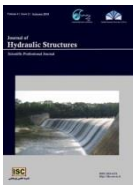
Journal of Hydraulic Structures
J. Hydraul. Struct., 2020; 6(4):75-98
DOI: 10.22055/jhs.2021.36629.1163



GIS-Based Flood Risk Zoning Based On Data-Driven Models

Seyed Ahmad Eslaminezhad [1]
Mobin Eftekhari[2]
Mohammad Akbari [3]

Abstract

Increasing the occurrence of floods, especially in cities, and the risks to human, financial, and environmental risks due to its, make flood risk zoning of great importance. The purpose of this study is to estimate the flood risk of the Maneh and Samalghan based on determining effective criteria and spatial and non-spatial data-driven models. The criteria used in this research include Modified Fournier Index, Topographic Position Index, Curve Number, Flow Accumulation, Slope, Digital elevation model, Topographic Wetness Index, Vertical Overland Flow Distance, Horizontal Overland Flow Distance, and Normalized difference vege-tation index. The novelty of this study is to present new combination approaches to determine the effective criteria in flood risk zoning (Maneh and Samalghan). In this regard, the geographically weighted regression (GWR) with exponential and bi-square kernels and artificial neural network (ANN) combined with a binary particle swarm optimization algorithm (BPSO). The best value of the fitness function (1-R2) for ANN, GWR with the exponential kernel, and GWR with bi-square kernel was obtained 0.1757, 0.0461, and 0.0097, respectively, which indicates higher compatibility of the bi-square kernel than the other models. It was also found that the criteria used have a significant effect on the rate of flooding in the study area.

Keywords: Flood Risk; Geographically Weighted Regression; Artificial Neural Network; Binary Particle Swarm Optimization Algorithm.

Received: 12 February 2021; Accepted: 21 March 2021

[1] Department of surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran, Iran; Ahmad.eslami73@ut.ac.ir
[2] Young Researchers and Elite Club, Mashhad Branch, Islamic Azad University, Mashhad, Iran; mobineftekhari@yahoo.com **(Corresponding Author)**
[3] Department of Civil Engineering, University of Birjand, Birjand, Iran; moakbari@birjand.ac.ir

1. Introduction

Flood is fact the increase of the water level of the river and the channel and the water coming out of it and occupying a part of the plains along the river, which can cause general damage by flooding the area and cause human and non-human losses [1],[2]. Flood risk zoning is one of the critical steps in river management. One way to prepare a flood zoning map is to use a spatial information system and integrate it with hydraulic models. Flood zoning maps are widely used in flood plain management studies. Applications of these maps include determining the bed and area of rivers, studying and economic justification of development plans, flood forecasting and warning, rescue operations and flood insurance [3],[4],[5]. The level of flood-prone areas of Iran is estimated at 91 million hectares. In other words, 55% of the country is involved in the production of direct and rapid runoff, of which about 22 million hectares have moderate to very high flooding intensity [6],[7]. Therefore, Iran is one of the flood-prone regions of the world and one of the ways to reduce the damage caused by floods is to prepare flood risk zoning maps to identify susceptible areas. Also, the preparation of these plans can be central to development activities and other future activities. According to previous research, flood risk zoning map production can be classified into two general categories of knowledge-driven and data-driven models [8],[9]. Data-driven models are highly effective in known areas or areas where the number of known evidence is statistically sufficient (references data). In these models, the purpose is to identify new locations for more detailed work, while in knowledge-driven models, they are effective in less known environments or where there are few targets in the area. Weight estimates and class estimates are based on expert judgment and do not require evidence of an answer [8],[9]. Numerous studies have been conducted on flood risk zoning with approaches based on knowledge-driven and data-driven integration models, to name a few:

