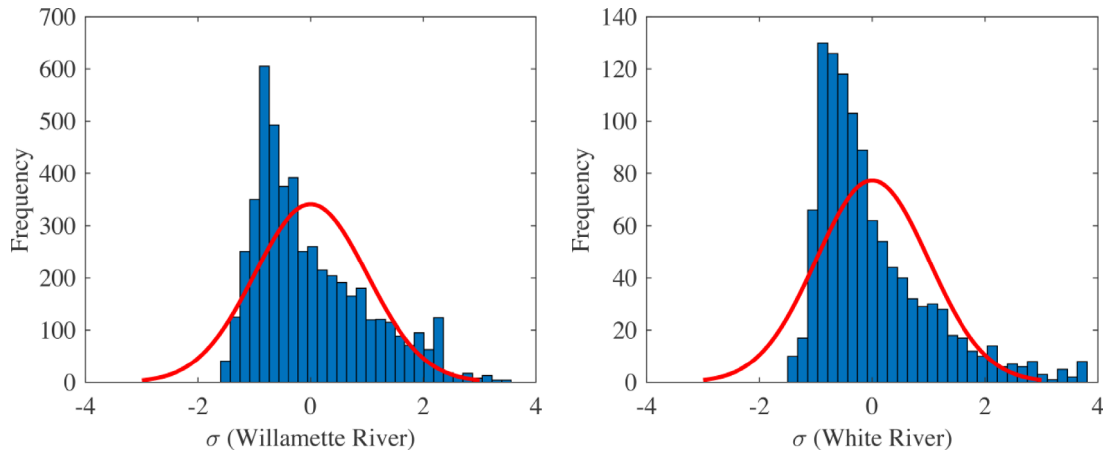


Figure 7. Histograms of the Willamette and White Rivers' FM datasets.

satellite-based signals were successfully calibrated with the in situ hydrometric gauge data, and the deleted signals were gap-filled using the calibrated regression model. Hence, all dataset gaps were filled with satisfactory values. The final results demonstrate a superior correlation between in situ measurements and satellite-based signals following the OCOR model application and can be used as a cost-effective method for tracking water levels in vast river networks during major flood events for early warning evacuation plans to save lives.

References

- Akhbari, A., Bonakdari, H., and Ebtehaj, I. (2017). Evolutionary prediction of electrocoagulation efficiency and energy consumption probing. *Desalin. Water Treat.*, 64:54–63.
- Alizadeh, M. J., Kavianpour, M. R., Kisi, O., and Nourani, V. (2017). A new approach for simulating and forecasting the rainfall-runoff process within the next two months. *J. Hydrol.*, 548:588–597.
- Birkinshaw, S. J., O'donnell, G., Moore, P., Kilsby, C., Fowler, H., and Berry, P. (2010). Using satellite altimetry data to augment flow estimation techniques on the Mekong River. *Hydrol. Process.*, 24(26):3811–3825.
- Bjerklie, D. M., Moller, D., Smith, L. C., and Dingman, S. L. (2004). Estimating discharge in rivers using remotely sensed hydraulic information. *J. Hydrol.*, 309(1):191–209.
- Bonakdari, H., Moradi, F., Ebtehaj, I., Gharabaghi, B., Sattar, A. A., Azimi, A. H., and Radecki-Pawlik, A. (2020). A non-tuned machine learning technique for abutment scour depth in clear water condition. *Water*, 12:301.
- Bonakdari, H., Zaji, A. H., Binns, A. D., and Gharabaghi, B. (2019). Integrated Markov chains and uncertainty analysis techniques to more accurately forecast floods using satellite signals. *J. Hydrol.*, 572:75–95.
- Brakenridge, G. R., Cohen, S., Kettner, A. J., De Groeve, T., Nghiem, S. V., Syvitski, J. P., and Fekete, B. M. (2012). Calibration of satellite measurements of river discharge using a global hydrology model. *J. Hydrol.*, 475:123–136.
- Brakenridge, G. R., Nghiem, S. V., Anderson, E., and Mic, R. (2007). Orbital microwave measurement of river discharge and ice status. *Water Resour. Res.*, 43(4):W04405.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. CRC Press, Boca Raton, Florida, USA.
- Calmant, S. and Seyler, F. (2006). Continental surface waters from satellite altimetry. *C. R. Geosci.*, 338(14):1113–1122.
- Chen, X. Y., Chau, K. W., and Busari, A. O. (2015). A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model. *Eng. Appl. Artif. Intell.*, 46:258–268.
- Coello, C. A. C. and Lechuga, M. S. (2002). MOPSO: a proposal for multiple objective particle swarm

