Evaluation of coupled ANN-GA model to prioritize flood source areas in ungauged watersheds

Naser Dehghanian, S. Saeid Mousavi Nadoushani, Bahram Saghafian and Morteza Rayati Damavandi

ABSTRACT

An important step in flood control planning is identification of flood source areas (FSAs). This study presents a methodology for identifying FSAs. Unit flood response (UFR) approach has been proposed to quantify FSAs at subwatershed and/or cell scale. In this study, a distributed ModClark model linked with Muskingum flow routing was used for hydrological simulations. Furthermore, a fuzzy hybrid clustering method was adopted to identify hydrological homogenous regions (HHRs) resulting in clusters involving the most effective variables in runoff generation as selected through factor analysis (FA). The selected variables along with 50-year rainfall were entered into an artificial neural network (ANN) model optimized via genetic algorithm (GA) to predict flood index (FI) at cell scale. The case studies were two semi-arid watersheds, Tangrah in northeastern Iran and Walnut Gulch Experimental Watershed in Arizona. The results revealed that the predicted values of FI via ANN-GA were slightly different from those derived via UFR in terms of mean squared error (MSE), mean absolute error (MAE), and relative error (RE). Also, the prioritized FSAs via ANN-GA were almost similar to those of UFR. The proposed methodology may be applicable in prioritization of HHRs with respect to flood generation in ungauged semi-arid watersheds.

Key words | artificial neural network (ANN), flood source areas (FSAs), hydrological homogenous regions (HHRs), ModClark, semi-arid ungauged watersheds, unit flood response (UFR)

Naser Dehghanian
S. Saeid Mousavi Nadoushani (corresponding author)
Department of Water Resources Management, Faculty of Civil, Water and Environmental Engineering,
Shahid Beheshti University,
Tehran,
I.R. Iran
E-mail: *sa_mousavi@sbu.ac.ir*

Bahram Saghafian
Department of Civil Engineering, Science and Research Branch,
Islamic Azad University,
Tehran,
I.R. Iran

Morteza Rayati Damavandi
Department of Technical and Engineering,
Islamic Azad University,
Qaemshahr,
Iran

INTRODUCTION

Identification of flood source areas (FSAs) through detailed simulation and analysis of runoff generation processes at cell scale is one of the most important tasks in flood management. The physically based distributed hydrological models (DHM) are often used to represent the complex, dynamic, and nonlinear rainfall-runoff (R-R) response; however, they require a large volume of data to run and sometimes have difficulty accounting for land-cover changes and parameter calibration. Moreover, hydrological models are inadequate for flood mapping studies for ungauged or poorly gauged watersheds. A valuable tool to overcome this modeling limitation is artificial neural network (ANN) application. Black box or data-driven models, such as ANN models, have also been developed to simulate R-R processes in flow prediction, because the rainfall-runoff process is a complex and nonlinear one. In addition, ANNs are often less expensive to train and simpler to implement in hydrologic applications than other types of conceptual/physical rainfall-runoff models (Abrahart *et al.* 2004). Awchi & Srivastava (2009) stated that although data-driven techniques often require similar data to physically based models, they need much

doi: 10.2166/nh.2020.141

less development time, are useful for real-time applications, and prove capable of accurately predicting streamflow.

In the last two decades, computational intelligence techniques, such as ANNs, have proven in many fields to be able to simulate complex and nonlinear systems because they have the ability of 'learning' from data. ANNs have received a great deal of attention in different aspects of hydrology, such as rainfall-runoff modeling (e.g., Sudheer *et al.* 2002; Bharti *et al.* 2017), river discharge prediction in semi-arid regions (e.g., Mwale *et al.* 2014; Babaei *et al.* 2019), and flood hazard mapping (e.g., Kourgialas & Karatzas 2017). Given their intrinsic nature, ANNs do not require detailed knowledge of physically complex processes of a system in order to recognize the relationship between input and output. ANN drawbacks are related to their black box approach (i.e., no physical meaning of the concerning parameters) and extrapolation capacity (He *et al.* 2014). Kişi *et al.* (2012) mentioned that computational models may offer promising alternatives for continuous streamflow prediction.

Some researchers have reported that the feedforward neural network (FFNN) model named multilayer perceptron (MLP) has better capability in rainfall-runoff modeling compared with other types of ANN models (e.g., Shamseldin 1997; Rezaeianzadeh *et al.* 2014; Bharti *et al.* 2017). MLP is perhaps the most popular ANN architecture which is believed to be capable of approximating arbitrary functions (Kisi 2005). This has been important in the study of nonlinear dynamics. MLPs are typically trained using the back propagation (BP) algorithm (Rumelhart *et al.* 1986) in which the network's interconnecting weights are iteratively changed to minimize the predefined error, which is mean square error (MSE) (Sudheer *et al.* 2002).

Genetic algorithm (GA) (Holland 1975) is a type of evolutionary algorithm that can be used for calibrating the rainfall-runoff models through an evolutionary search on a population of randomly generated individuals over a number of generations (e.g., Khazaei *et al.* 2014). The GA has also been applied to optimize the training of data resulting in improved ANN simulation of a hydraulic flow model (Kamp & Savenije 2006). Hence, the hybrid ANN-GA is a powerful method for simulation and forecasting of surface flows (Bozorg-Haddad *et al.* 2016).

In recent decades, different geostatistical and mathematical methods along with distributed hydrological models (DHM) have been developed for watershed partitioning into homogenous hydrological regions (HHRs). Clustering, as a data-driven model, has been used to identify HHRs using hydrologic or watershed characteristics. Also, clustering as a multivariate technique attempts to discover structures or certain patterns in a data set, where the objects inside each cluster show a certain degree of similarity (Araghinejad 2014). Fuzzy c-means (FCM) (Bezdek 1981) has been widely applied in numerous applications involving discretization of watersheds into groups (e.g., Rao & Srinivas 2006) and in regionalization studies (e.g., Basu & Srinivas 2015).

Hybrid self-organizing feature mapping (SOFM) (Kohonen 1982) coupled with the FCM clustering method, denoted as SOMFCM in this study, has been successfully applied in the regionalization of sites or watersheds (e.g., Srinivas *et al.* 2008; Singh & Singh 2017). It is believed that the performance of the FCM algorithm is improved when coupled with the SOFM method. SOMFCM can identify homogeneous regions more quickly and accurately while minimizing the FCM objective function (Nadoushani *et al.* 2018).

A prerequisite to any flood control strategy is the evaluation of spatial distribution of surface runoff generation and subsequent flood translation-attenuation throughout a watershed–stream network. Saghafian & Khosroshahi (2005) proposed a unit flood response (UFR) approach to identify and prioritize FSAs of a given watershed on the basis of their contribution to the flood response at the main outlet. In the UFR approach, unit areas at the desired scale in the watershed are successively removed in the simulation process, and their effects at the outlet are quantified. There have been a number of UFR applications for spatial prioritization of runoff sources (e.g., Saghafian *et al.* 2010, 2012, 2015; Sanyal *et al.* 2014). As such, UFR may help managers to effectively select areas for flood control practices through identifying critical FSAs.

The main objective of this study is to assess the capability of a fuzzy hybrid clustering method coupled with ANN-GA in prioritizing FSAs (or identifying high ranks of FSAs) which are the most important areas for flood management in a gauged or ungauged semi-arid watershed. Prioritization is based on flood index (FI) quantity (i.e., FSAs map) as simulated by a DHM and resulting from the application of UFR approach at cell scale in a gauged watershed. The study consists of three main parts. First, the FSAs are identified/quantified via a calibrated hydrological model

and implementing the UFR approach. Second, HHRs are derived by SOMFCM. Then, the coupled ANN-GA is applied to predict FI in HHRs at cell scale using some dominant input variables. Third, the transferability of the proposed methodology (ANN-GA) in FI prioritization is evaluated in two semi-arid watersheds.

METHODOLOGY

Hydrological modeling

A combined DHM and flow routing method was implemented in HEC-HMS to identify the FSAs. The SCS-CN excess rainfall method, a modified version of the Clark rainfall-runoff model (ModClark) at cell scale, and Muskingum river flow routing method were the choices activated in the HEC-HMS.

Rainfall excess determined for each grid cell is lagged based on the travel time to the outlet for that grid cell (Equation (1)). The travel time for each cell is based on the maximum travel time to the watershed outlet which follows this relationship (Kull & Feldman 1998):

$$t_{cell} = T_c\left(\frac{l_{cell}}{l_{max}}\right) \tag{1}$$

where t_{cell} is the travel time of each cell to the outlet (in s), T_c is the time of concentration of the watershed (in s), l_{cell} is the flow distance from each cell to the outlet (in m), and l_{max} is the maximum flow length in the watershed (in m). Time variation of excess rainfall can also be taken into account in the runoff hydrograph determination. The time–area hydrograph is then given by:

$$Q_j = \sum_{k=1}^{j} E_k A_{j-k+1} \tag{2}$$

where j is the time step number, Q is the watershed outlet runoff (in m^3/s or cms), E is the excess rainfall intensity (in mm/s), and A is the area between adjacent isochrones (in m^2). Once the time–area hydrograph is determined, it is routed in a linear reservoir with a storage coefficient R (in s) to account for the effects of watershed storage.

T_c and R are main ModClark model parameters at subwatershed scale. In this study, empirical well-known relationships such as Kirpich, Bransby Williams and Kerby were used to initially estimate T_c values based on physical features of the subwatersheds. Overall, three parameters, initial abstraction ratio (I_a), T_c, and R, were calibrated along with two Muskingum parameters, X and K.

UFR approach

To quantify flood impact of subwatershed cells on watershed outlet, the UFR approach (Saghafian & Khosroshahi 2005) was adopted. UFR, mostly implemented in conjunction with variants of the time-area method (e.g., ModClark), defines two indices to prioritize subwatershed cells on the basis of the quantity of their contribution to the flood peak at the main outlet of the watershed. In this study, the following index was used:

$$FI_i = \frac{Q_p - Q_i}{A_i} \tag{3}$$

where FI_i (hereafter FI) is the unit flood index of the ith cell (in cm/km^2), Q_p is the simulated peak flood discharge at the outlet with all cells in the base simulation (in cm), Q_i is the peak flood discharge at the outlet after removing the ith cell, and A_i is the area of ith cell (in km^2). The A_i corresponding to cell area is uniform for all cells and equal to $0.5 \times 0.5 = 0.25$ (km^2) and $0.15 \times 0.15 = 0.0225$ (km^2) for the Tangrah and Walnut Gulch Experimental Watersheds (WGEW), respectively.