Elsheikh et al. [10] used Geographic Information System (GIS) to evaluate areas with potential flood risk in Terengganu. They calculated the importance of each layer (Annual rainfall, basin slope, drainage network, and soil type) using the Analytic hierarchy process (AHP) method. The percentage related to the criteria affecting annual rainfall was 38.7%, drainage network was 27.5%, river basin slope was 19.8% and soil type was 14%, and finally, they prepared a flood risk map. Papaioannou et al. (2015) used GIS, fuzzy logic, and multi-criteria assessment methods to represent flood areas. They divided the criteria used into independent groups. They also used GIS and numerical simulation to collect and process geographic data. Finally, with multi-criteria analysis (MCA), including the AHP and fuzzy processes, they created the zoning map of flood vulnerable areas. Xiao et al. [11] proposed a multi-criteria analysis (MCA) framework using the fuzzy analytic hierarchy process (FAHP), and the ordered weighted averaging method (OWA) for flood risk assessment. The relative importance of the criteria was determined by the fuzzy AHP method. The OWA method was also used to analyze the effects of different decision-maker risk attitudes on the evaluation results. Al-Juaidi et al. [12] used a logistic regression model for flood susceptibility mapping in southern Gaza Strip areas. They used the criteria including topographic slope, Digital elevation model (DEM), rainfall, flow accumulation, soil type, and land use/land cover (LULC). The results showed that the prediction and success rates are 76 and 81%, respectively. Kanani-sadat et al. [6] provided a framework for mapping flood-prone areas of Kurdistan province by integrating GIS, fuzzy logic, and multi-criteria decision making. To achieve this purpose, a set of geophysical, geomorphological, meteorological, hydrological, and geographical criteria were examined. Then, considering the uncertainty and ambiguity of the experts' opinions, they merged the fuzzy theory with DEMATEL. Finally, the analytic network process (ANP) was used to

calculate the final weight of each criterion. To assess the flood vulnerability map in the study area, the flood reference map in Kurdistan province was used. About 85% of the validation area was classified as "very high floods", which referred to the proposed framework's efficiency. Also, the AHP method was implemented to evaluate the proposed framework's performance compared with traditional approaches. Validation results showed that the fuzzy ANP-DEMATEL model had higher performance accuracy compared to the AHP model. Ardiansyah and Sumunar [13] using weighting and scoring methods (based on the quantity or quality of the criteria affecting the flood), considered the score for each of the classes (ranges) defined as the effective and weighted criteria for the effective criteria. Finally, the weights and scores assigned to the effective criteria were entered into the overlap analysis to identify flood vulnerabilities. The results of this study showed that there are three levels of flood vulnerability, namely low, medium, and high vulnerability in the study area.

Flood risk zoning map production is a topic that has received a lot of attention so far, but among the studies conducted, some points have received less attention; first, none of these studies provide an adequate combination of criteria for flood risk zoning. Second, proper analysis has not been used to determine the optimal combination of effective criteria and to prepare a flood risk zoning map based on the effective criteria. In this study to recognizing the vulnerability of Maneh and Samalghan to flood risk, due to the availability of flood reference map in the region, the combination of spatial and non-spatial data-driven models including artificial neural network (ANN) and geographically weighted regression (GWR) with binary particle swarm optimization (BPSO) were used to prepare a flood risk zoning map based on determining the optimal combination of effective criteria.

2. Materials and Methods

2.1. Study Area

Maneh and Samolghan cities are located in the northwest of North Khorasan province, which in terms of area is 21% and in terms of population is 11.8% of the total population of the province (Fig. 1). The city shares an 8-kilometer border with Turkmenistan. In terms of geographical location, the city is located between 37˚17' to 38˚7' north latitude of the equator and 55˚ 59' to 57˚ 17' east longitude of the Greenwich meridian. The hydrographic network in Maneh and Samolghan includes seasonal and permanent rivers that are part of the Atrak catchment. These rivers originate from inside the heights and outside the city and flow along the general slope. Generally, part of the precipitation on the morphological features flows in surface water, which accumulates in these rivers. The source of most of these rivers is mountainous and with the melting of snow, the river is fed and strengthened. Also, Fig. 2 shows the flowchart of the proposed model in this study.