- optimization. In *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on, Honolulu, USA*, pages 1051–1056. IEEE, New York, USA.
- Darras, T., Kong-A-Siou, L., Vayssade, B., Johannet, A., and Pistre, S. (2017). Karst flash flood forecasting using recurrent and non-recurrent artificial neural network models: the case of the Lez Basin (Southern France). In Renard, P. and Bertrand, C., editors, *EuroKarst 2016, Neuchâtel: Advances in the Hydrogeology of Karst and Carbonate Reservoirs*, pages 169–177. Springer International Publishing, Cham.
- De Groeve, T. and Riva, P. (2009a). Early flood detection and mapping for humanitarian response. In *Proceedings of the 6 International ISCRAM Conference, Gothenburg, Sweden*. Information Systems for Crisis Response and Management (ISCRAM), Gothenburg, Sweden.
- De Groeve, T. and Riva, P. (2009b). Global real-time detection of major floods using passive microwave remote sensing. In *Proceedings of the 33rd International Symposium on Remote Sensing of Environment, Stresa, Italy*, pages 4–8. International Center for Remote Sensing of Environment (ICRSE), Stresa, Italy.
- Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In *International Conference on Parallel Problem Solving From Nature*, pages 849–858. Springer.
- Ebtehaj, I., Bonakdari, H., and Gharabaghi, B. (2019). Closure to “An integrated framework of Extreme Learning Machines for predicting scour at pile groups in clear water condition by Ebtehaj, I., Bonakdari, H., Moradi, F., Gharabaghi, B., Khozani, Z. S.”. *Coast. Eng.*, 147:135–137.
- Ebtehaj, I., Bonakdari, H., Moradi, F., Gharabaghi, B., and Khozani, Z. S. (2018). An integrated framework of Extreme Learning Machines for predicting scour at pile groups in clear water condition. *Coast. Eng.*, 135:1–15.
- Espinoza-Villar, R., Martinez, J.-M., Armijos, E., Espinoza, J.-C., Filizola, N., Dos Santos, A., Willems, B., Fraizy, P., Santini, W., and Vauchel, P. (2018). Spatio-temporal monitoring of suspended sediments in the Solimoes River (2000–2014). *C. R. Geosci.*, 350(1–2):4–12.
- Frasson, R. P. d. M., Wei, R., Durand, M., Minear, J. T., Domeneghetti, A., Schumann, G., Williams, B. A., Rodriguez, E., Picamill, C., Lion, C., Pavelski, T., and Garamois, P. A. (2017). Automated river reach definition strategies: applications for the surface water and ocean topography mission. *Water Resour. Res.*, 53(10):8164–8186.
- Gholami, A., Bonakdari, H., Samui, P., Mohammadian, M., and Gharabaghi, B. (2019). Predicting stable alluvial channel profiles using emotional artificial neural networks. *Appl. Soft Comput. J.*, 78:420–437.
- Han, J., Pei, J., and Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier Inc., Amsterdam, Netherlands.
- He, Z., Wen, X., Liu, H., and Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J. Hydrol.*, 509:379–386.
- Hotelling, H. (1992). The generalization of Student's ratio. In *Breakthroughs in Statistics*. Springer, New York, USA.
- Jiang, S., Ren, L., Hong, Y., Yang, X., Ma, M., Zhang, Y., and Yuan, F. (2014). Improvement of multi-satellite real-time precipitation products for ensemble streamflow simulation in a middle latitude basin in South China. *Water Resour. Manage.*, 28(8):2259–2278.
- Kennedy, J. F., Kennedy, J., Eberhart, R. C., and Shi, Y. (2001). *Swarm Intelligence*. Morgan Kaufmann, San Francisco, California, USA.
- Khan, S. I., Hong, Y., Vergara, H. J., Gourley, J. J., Brakenridge, G. R., De Groeve, T., Flamig, Z. L., Policelli, F., and Yong, B. (2012). Microwave satellite data for hydrologic modeling in ungauged basins. *IEEE Geosci. Remote Sens. Lett.*, 9(4):663–667.
- Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., Brakenridge, G. R., Policelli, F., Habib, S., and Irwin, D. (2011). Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: implications for hydrologic prediction in ungauged basins. *IEEE Trans. Geosci. Remote Sens.*, 49(1):85–95.
- Kugler, Z. and De Groeve, T. (2007). The global flood detection system. Office for Official Publications of the European Communities, Luxembourg, 45.
- Mason, R. L. and Young, J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications*. Society for Industrial Mathematics, Philadelphia, USA.
- Mitra, A. (2009). Data transformation for normaliza-