Fuzzy hybrid clustering method (SOMFCM)

Among fuzzy clustering methods, the FCM algorithm (Bezdek 1981) is simple and comprehensive, wherein each data point belongs to a cluster to some degree that is specified by a membership grade. Similar to most clustering methods, the main disadvantage of the FCM algorithm is its sensitivity to m (the weight exponent or fuzzification parameter) and c (number of clusters), both of which should be optimized simultaneously. To do this, some proposed techniques along with applying selected conventional fuzzy cluster validation indices (Pal & Bezdek 1995) have been used.

A particular category of ANN, known as SOFM (Kohonen 1982), has been used as a clustering tool in order to convert the complex nonlinear statistical

relationship among high-dimensional input data into a simple and geometric relationship on a low-dimensional display (Nourani *et al.* 2015).

In this study, SOMFCM, a fuzzy hybrid clustering approach in which the SOFM method provides the initial cluster centers to improve performance of the FCM algorithm for discretization of watersheds into HHRs was assembled using features of R programming language.

ANN-GA model

The GA is an evolutionary algorithm whose logic mimics evolutionary processes. An objective function is calculated based on the decision variables, which are initialized randomly. In this paper, the GA was used to determine the ANN parameters such as the number of hidden layers and the number of neurons in each layer. In addition, in nonlinear functions optimization using traditional methods, it is possible that the algorithm converges to local optima instead of finding the global optimal solution, whereas the GA has the potential of avoiding those local optima.

An ANN is a simulation technique based on the human nervous system. It has a structure that mimics the human brain and the body's nervous system. ANN is a powerful tool for forecasting complex phenomena (Govindaraju 2000). The ANN structure consists of neurons arranged in layers. In an ANN, each neuron operates independently while the whole network's capability is based on the performance of each individual neuron. In general, the number of neurons, layers (consisting of input, hidden, and output layers), and weights should be defined to create the structure of an ANN model. The number of hidden layers and number of neurons in each hidden layer should be determined using experimental and trial-and-error methods to avoid under- and/or over-fitting problems (Valipour *et al.* 2013). Figure 1 shows the schematic view of the ANN model used in this study.

In this study, a number of multilayer perceptron (MLP) networks with one input layer, one, two, or without any hidden layers, and one output layer along with back propagation algorithm was employed in order to predict FI value at cell scale. The inputs to the ANN model were watershed characteristics' layers while the corresponding output was the FI as simulated via UFR. In addition, tangent sigmoid and linear were the transition functions for hidden and output layers, respectively. The Levenberg–Marquardt algorithm is believed to produce reasonable results for most ANN applications (Awchi 2014), therefore it was selected for this study.

In this study, the parameters of ANN architecture (i.e., the number of hidden layers and the number of neurons in each layer) were optimized via GA for each watershed. Maximum number of the epochs was selected as 1,000.

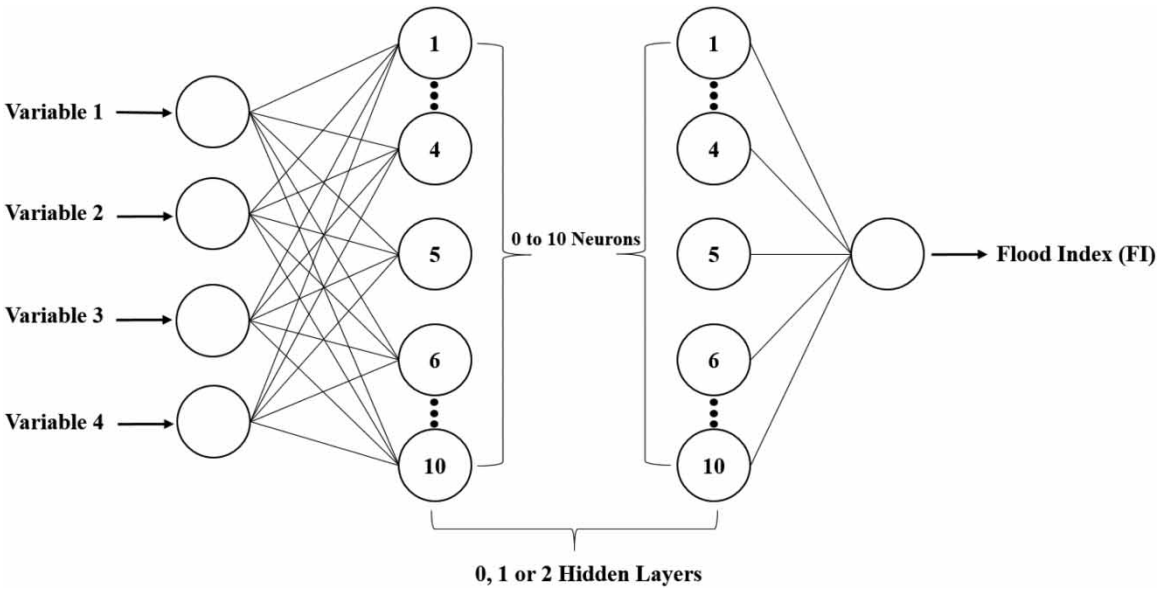

Figure 1 | Schematic view of the ANN model used in the study.

Furthermore, the performance of the ANN model was evaluated using mean squared error (MSE), mean absolute error (MAE), and relative error (RE), calculated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(FI_{i\_UFR} - FI_{i\_ANN})^2 \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|FI_{i\_UFR} - FI_{i\_ANN}| \quad (5)$$

$$RE = \frac{|FI_{i\_UFR} - FI_{i\_ANN}|}{FI_{i\_UFR}} \times 100 \quad (6)$$

where FI_{i_UFR} is FI value simulated via UFR, FI_{i_ANN} is FI value predicted via ANN for each cell, and n is number of total cells belonging to each clustered HHR, predictable homogenous zone (PHZ), or subwatershed.

Study areas and spatial data

Case study 1

The Tangrah watershed, covering 1,970 km^2 in area, is located in northeastern Iran. For this study, the watershed was divided into six major subwatersheds (Figure 2).

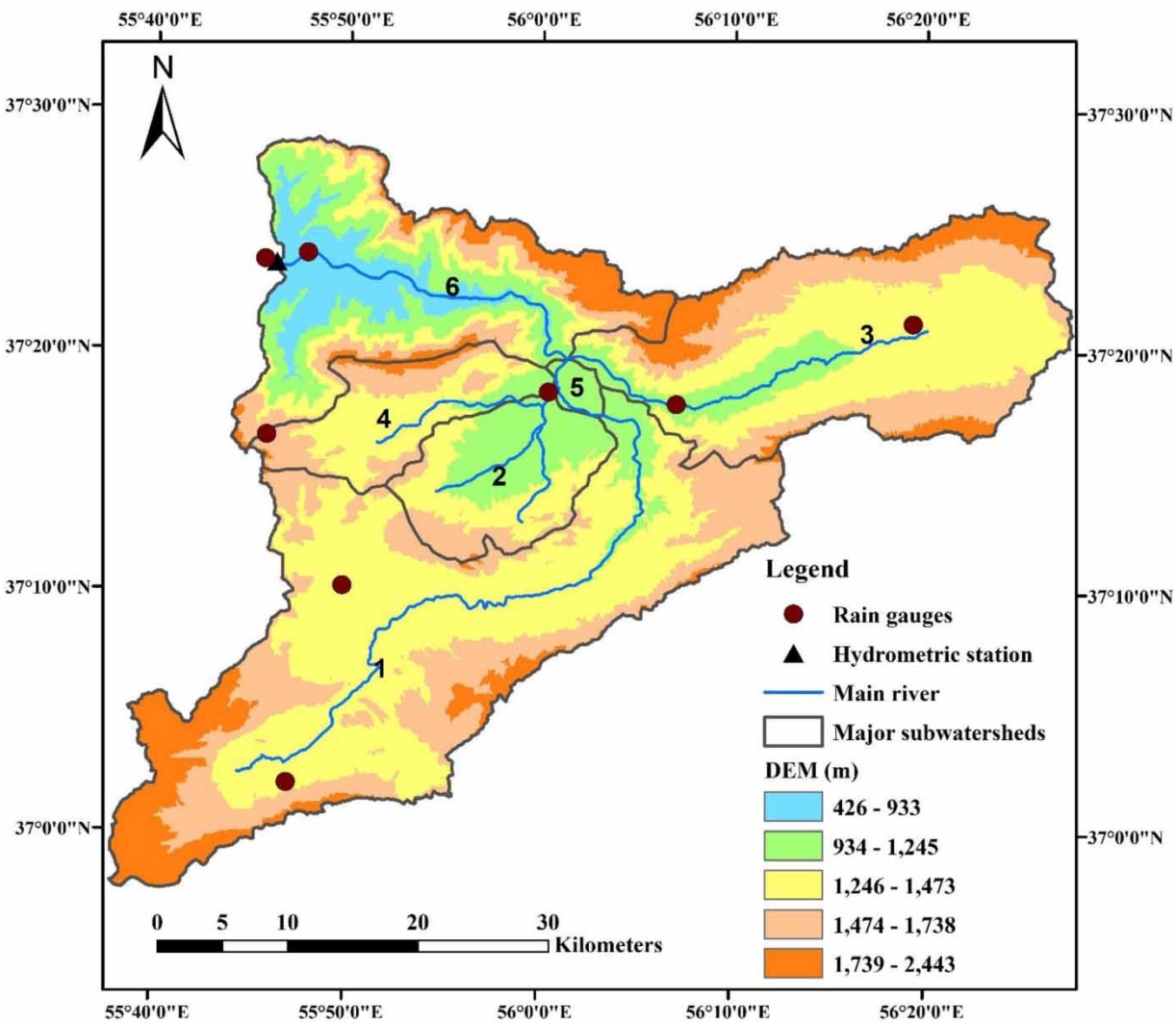

Figure 2 | Map of Tangrah watershed.

The dominant climate is of semi-arid type. Although most rainfall occurs in winter, summers are not entirely dry, such that floods of large magnitudes mainly form in summer thanks to high intensity storms. The watershed-average annual rainfall is about 400 mm. Vegetation varies from poor range in the east and south to dense forest at subwatersheds 4 and 6 (Saadat *et al.* 2011). Also, soil texture consists of loamy sand and sand, loam, sandy clay loam, and clay loam. Table 1 lists the main characteristics of all subwatersheds.