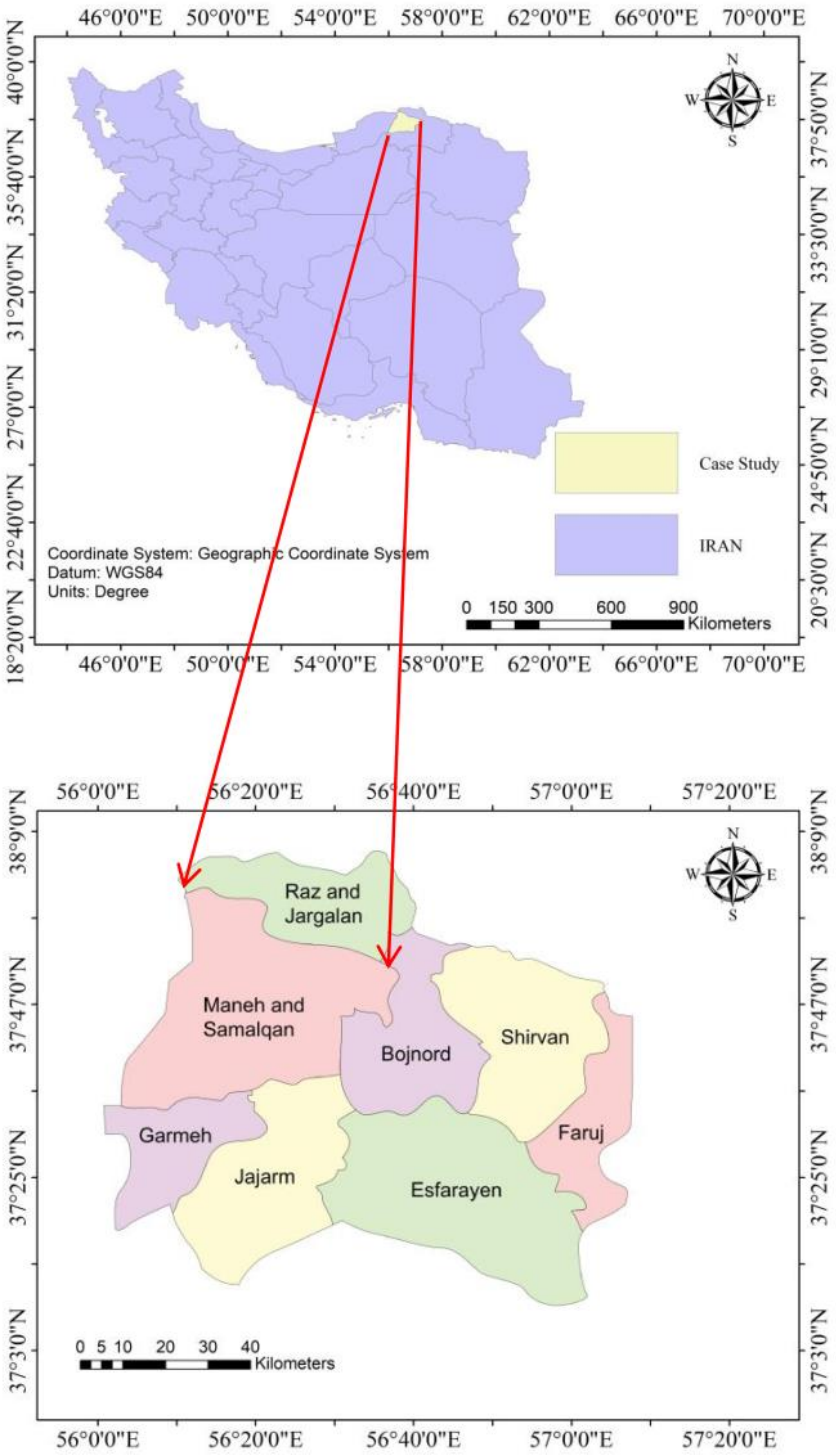


Figure 1. Geographical location of the study area

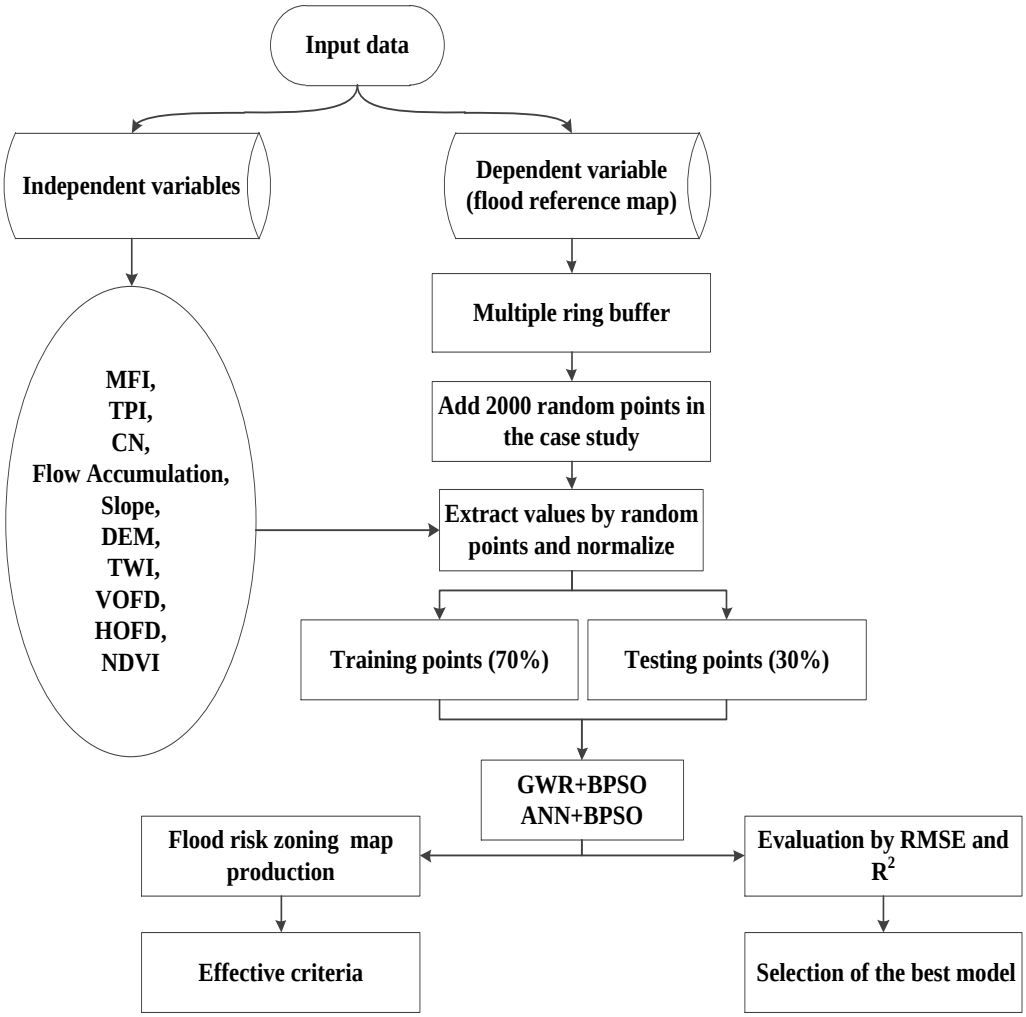


Figure 2. Flowchart showing the methodology used in this study

2.2 Spatial criteria affecting floods

Flood zoning depends on various criteria that the following criteria have been used in this study:

2.2.1. Horizontal Overland Flow Distance (HOFD)

The horizontal overland flow distance is the actual movement of water from one cell to another. The shorter the horizontal distance from the stream, the greater the potential for flooding [14].

2.2.2. Vertical Overland Flow Distance (VOFD)

The vertical overland flow distance indicates the vertical distance between each cell's height and the height calculated for the flow network. The shorter the vertical distance from the stream, the greater the potential for flooding [4],[6].

2.2.3. Normalized Difference Vegetation Index (NDVI)

This index estimates the vegetation criteria. The lower the value, the greater the risk of flooding, which is obtained from Eq. 1 [15]:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

Where NIR is the near-infrared band and R is the red band of satellite imagery.

2.2.4. Topographic Wetness Index (TPI)