- tion. In *Encyclopedia of Data Warehousing and Mining*, pages 566–571. IGI Global, Hershey, USA, 2nd edition.
- Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., and Nghiem, S. V. (2003). Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sens.*, 41(2):215–229.
- Quinlan, J. R. (1986). Induction of decision trees. *Mach. Learn.*, 1(1):81–106.
- Quinlan, J. R. (1987). Simplifying decision trees. *Int. J. Man-Mach. Stud.*, 27(3):221–234.
- Quinlan, J. R. and Rivest, R. L. (1989). Inferring decision trees using the minimum description length principle. *Info. Comput.*, 80(3):227–248.
- Salvia, M., Grings, F., Ferrazzoli, P., Barraza, V., Douna, V., Perna, P., and Karszenbaum, H. (2011). Estimating flooded area and mean water level using active and passive microwaves: the example of Paran River delta floodplain. *Hydrol. Earth Syst. Sci. Discuss.*, 8(2):2895–2928.
- Sattar, M. A., Bonakdari, H., Gharabaghi, B., and Radecki-Pawlik, A. (2019). Hydraulic modeling and evaluation equations for the incipient motion of sandbags for levee breach closure operations. *Water*, 11(279):1–22.
- Schmugge, T. J. (1980). Microwave approaches in hydrology. *Photogramm. Eng. Remote Sens.*, 46:495–507.
- Shabbak, A., Midi, H., and Hassan, M. N. (2011). The performance of robust multivariate statistical control charts based on different cutoff-points with sustained shifts in mean. *J. Appl. Sci.*, 11(1):56–65.
- Shaghaghi, S., Bonakdari, H., Gholami, A., Ebtehaj, I., and Zeinolabedini, M. (2017). Comparative analysis of GMDH neural network based on genetic algorithm and particle swarm optimization in stable channel design. *Appl. Math. Comput.*, 313:271–286.
- Shiklomanov, A., Lammers, R., and Vörösmarty, C. J. (2002). Widespread decline in hydrological monitoring threatens pan-Arctic research. *EOS Trans. AGU.*, 83(2):13–17.
- Sivapalan, M. (2003). Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrol. Process.*, 17(15):3163–3170.
- Srivastava, P. K. (2017). Satellite soil moisture: review of theory and applications in water resources. *Water Resour. Manage.*, 31(10):3161–3176.
- Stokstad, E. (1999). Scarcity of rain, stream gages threatens forecasts. *Science*, 285(5431):1199–1200.
- Su, F., Hong, Y., and Lettenmaier, D. P. (2008). Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. *J. Hydrometeorol.*, 9(4):622–640.
- Sullivan, J. H. and Woodall, W. H. (1996). A comparison of multivariate control charts for individual observations. *J. Qual. Tech.*, 28(4):398–408.
- Tekeli, A. E. and Fouli, H. (2017). Reducing false flood warnings of TRMM rain rates thresholds over Riyadh City, Saudi Arabia by utilizing AMSR-E Soil moisture information. *Water Resour. Manage.*, 31(4):1243–1256.
- Temimi, M., Lacava, T., Lakhankar, T., Tramutoli, V., Ghedira, H., Ata, R., and Khanbilvardi, R. (2011). A multi-temporal analysis of AMSR-E data for flood and discharge monitoring during the 2008 flood in Iowa. *Hydrol. Process.*, 25(16):2623–2634.
- Temimi, M., Leconte, R., Brissette, F., and Chaouch, N. (2007). Flood and soil wetness monitoring over the Mackenzie River Basin using AMSR-E 37 GHz brightness temperature. *J. Hydrol.*, 333(2):317–328.
- Theis, S. W., McFarland, M. J., Rosenthal, W. D., and Jones, C. L. (1982). *Microwave Remote Sensing of Soil Moistures*. Remote Sensing Center, Texas A&M University, College Station, TX. RSC-3458-129.
- Ulaby, F. T., Batlivala, P. P., and Dobson, M. C. (1978). Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I-bare soil. *IEEE Trans. Geosci. Elec.*, GE-16(4):286–295.
- Wang, J., O'Neill, P., and Engman, E. (1982). Remote Measurements of Soil Moisture by Microwave Radiometers at BARC Test Site II. 83954, NASA.
- Weigend, A. S. (2018). *Time Series Prediction: Forecasting the Future and Understanding the Past*. Routledge, Taylor and Francis, New York, USA.
- Yaseen, Z. M., Ebtehaj, I., Bonakdari, H., Deo, R. C., Danandeh Mehr, A., Melini Wan Mohtar, W. H., Diop, L., El-shafieh, A., and Singh, V. P. (2017). Novel approach for streamflow forecasting using a hybrid ANFIS-FFa model. *J. Hydrol.*, 554:263–276.
- Young, C. C., Liu, W. C., and Wu, M. C. (2017). A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Appl. Soft Comput.*, 53:205–216.
- Zaji, A. H. and Bonakdari, H. (2019). Robustness lake water level prediction using the search heuristic