Case study 2

The eastern part of the well-gauged WGEW with an area of 93.4 km^2 was selected as the second study area (Figure 3). The climate is classified as semi-arid, with mean annual temperature of 17.7 °C and mean annual precipitation of 350 mm. WGEW is managed by the Southwest Watershed Research Center (SWRC) in Tucson. Grassland is the primary land use while soils are dominantly sandy and gravelly loams. Almost all the runoff is generated by summer thunderstorms, and runoff volumes and peak flow rates vary greatly with area and on an annual basis (USDA 2007). Subwatersheds' characteristics are presented in Table 2.

Tangrah and WGEW study areas have similarities including high flood potential, especially in the summer period, while they enjoy a sufficient number of recorded rainfall-runoff data for model calibration. In this study, slope, curvature, CN, distance from the nearest stream (Distance), flow length, normalized difference vegetation index (NDVI), soil saturated hydraulic conductivity (Ksat), and Topographic Position Index (TPI) as an index reflecting tendency to saturate (Weiss 2001) were selected as the layers that influence rainfall-runoff processes.

The flow chart of the study procedure is presented in Figure 4.

According to Figure 4, the FA, SOMFCM, and ANN-GA were applied for FI prediction in HHRs using effective input features (right path in Figure 4) while hydrological modeling and UFR approach were used to generate the FSAs map (left path in Figure 4). As HHRs have an almost similar response to rainfall, they are to be expected to facilitate ANN in the training phase. All steps in both parts are carried out at cell scale. Then, simulated (UFR) and predicted FI (ANN) are compared to each other.

Table 1 | Characteristics of Tangrah subwatersheds

Subwatershed	Area (km^2)	Mean slope (%)	Length of main river (km)	Curve number (CN II)
1	745.3	14.2	72.4	76.1
2	144.3	13.1	23.3	71
3	451.8	14.9	48.4	70.5
4	125.7	18.4	27.9	80.7
5	13.1	11.8	8.1	78.7
6	306.9	36.1	36.7	81.1

RESULTS AND DISCUSSION

R-R model calibration

In this study, two and seven storm events in two different study areas (i.e., Tangrah and WGEW) over the 2000–2018 period were used for calibration, respectively. Through inspection of recorded rainfall and runoff data, most of the annual maximum peak flows have occurred during summer in both watersheds. As a result, the dominant antecedent moisture condition (AMC) was selected as dry. Also, the cell sizes were selected as 500 and 150 m for Tangrah and WGEW, respectively. Based on the ratio of simulated to observed discharge (Qobs./Qsim.), manual calibration focusing on peak flows was carried out at 1-hour and 10-minute time steps for Tangrah and WGEW, respectively (Tables 3 and 4). In Tangrah, most of the calibrated parameters were similar to the ones obtained by Saghafian *et al.* (2014). Figure 5(a)–5(d) show some selected hydrographs at the main outlets of the study areas.

As shown in Figure 5, the ModClark model and Muskingum flow routing method were well calibrated in both watersheds. Moreover, the average ratio of simulated to observed discharge (Qobs./Qsim.) was determined as 87.8 and 73.4 percent as the best result of manual calibration for Tangrah and WGEW, respectively. The results of calibration were also similar to Golian *et al.* (2010) (for Tangrah) and Foda *et al.* (2017) (for WGEW).

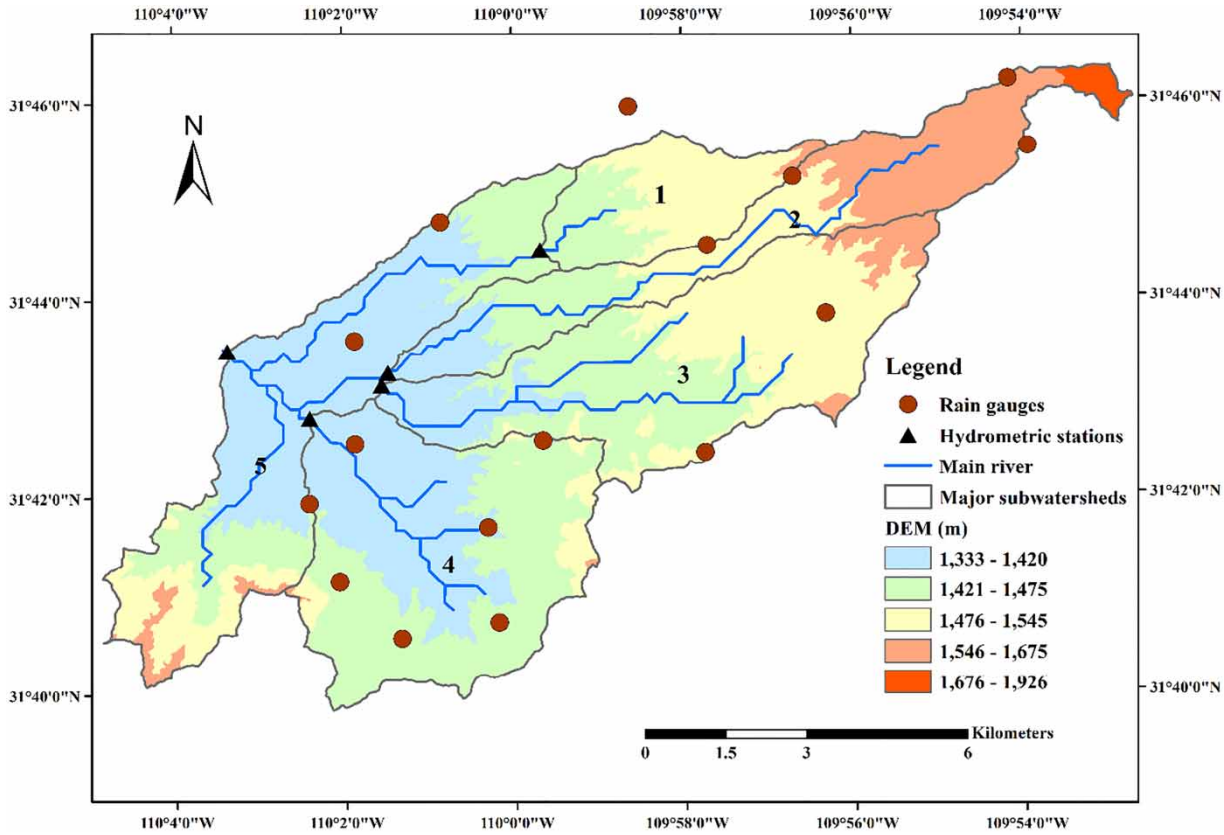

Figure 3 | Map of WGEW watershed.

Table 2 | Characteristics of WGEW subwatersheds

Subwatershed (runoff measuring flume No.)	Area (km^2)	Mean slope (%)	Maximum stream length to the outlet (km)	Curve number (CN II)
1 (11)	7.8	9.1	3.8	58.1
2 (10)	15.6	9.6	14.8	61.9
3 (9)	23.8	10.7	11.9	61.3
4 (15)	23.6	6.8	6.4	68.7
5 (6)	22.5	12.7	7.4	70.7

Application of UFR

Rainfall depth corresponding to a 50-year return period (denoted as P50) was derived from the rainfall frequency analysis of available long-term annual maximum time series for some rain gauge stations. The point-wise 50-year design storm was spatially interpolated over the watershed using the inverse distance weighted (IDW) method. The Huff 2nd quartile curve was derived and applied to represent the temporal pattern of the design storm in each subwatershed. Storm duration was approximately set equivalent to the watershed time of concentration (i.e., 12 hours for Tangrah and 75 minutes for WGEW). According to Saghafian *et al.* (2014), the spatially averaged 5-day rainfall accumulations prior to the occurrence date of annual maximum storm events for all subwatersheds were determined in both study areas over a long-term period. The results revealed that most of the maximum annual daily rainfall events were categorized as AMC I (dry).

Since the UFR target point was set to the main outlet of the watersheds, there was a need to perform Muskingum flow routing through river reaches. Therefore, for UFR application, the calibrated model (Tables 3 and 4) was subjected to the 50-year return period rainfall under dry AMC condition. The peak flow discharge corresponding to the 50-year rainfall was simulated as 233.6 and 82.2 (cm) for Tangrah and WGEW, respectively. Then, the UFR approach

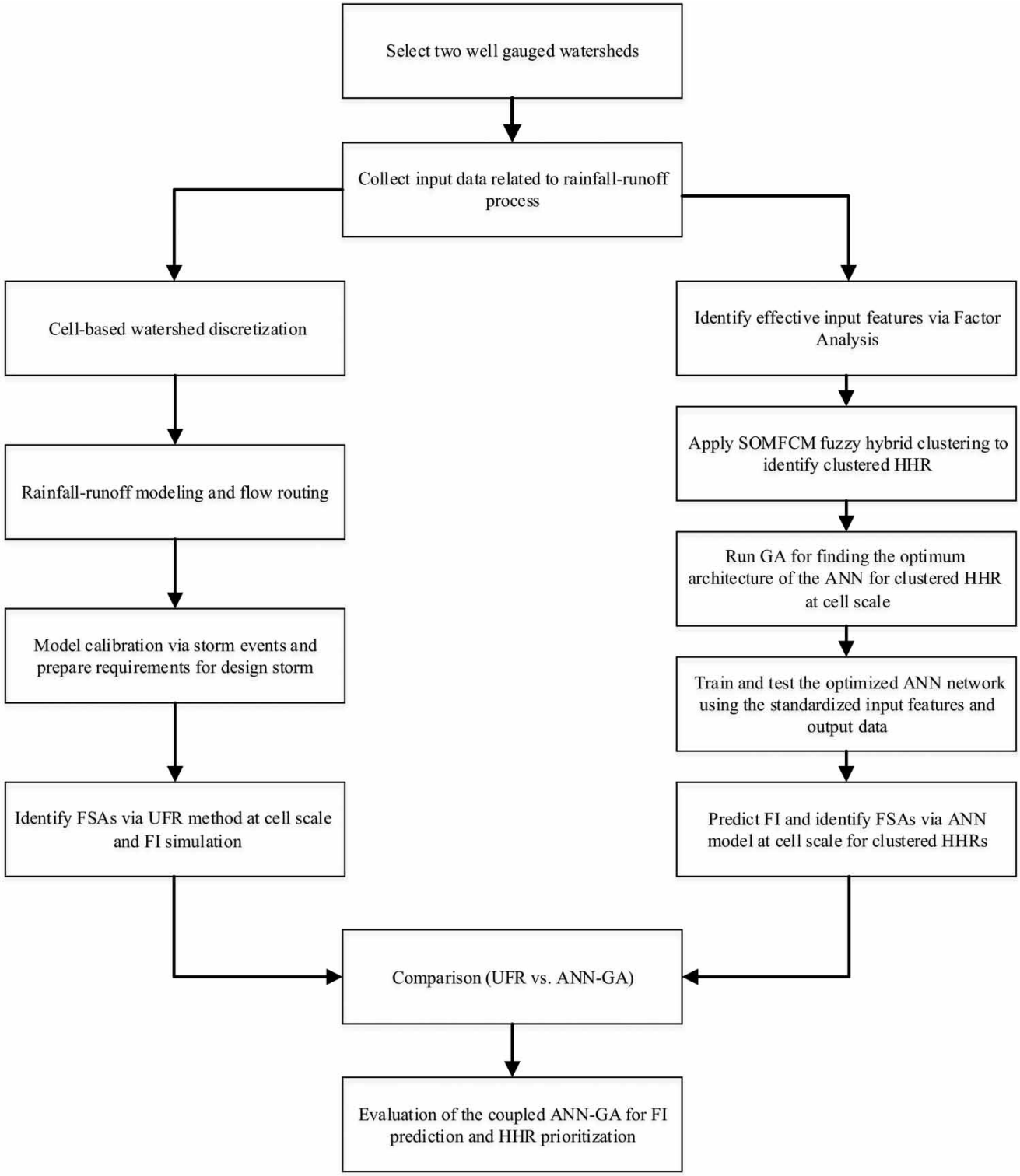

Figure 4 | Flowchart of the study procedure.