A Topographic moisture index is a useful and common tool to describe the humidity conditions at the basin scale. The lower the value of this criterion, the greater the risk of flooding, which is obtained according to Eq. 2 [16]:

$$TWI = \ln\left(\frac{A_s}{\tan\beta}\right) \tag{2}$$

Where As is the area of the upstream sphere and is the angle of inclination.

2.2.5. Digital Elevation Model (DEM)

The digital elevation model includes the values representing the height of the ground. The lower the criterion, the greater the risk of flooding [1],[3].

2.2.6. Slope

The presence of steep slopes indicates that the area in question has less flat land, which can cause devastating floods in the event of heavy rains, which are obtained according to Eq. 3 and Fig. 3 [1]:

$$Slope = a\tan\left(\sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}\right)$$
$$[dz/dx] = ((c + 2f + i) - (a + 2d + g)) / 8 \tag{3}$$
$$[dz/dy] = ((g + 2h + i) - (a + 2b + c)) / 8$$

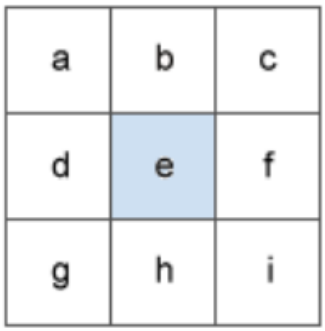


Figure 3. Sample cells to calculate the slope [1]

2.2.7. Flow Accumulation

The flow aggregation network contains the cumulative number of cells upstream of a cell and shows the amount of current flowing from the upstream cells to that cell. According to Fig. 4, the higher the value of this criterion, the greater the potential for flooding [17].