- based artificial intelligence methods. *ISH J. Hydraul. Eng.*, 25(3):316–324.
- Zaji, A. H., Bonakdari, H., and Gharabaghi, B. (2018). Remote sensing satellite data preparation for simulating and forecasting river discharge. *IEEE Trans. Geosci. Remote Sens.*, 56(6):3432–3441.
- Zaji, A. H., Bonakdari, H., and Gharabaghi, B. (2019a). Applying upstream satellite signals and a 2-D Error minimization algorithm to advance early warning and management of flood water levels and river discharge. *IEEE Trans. Geosci. Remote Sens.*, 99:1–9.
- Zaji, A. H., Bonakdari, H., and Gharabaghi, B. (2019b). Developing an AI-based method for river discharge forecasting using satellite signals. *Theor. Appl. Climatol.*, 138:347–362.
- Zaji, A. H., Bonakdari, H., Shamshirband, S., and Qasem, S. N. (2015). Potential of particle swarm optimization based radial basis function network to predict the discharge coefficient of a modified triangular side weir. *Flow Measur. Instrum.*, 45:404–407.
- Zeynoddin, M., Bonakdari, H., Azari, A., Ebtehaj, I., Gharabaghi, B., and Riahi, H. M. (2018). Novel hybrid linear stochastic with non-linear extreme learning machine methods for forecasting monthly rainfall a tropical climate. *J. Environ. Manage.*, 222:190–206.