Table 3 | Average values of calibrated parameters in Tangrah

Rainfall-runoff parameters				Flow routing parameters		
Subwatershed	Tc (min)	R (min)	Ia	Reach No. (length in km)	K (min)	X
1	470	470	0.12	1 (8.1)	35	0.2
2	155	155	0.12	2 (36.7)	150	0.2
3	307	307	0.12	–	–	–
4	165	165	0.12	–	–	–
5	54	54	0.12	–	–	–
6	195	195	0.12	–	–	–

Note: Calibration events: 10/08/2001 and 09/08/2005.

Table 4 | Average values of calibrated parameters in WGEW

Rainfall-runoff parameters				Flow routing parameters		
Subwatershed	Tc (min)	R (min)	Ia	Reach No. (length in km)	K (min)	X
1	17	9	0.06	1 (7.6)	28	0.25
2	54	27	0.13	2 (3.93)	19	0.25
3	43	22	0.04	3 (3.91)	14	0.25
4	38	19	0.12	4 (2.4)	12	0.25
5	27	14	0.18	–	–	–

Note: Calibration events: 11/08/2000, 27/08/2003, 17/08/2006, 20/07/2007, 19/07/2008, 25/07/2008, and 03/07/2012.

was performed via removing one-by-one watershed cells. Therefore, one model run was undertaken for each cell (i.e., 7,380 model runs for Tangrah and 4,004 model runs for WGEW) in order to determine FI for each cell.

SOMFCM clustering

At first, factor analysis (FA) was applied in order to reduce the dimensionality of the data set. The total number of cells entering into FA was equal to 7,380 and 4,004 for each selected clustering input variable in Tangrah and WGEW, respectively. In this study, 8 and 12 distributed variables of watershed features were entered into FA for Tangrah and WGEW, respectively. Three superior variables were DEM, TPI, and Slope (for Tangrah) and DEM, NDVI, and TWI (for WGEW). Since rainfall is indeed the most important factor in rainfall-runoff events, it must be taken into account in the prioritization of the regions in terms of potential runoff generation (Saghafian & Khosroshahi 2005). Thus, the P50 layer must be inserted into the clustering procedure along with the most effective variables resulting from the FA.

Different numbers of clusters and SOFM configurations were evaluated through several runs based on the silhouette width (Rousseeuw 1987). In summary, the best configuration of lattice neighborhood structure corresponding to each number of clusters was entered into the fuzzy clustering for both watersheds. Then, the FCM clustering was implemented in various possible runs (i.e., for $1.1 \leq m \leq 2.5$ and for $4 \leq c \leq 9$) in order to determine optimum m and c. For this purpose, six well-known conventional validation indices including V_{EXB}, the extended FCM Xie–Beni index (Pal & Bezdek 1995), V_K, Kwon index (Kwon 1998), V_{FS}, Fukuyama and Sugeno validity measure (Fukuyama & Sugeno 1989), V_{PC}, partition coefficient (Bezdek 1974), V_{PE}, partition entropy (Bezdek 1974), and *CI*, confusion index (Burrough *et al.* 1997) were calculated for different m and c. Table 5 presents the range of each aforementioned indices.

Since it was difficult to determine optimum m and c based on the results of fuzzy validation indices, other validation techniques (e.g., Nadoushani *et al.* 2018) were applied. In this study, optimum values of $m = 1.6$ and $c = 5$ were dominant in most clustering runs in both watersheds, as presented in Table 5. Therefore, Tangrah and WGEW may be divided into five clusters or HHRs based on the optimum SOMFCM clustering (Figures 6(a) and 7(a)). In other words, each watershed cell is assigned to one of five clusters. Therefore, the ANN would be developed in these clusters. Furthermore, the UFR-derived FSAs map is shown in Figures 6(b) and 7(b).

Identification of areas with high flood potential is important in flood control studies. According to Figures 6(b) and 7(b), most parts of the study areas have low to very low flood potential so that only 1.1% of Tangrah and 2.3% of WGEW fall within high to very high FSA classes. Comparison of Figures 6(a) and 7(a) shows that areas of high flood classes fall within clusters 4 and 5 in Tangrah and clusters 1 and 2 in WGEW. Therefore, in order to identify the most critical SOMFCM clusters in terms of flood potential, dominant surface factors resulting from FA should be taken into account.

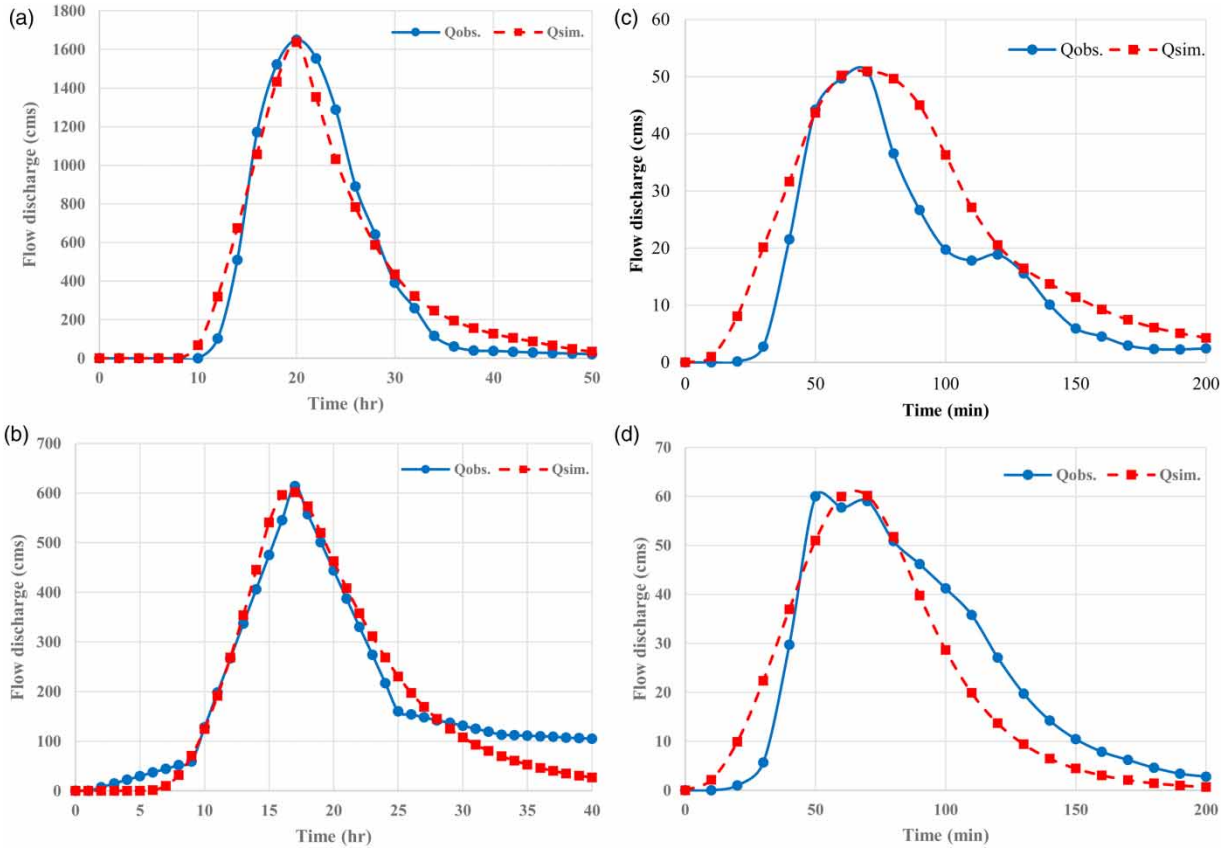

Figure 5 | Observed and simulated hydrographs at the main outlet of Tangrah: (a) event of 10/08/2001 and (b) event of 09/08/2005; and WGEW: (c) event of 27/08/2003 and (d) event of 19/07/2008.

Table 5 | Selected results of conventional fuzzy validation indices

Watershed	Index	V_{EXB}	V_K	V_{FS}	V_{PC}	V_{PE}	CI
Tangrah	min	0.19	1471	−16146	0.67	0.25	0.15
	max	0.51	3783	−11,562	0.87	0.72	0.34
	m = 1.6 and c = 5	0.49	3676	−12,519	0.73	0.45	0.28
WGEW	Min	0.32	1310	−12,001	0.41	0.03	0.23
	max	1.08	4360	−2671	0.98	1.19	0.464
	m = 1.6 and c = 5	0.4	1628	−8349	0.74	0.53	0.27

Application of coupled ANN-GA

In this study, the watershed cells which have common features in flood generation are denoted by PHZ. In fact, PHZs are parts of HHRs (as derived by SOMFCM) distributed throughout the subwatersheds (Tables 6 and 7 and Figures 6(a) and 7(a)).

As presented in Tables 6 and 7, there are 20 PHZs in each study watershed in which the ANN model is trained and/or predicted at cell scale. For instance, as shown in Figure 6(a), five PHZs are located in subwatershed 4. Also, in WGEW (Figure 7(a)), subwatershed 1 has only three PHZs. On the other hand, some PHZs are only used in the prediction phase, because their number of

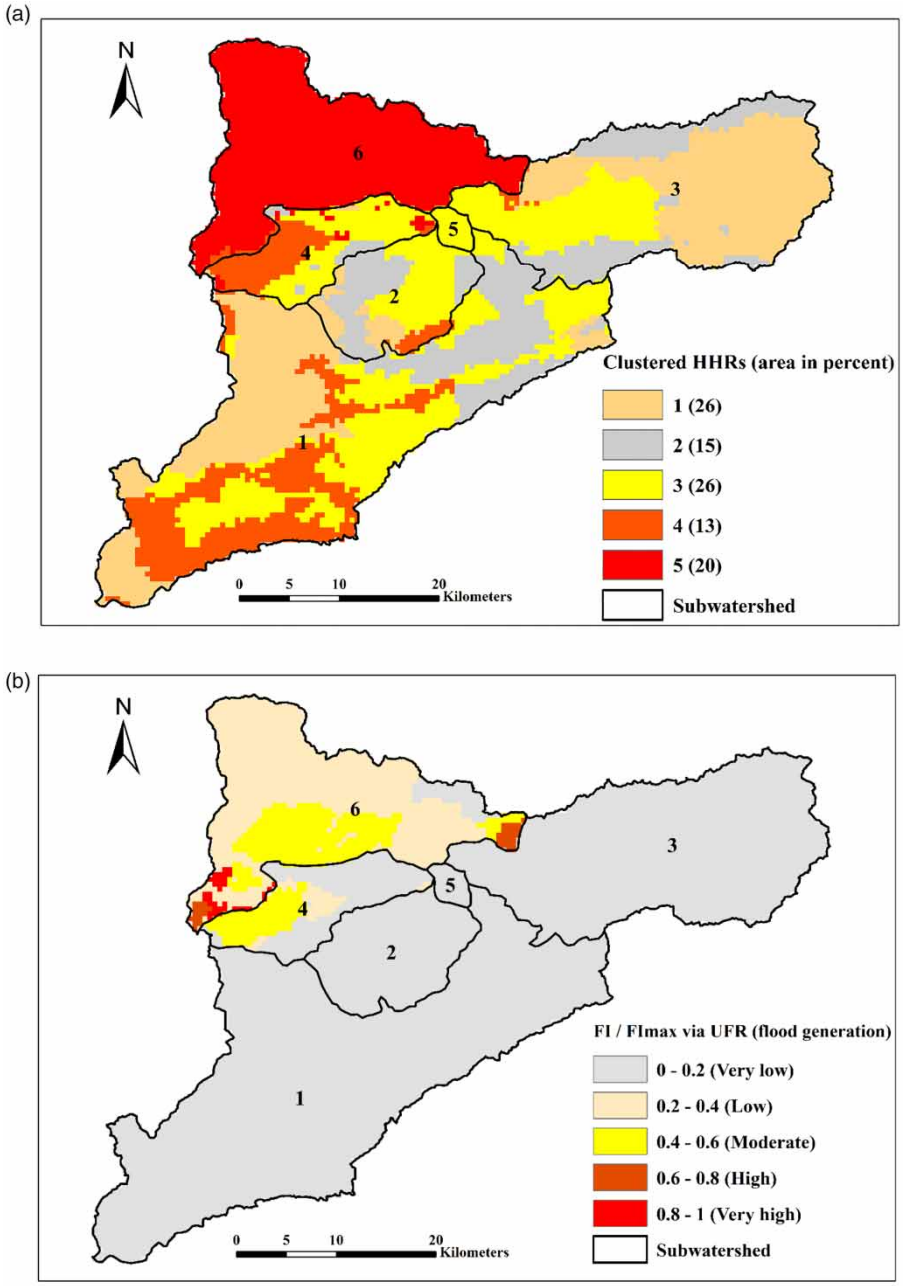

Figure 6 | (a) Map of delineated HHRs via SOMFCM clustering method and (b) map of UFR-derived FI/FImax in Tangrah.

cells is less than the predefined threshold number (i.e., in this study, equal to 20% of the total number of cells in the corresponding cluster). As a result, PHZs larger than the aforementioned threshold could be used in both training and prediction phases. For example, in Tangrah, PHZ 1 is trained and PHZs 5, 9, and 13 are predicted, because they are all in the same cluster. Then, PHZ 9 is trained and PHZs 1, 5, and 13 are predicted. Here, PHZs 5 and 13 are only predicted based on the trained ANN (PHZs 1 and 9). Due to the aforementioned threshold, PHZ 17 (in Tangrah) and PHZ 20 (in WGEW) were used only for prediction.

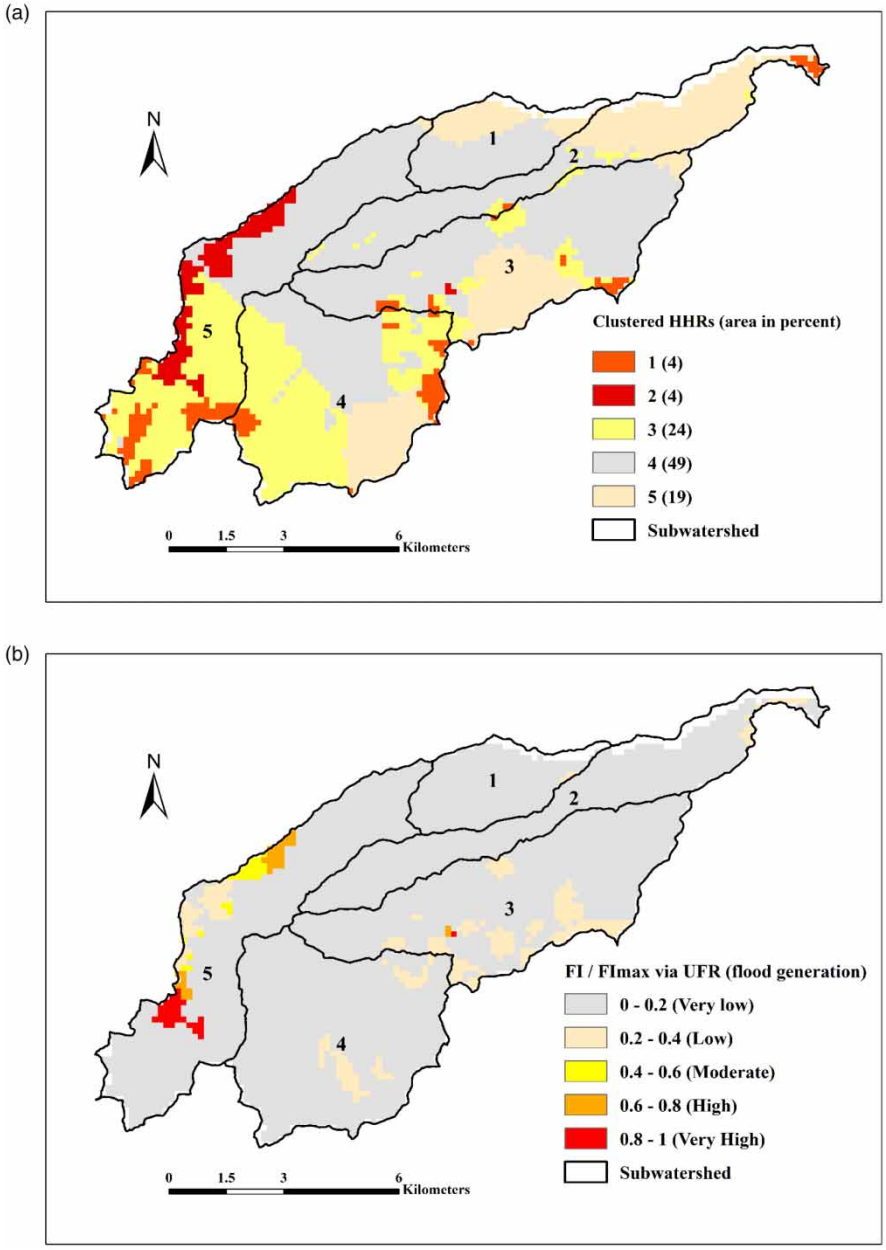

Figure 7 | (a) Map of delineated HHRs via SOMFCM clustering method and (b) map of UFR-derived FI/FImax in WGEW.

As there is no default selection for an ungauged subwatershed, PHZs with different values of mean FI for each subwatershed could be separately tested for prediction. The given ungauged subwatershed may include at least one to up to a number of clusters. In this ration, all features of the gauged and ungauged watersheds must be entered into the clustering stage to form common HHRs (i.e., clusters) before the ANN procedure is implemented. As such, transferability of the trained ANN is evaluated for FI prediction in an adjacent ungauged subwatershed having some PHZs.

FI prediction was performed for PHZs by the superior variables entered into the clustering at cell scale. The contributing variables in the prediction were DEM-TPI-Slope-P50 for Tangrah and DEM-TWI-NDVI-P50 for WGEW. The standardized inputs (i.e., four variables for each watershed)

Table 6 | Characteristics of PHZs in Tangrah

PHZ No.	PHZ description	No. of cells	FI * 0.01 (cm/km^2)	Rank (flood generation)	Used in phase
1	Sub1_cluster1	879	0.043	20	Training and prediction
2	Sub1_cluster2	488	0.060	17	Training and prediction
3	Sub1_cluster3	937	1.709	11	Training and prediction
4	Sub1_cluster4	703	4.167	9	Training and prediction
5	Sub2_cluster1	99	0.430	15	Prediction
6	Sub2_cluster2	219	0.044	19	Prediction
7	Sub2_cluster3	228	1.962	10	Prediction
8	Sub2_cluster4	34	6.457	7	Prediction
9	Sub3_cluster1	953	0.206	16	Training and prediction
10	Sub3_cluster2	393	0.044	18	Training and prediction
11	Sub3_cluster3	439	1.323	12	Training and prediction
12	Sub3_cluster4	12	4.737	8	Prediction
13	Sub4_cluster1	10	13.088	5	Prediction
14	Sub4_cluster2	34	1.039	13	Prediction
15	Sub4_cluster3	248	21.213	4	Prediction
16	Sub4_cluster4	175	60.532	2	Training and prediction
17	Sub4_cluster5	27	12.982	6	Prediction
18	Sub5_cluster3	51	0.758	14	Prediction
19	Sub6_cluster4	15	142.013	1	Prediction
20	Sub6_cluster5	1,436	48.107	3	Training

and one output (i.e., FI simulated via UFR) of the two case studies were randomly used to evaluate the proposed ANN-GA for FI prediction.

GA was separately applied on PHZ cells in each PHZ for finding the best ANN architecture. The performance of GA optimization model depends on proper selection of GA parameters. Thus, it is necessary to calibrate the GA parameters, such as crossover rate (P_c) and mutation rate (P_m). A sensitivity analysis consisting of several short runs yielded $P_c = 0.7$ and $P_m = 0.3$. Therefore, the GA was run for 100 iterations to determine the optimum network architecture including the number of hidden layers (0 to 2) and the number of neurons (0 to 10) in each hidden layer in each study area. ANN training runs for PHZs obtained the average optimal ANN architecture as 7–5 and 4–3 for Tangrah and WGEW, respectively.

The optimized architecture for each PHZ is different from other PHZs. Therefore, the ANN was applied on PHZ cells for all PHZs that participated in the training phase. For this purpose, different splits of the existing data set were examined for the ratio of training/testing varying from 60/40 to 80/20 percent. According to the evaluation criteria (i.e., MSE, MAE, and RE), the obtained results for the 80/20 percent ratio were marginally better than other cases. Then, the trained and tested ANN was used for FI prediction in PHZs which have a common cluster number and could participate in the prediction phase. For instance, in Tangrah, PHZ 1 was selected for training and PHZs 5, 9, and 13 were used for prediction (Table 6).

The overall ANN results for the 80/20 percent ratio along with the corresponding statistical criteria are presented in Table 8.

As there are several PHZs in each cluster, the PHZ averages of the performance criteria are presented in Table 8. Although FI values used in the training phase are subject to uncertainty of model calibration and input data, overall results indicate a good performance. Taking the values of MSE, MAE, and RE into account, the results indicate that FI values were accurately predicted via ANN, especially in the prediction phase in clusters of

Table 7 | Characteristics of PHZs in WGEW

PHZ No.	PHZ description	No. of cells	FI (cm/km^2)	Rank (flood generation)	Used in phase
1	Sub1_cluster3	19	0.814	12	Prediction
2	Sub1_cluster4	239	1.006	8	Training and prediction
3	Sub1_cluster5	58	1.335	5	Prediction
4	Sub2_cluster2	12	0.335	14	Prediction
5	Sub2_cluster3	155	0.158	18	Prediction
6	Sub2_cluster4	170	0.152	19	Prediction
7	Sub2_cluster5	294	0.520	13	Training and prediction
8	Sub3_cluster1	3	10.152	1	Prediction
9	Sub3_cluster2	119	3.475	3	Prediction
10	Sub3_cluster3	226	1.176	7	Training and prediction
11	Sub3_cluster4	570	0.839	11	Training and prediction
12	Sub3_cluster5	127	2.005	4	Training and prediction
13	Sub4_cluster2	581	0.959	9	Training and prediction
14	Sub4_cluster3	337	0.244	15	Training and prediction
15	Sub4_cluster4	44	0.223	16	Prediction
16	Sub4_cluster5	68	0.904	10	Prediction
17	Sub5_cluster1	137	7.291	2	Training
18	Sub5_cluster2	385	1.270	6	Training and prediction
19	Sub5_cluster3	432	0.165	17	Training and prediction
20	Sub5_cluster4	28	0.152	20	Prediction

Table 8 | Results of ANN application in the study areas

	SOMFCM			ANN					
							Prediction		
Watershed	Cluster	No. of cells	UFR Mean FI (cm/km^2)	Input variables	Training Mean MSE (cm/km^2)^2	Test Mean MSE (cm/km^2)^2	Mean MSE (cm/km^2)^2	Mean MAE (cm/km^2)	Mean RE (%)
Tangrah	1	1,941	0.0021	DEM; TPI; Slope; P50	0	0	0.01	0.04	34.1
	2	1,134	0.0008		0	0	0	0.002	25
	3	1,903	0.04		0	0	0	0.02	25.8
	4	939	0.17		0.01	0.005	0.02	0.06	18.9
	5	1,463	0.47		0.01	0.04	0	0.06	15.5
WGEW	1	140	7.35	DEM; NDVI; TWI; P50	0.59	1.58	1.32	1.51	3.84
	2	1,097	1.33		0.27	0.41	1.04	0.76	23.5
	3	1,169	0.39		0	0	0.04	0.11	40.7
	4	1,051	0.72		0.01	0.02	0.1	0.16	35.2
	5	547	0.99		0.96	1.07	2.14	1.04	55.2

high FI (i.e., cluster 5 in Tangrah and cluster 1 in WGEW).

Table 8 also reveals that the values of MSE, MAE, and RE in the prediction phase for Tangrah were generally better than those of the WGEW. One reason may be use of more input data in Tangrah compared to WGEW (i.e., 7,380 cells versus 4,004 cells). Moreover, mean RE for all PHZs was calculated as 24.9 and 35.2% for

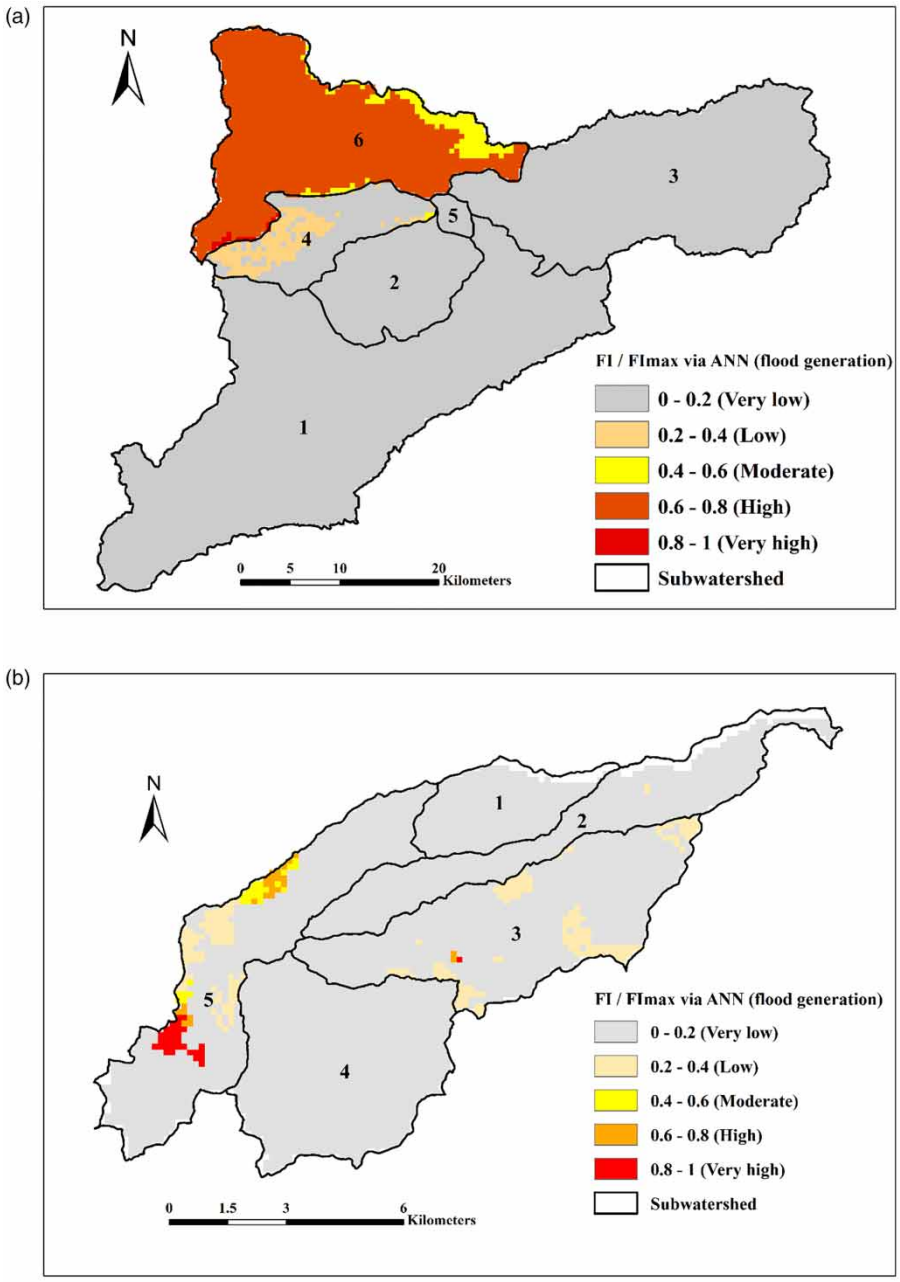

Figure 8 | (a) Map of predicted FI/FImax via ANN for (a) Tangrah and (b) WGEW.

Tangrah and WGEW, respectively. On the other hand, mean RE for PHZ cells corresponding to clusters of the two highest flood potential areas were better than others in both study areas. This indicates that the predicted FI values have acceptable accuracy in high flood clusters, in spite of using the FI values (obtained from UFR) in the ANN training phase.

FSAs predicted versus simulated maps

Figures 8(a) and 8(b) show the distribution of predicted FI for all PHZ cells in both study areas.

Comparison between Figures 8(a) and 6(b) (Tangrah) and Figures 8(b) and 7(b) (WGEW) indicates that the ANN properly predicted FI in most flood potential classes,

especially in higher classes. Also, there is a similarity between the map of FI simulated via UFR and the map of FI predicted via ANN in both study areas.

Most parts of both study watersheds have low FI and, consequently, low effect on the peak flow at the main outlet, because most areas in both watersheds consist of plains receiving lower rainfall depths and subject to high infiltration. However, maximum FI values may also be observed in downstream areas closer to the main outlet. In such areas, there is a slight difference between simulated FI (UFR) and predicted FI (ANN).

The lower the FI value, the less the contribution of the cell to the peak flood at the outlet. Mean of FI in Tangrah and in WGEW were determined as 0.127 and 1.063 (cm/km^2) via the UFR approach, respectively, while the unit cell area in Tangrah is almost 11 times that of the WGEW. However, the FI values of WGEW are much greater than those of Tangrah, because WGEW produces high runoff generated by summer thunderstorms over a small area.

UFR versus ANN in FSAs' prioritization

Figures 9(a) and 10(a) show the average FI for PHZ cells. Moreover, aggregation of the FI values resulting from PHZ cells yields the average FI for each subwatershed (Figures 9(b) and 10(b)).

As evident in Figures 9 and 10, there is high agreement between the simulated (UFR) and predicted (ANN) FI values in both watersheds. Moreover, as Figures 9(a) and 10(a) show, although there are a few exceptions, the predicted FI via ANN are almost equal to those of UFR, especially for PHZs having high FI (i.e., clusters 4 and 5 in Tangrah and clusters 1 and 2 in WGEW). On the other hand, the results of ANN FI prediction are relatively poor in a few PHZs, which correspond to the clusters of low or very low flood potential classes; thus, they are of little importance in flood control planning.

In terms of flood potential prioritization in all PHZs (Figures 9(a) and 10(a)), it may be said that the methodology has been successful in 90% and 80% of Tangrah and WGEW areas, respectively. This means that among 20 PHZs in both watersheds, 18 and 16 PHZs have been correctly prioritized in Tangrah and WGEW, respectively. In addition, the accuracy of PHZs' prioritization corresponding to rank 1 and 2 clusters was 100%. This indicates that ANN outcome could be transferable to each clustered HHR in order to predict FI in an adjacent ungauged subwatershed having common clusters (HHRs) with gauged subwatersheds. As a result, the ANN could correctly prioritize HHRs (at least ranks of 1 and 2) in an ungauged subwatershed.

FI values resulting from PHZ cells were aggregated for prioritization at subwatershed scale (Figures 9(b) and 10(b)). Generally, the ANN could correctly prioritize most subwatersheds in both study areas except WGEW subwatershed 1 ranking third in FI. Moreover, the ANN overestimated the FI in most subwatersheds in both watersheds at cell scale, with the exception of subwatershed 4 in Tangrah and subwatershed 5 in WGEW.

No correlation was found between FI and any of the ANN input variables in Tangrah and WGEW at both PHZ and cell scales. In Tangrah, the mean R^2 values were calculated as 0.14 and 0.07 at PHZ and cell scales, respectively. Similarly, in WGEW, the mean R^2 values were determined as 0.13 and 0.05 at PHZ and cell scales, respectively. As a result, it is not possible to predict FI based on only one variable via ANN at cell or PHZ scales in either study area.

SUMMARY AND CONCLUSIONS

In this study, the UFR approach was carried out through successive implementations to evaluate the effect of watershed cells on the peak flow at the main outlet of two semi-arid watersheds. The result was the FI cellular or FSA map. On the other hand, different physiographic features along with 50-year rainfall were selected to be used in the SOMFCM clustering technique as well as in the ANN-GA model. The most effective variables that contributed to SOMFCM were identified through FA. As a result, the clustered HHRs' map was prepared to facilitate ANN in the training phase. Also, ANN architecture was optimized with GA. Major conclusions are as follows:

- Among various geomorphological and hydrological inputs impacting non-linear and complex rainfall-runoff processes, DEM, TPI, and slope (for Tangrah) and DEM, NDVI, and TWI (for WGEW) along with P50

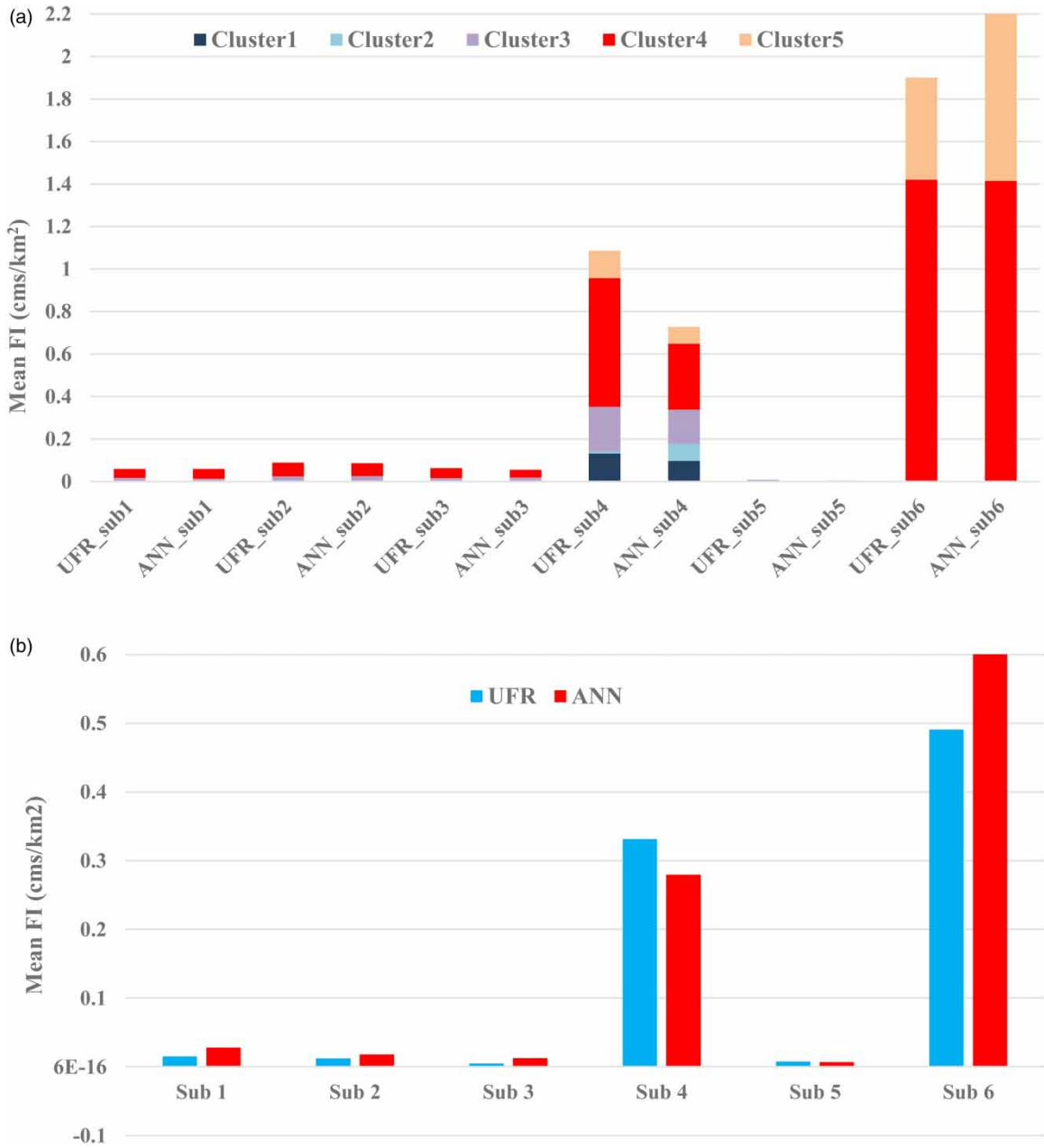

Figure 9 | Comparison of FI for Tangrah derived by UFR and ANN for (a) clusters (PHZs) and (b) subwatersheds.

were dominant factors in identifying the FSAs in semi-arid regions. In other words, input variables used in the UFR approach are the most effective factors for FI prediction via ANN at cell scale.

- As expected in semi-arid watersheds, most parts of the study areas do not significantly contribute to the flood reaching the outlet. However, in both study areas, the spatial pattern of the FI map resulting from UFR was similar to that of the ANN model. Moreover, the locations of critical FSAs with maximum FI values were identified almost accurately via ANN.
- Although the results of ANN in the Tangrah watershed (with larger area and coarser grid cell size) were better than those of the WGEW, overall results proved

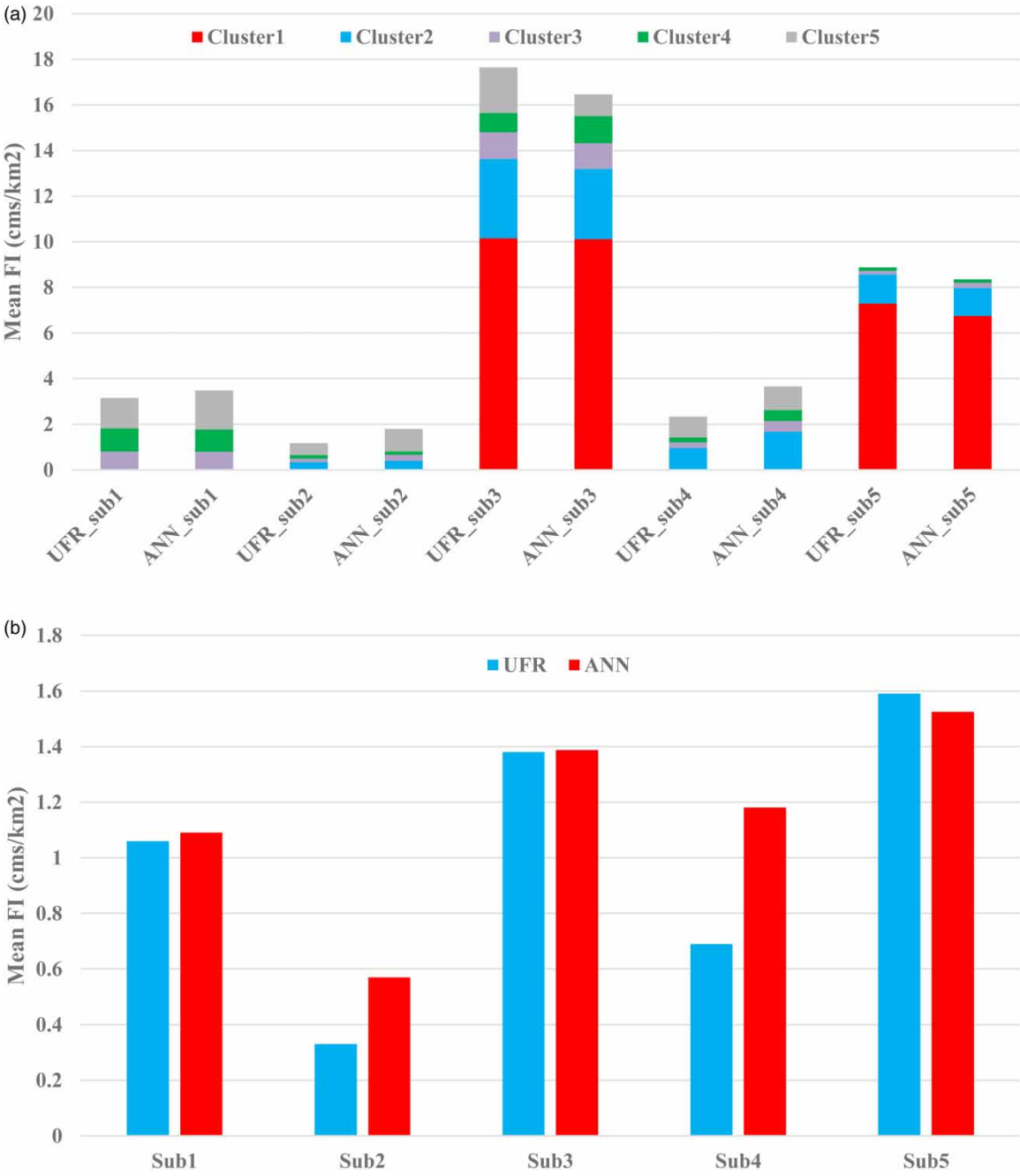

Figure 10 | Comparison of FI for WGEW derived by UFR and ANN for (a) clusters (PHZs) and (b) subwatersheds.

transferability and effectiveness of the ANN-GA model in FI prediction in semi-arid watersheds. However, ANN validity for FI prediction in humid watersheds must be studied.

- In spite of the fact that there is a nonlinear and complex relationship between input variables and FI, the results of the ANN-GA model coupled with SOMFCM clustering were promising. Therefore, the methodology may be transferable to each of the clustered HHRs in order to predict FI in an adjacent ungauged subwatershed having a common cluster (or HHR) with the other gauged subwatersheds.

- Since no correlation or direct relationship was found between FI and any of the input variables participating in prediction via ANN, it is not possible to predict FI based on only one variable at cell or PHZ scales in either study area. However, the methodology was capable of correctly prioritizing high-rank HHRs (i.e., ranks 1 and 2) in ungauged subwatersheds.
- One should not expect that the methodology predicts FI with high accuracy for all cells, because FI values used in the training phase are subject to uncertainty in model calibration and input data. However, the use of four input variables via ANN-GA had acceptable accuracy for FI prediction in HHRs having high priority in flood generation of gauged and ungauged watersheds.

REFERENCES

Abrahart, R. J., Kneale, P. E. & See, L. M. 2004 *Neural Networks for Hydrological Modelling*. Taylor & Francis, London, UK.

Araghinejad, S. 2014 *Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering*. Water, Science and Technology Library vol. 67. Springer, Dordrecht, The Netherlands. https://doi.org/10.1007/978-94-007-7506-0

Awchi, T. A. 2014 River discharges forecasting in northern Iraq using different ANN techniques. *Water Resources Management* **28** (3), 801–814.

Awchi, T. A. & Srivastava, D. K. 2009 Analysis of drought and storage for Mula project using ANN and stochastic generation models. *Hydrology Research* **40** (1), 79–91. doi:10.2166/nh.2009.012.

Babaei, M., Moeini, R. & Ehsanzadeh, E. 2019 Artificial neural network and support vector machine models for inflow prediction of dam reservoir (case study: Zayandehroud Dam Reservoir). *Water Resources Management* **33**, 2203. https://doi.org/10.1007/s11269-019-02252-5.

Basu, B. & Srinivas, V. V. 2015 Analytical approach to quantile estimation in regional frequency analysis based on fuzzy framework. *Journal of Hydrology* **524**, 30–43.

Bezdek, J. C. 1974 Cluster validity with fuzzy sets. *Journal of Cybernetics* **3**, 58–74.

Bezdek, J. C. 1981 *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, USA.

Bharti, B., Pandey, A., Tripathi, S. K. & Kumar, D. 2017 Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models. *Hydrology Research* **48** (6), 1489–1507. https://doi.org/10.2166/nh.2017.153.

Bozorg-Haddad, O., Zarezadeh-Mehrizi, M., Abdi-Dehkordi, M., Loáiciga, H. A. & Mariño, M. A. 2016 A self-tuning ANN model for simulation and forecasting of surface flows. *Water Resources Management* **30**, 2907. https://doi.org/10.1007/s11269-016-1301-2.

Burrough, P. A., van Gaans, P. F. M. & Hootsmans, R. 1997 Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma* **77**, 115–135.

Foda, R. F., Awadallah, A. G. & Gad, M. A. 2017 A fast semi distributed rainfall runoff model for engineering applications in arid and semi-arid regions. *Water Resources Management* **31** (15), 4941–4955.

Fukuyama, Y. & Sugeno, M. 1989 A new method of choosing the number of clusters for the fuzzy c-means method. In: *Proceedings of the Fifth Fuzzy Systems Symposium*, Ankara, Turkey, pp. 247–250.

Golian, S., Saghafian, B. & Maknoon, R. 2010 Derivation of probabilistic thresholds of spatially distributed rainfall for flood forecasting. *Water Resources Management* **24** (13), 3547–3559.

Govindaraju, R. S. 2000 Artificial neural network in hydrology. I: preliminary concepts. *Journal of Hydrologic Engineering* **5** (2), 115–123.

He, Z., Zhao, W. & Liu, H. 2014 Comparing the performance of empirical black-box models for river flow forecasting in the Heihe River Basin, Northwestern China. *Hydrological Processes* **28**, 1–7.

Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbor, MI, USA.

Kamp, R. G. & Savenije, H. H. G. 2006 Optimising training data for ANNs with genetic algorithms. *Hydrology and Earth System Sciences Discussions* **3**, 285–297.

Khazaei, M. R., Zahabiyoun, B., Saghafian, B. & Ahmadi, S. 2014 Development of an automatic calibration tool using genetic algorithm for the ARNO conceptual rainfall-runoff model. *Arabian Journal for Science and Engineering* **39** (4), 2535–2549.

Kisi, O. 2005 Daily river flow forecasting using artificial neural networks and auto regressive models. *Turkish Journal of Engineering and Environmental Sciences* **29**, 9–20.

Kişi, O., Nia, A. M., Gosheh, M. G., Tajabadi, M. R. & Ahmadi, A. 2012 Intermittent streamflow forecasting by using several data driven techniques. *Water Resources Management* **26**, 457–474. doi:10.1007/s11269-0119926-7.

Kohonen, T. 1982 Self-organized formation of topologically correct feature maps. *Biological Cybernetics* **43**, 59–69.

Kourgialas, N. N. & Karatzas, G. P. 2017 A national scale flood hazard mapping methodology: the case of Greece–Protection and adaptation policy approaches. *Science of the Total Environment* **601**, 441–452.

Kull, D. W. & Feldman, A. D. 1998 Evaluation of Clark's unit graph method to spatially distributed runoff. *ASCE Journal of Hydrologic Engineering* **3** (1), 9–19. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:1(89).

Kwon, S. H. 1998 Cluster validity index for fuzzy clustering. *Electronics Letters* **34** (22), 176–177.

Mwale, F. D., Adeloye, A. J. & Rustum, R. 2014 Application of self-organising maps and multi-layer perceptron-artificial neural

networks for streamflow and water level forecasting in data-poor catchments: the case of the Lower Shire floodplain, Malawi. *Hydrology Research* **45** (6), 838–854. https://doi.org/10.2166/nh.2014.168.

Nadoushani, S. S. M., Dehghanian, N. & Saghafian, B. 2018 A fuzzy hybrid clustering method for identifying hydrologic homogeneous regions. *Journal of Hydroinformatics* **20** (6), 1367–1386. https://doi.org/10.2166/hydro.2018.004.

Nourani, V., Alami, M. T. & Daneshvar, F. 2015 Self-organizing map clustering technique for ANN-based spatiotemporal modeling of groundwater quality parameters. *Journal of Hydroinformatics* **18** (2), 288–309.

Pal, N. R. & Bezdek, J. C. 1995 On cluster validity for fuzzy c-means model. *IEEE Transactions on Fuzzy Systems* **3**, 370–379.

Rao, A. R. & Srinivas, V. V. 2006 Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology* **318** (1–4), 57–79.

Rezaeianzadeh, M., Tabari, H., Yazdi, A. A., Isik, S. & Kalin, L. 2014 Flood flow forecasting using ANN, ANFIS and regression models. *Neural Computing & Applications* **25** (1), 25–37. https://doi.org/10.1007/s00521-013-1443-6.

Rousseeuw, P. J. 1987 Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53–65.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986 Learning representations by back-propagating errors. *Nature* **323**, 533–536.

Saadat, H., Adamowski, J., Bonnell, R., Sharifi, F., Namdar, M. & Ale-Ebrahim, S. 2011 Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **66** (5), 608–619.

Saghafian, B. & Khosroshahi, M. 2005 Unit response approach for priority determination of flood source areas. *ASCE Journal of Hydrologic Engineering* **25**, 270–277. https://doi.org/10.1061/(ASCE)1084-0699(2005)10:4(270).

Saghafian, B., Ghermezcheshmeh, B. & Kheirkhah, M. M. 2010 Iso-flood severity mapping: a new tool for distributed flood source identification. *Natural Hazards* **55**, 557–570.

Saghafian, B., Sima, S., Sadeghi, S. & Jeirani, F. 2012 Application of unit response approach for spatial prioritization of runoff and sediment sources. *Agricultural Water Management* **109**, 36–45.

Saghafian, B., Golian, S. & Ghasemi, A. 2014 Flood frequency analysis based on simulated peak discharges. *Natural Hazards* **71**, 403–417. https://doi.org/10.1007/s11069-013-0925-2.

Saghafian, B., Meghdadi, A. & Sima, S. 2015 Application of the WEPP model to determine sources of run-off and sediment in a forested watershed. *Hydrologic Processes* **29**, 481–497.

Sanyal, J., Densmore, A. L. & Caronneau, P. 2014 Analyzing the effect of land-use/cover changes at sub-catchment levels on downstream flood peaks: a semi-distributed modelling approach with sparse data. *Catena* **118**, 28–40. https://doi.org/10.1016/j.catena.2018.09.001.

Shamseldin, A. Y. 1997 Application of a neural network technique to rainfall-runoff modeling. *Journal of Hydrology* **199**, 272–294.

Singh, K. K. & Singh, A. 2017 Identification of flooded area from satellite images using Hybrid Kohonen Fuzzy C-Means sigma classifier. *The Egyptian Journal of Remote Sensing and Space Science* **20**, 147–155. https://doi.org/10.1016/j.ejrs.2016.04.003.

Srinivas, V. V., Tripathi, S., Rao, A. R. & Govindaraju, R. S. 2008 Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. *Journal of Hydrology* **348**, 148–166.

Sudheer, K. P., Gosain, A. K. & Ramasastri, K. S. 2002 A data-driven algorithm for constructing artificial neural network rainfall–runoff models. *Hydrologic Processes* **16**, 1325–1330.

USDA, United States Department of Agriculture 2007 Agricultural Research Service, Southwest Watershed Research Center, Tucson, AZ, USA. http://www.tucson.ars.ag.gov/swrc_site/research/wgew (accessed August 17 2018).

Valipour, M., Banihabib, M. E. & Behbahani, S. M. R. 2013 Comparison of the ARMA, ARIMA, and the autoregressive artificial neural. *Journal of Hydrology* **476**, 433–441.

Weiss, A. 2001 Topographic Position and Landforms Analysis. In *ESRI User Conference*, San Diego, CA, USA.

First received 21 August 2019; accepted in revised form 6 January 2020. Available online 7 February 2020