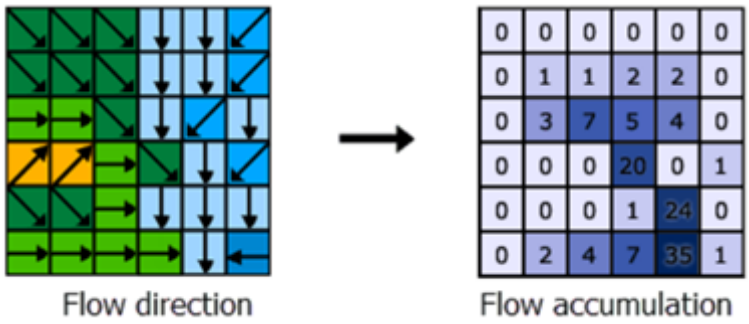


Figure 4. Sample cells to calculate the flow accumulation [17]

2.2.8. Curve Number (CN)

Curve number defines the runoff potential of the basin. The higher the value of this criterion, the greater the risk of flooding. This criterion is derived from Eq. 4 [18]:

Where R is the height of the direct runoff and P is the height of the rainfall.

2.2.9. Topographic Position Index (TPI)

A topographic position index shows the difference between the height of each cell and the average height of its neighboring cells. The lower the value of this criterion, the greater the risk of flooding, which is obtained according to Eq. 5 [19]:

$$TPI = Z_0 - \frac{\sum_{i=1}^{n} Z_i}{n} \quad (5)$$

Where Z_0 is the height of the desired cell and $\frac{\sum_{i=1}^{n} Z_i}{n}$ is the average height of neighboring cells.

2.2.10. Modified Fournier Index (MFI)

The rainfall intensity map is created using the modified Fournier index method. The higher the value of this criterion, the greater the risk of flooding, which is obtained according to Eq. 6 [20]:

$$MFI = \frac{\sum_{i=1}^{12} p_i^{\,2}}{p} \quad (6)$$

Where Pi is the average annual rainfall for the first month, and p is the av-erage annual rainfall.

2.3. Artificial neural network

Artificial neural networks are one of the computational methods inspired by the neural system of the human brain. One of the remarkable characteristics of this type of network is their ability to learn and the ability to generalize this learning, because of this feature, they make it possible to learn to understand patterns [21]. The most important advantage of artificial neural networks over regression methods for modeling a pattern is that there is no need for an initial model in linking input and output data [22]. Based on the intrinsic relationships between data, a linear or nonlinear model is established between independent and dependent variables.

In this study, a multilayer perceptron neural network has been used to model flood risk. This type of neural network consists of a set of neurons arranged in different layers in a row. The law of multilayer perceptron learning is called the error propagation rule, which is used to estimate unknown network parameters. The multilayer perceptron works in such a way that a pattern is supplied to the network and its output is calculated. Actual output values and desired output cause the network coefficients to change; in such a way that a more accurate output is obtained in later stages. To succeed in network training, its output must be gradually brought closer to the desired output and the error rate must be re-duced. The design ANN used in this study is illustrated in Fig. 5.

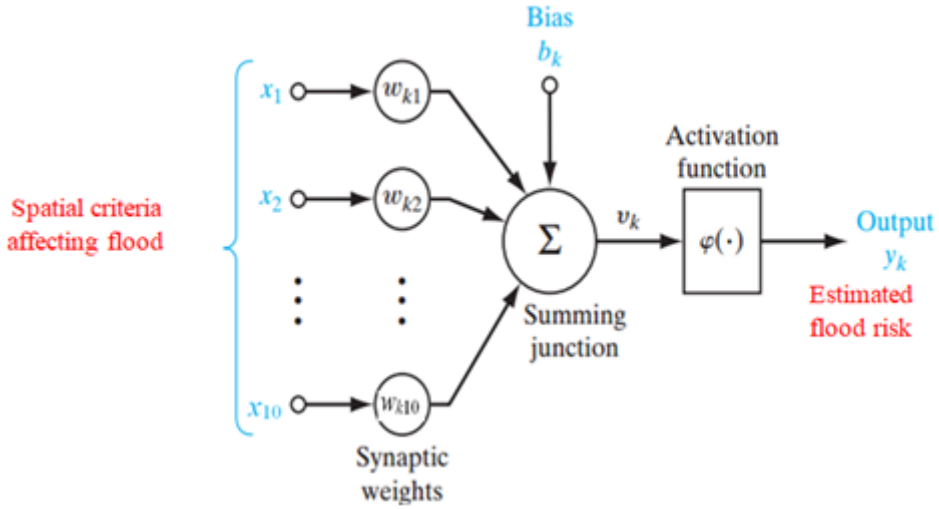


Figure 5. ANN architecture applied in this study [22]

2.4. Geographically weighted regression

According to spatial autocorrelation and spatial non-stationarity properties for spatial data, it is less possible to use basic global regressions such as Ordinary Least square [23]. In this model, the spatial dependencies between the events are considered as weight matrices, and due to the heterogeneity of the environmental factors and the existence of local variation, regression coefficients of the GWR model for observation are measured locally [24]. The equation of the GWR model is calculated as Eq. 7 [25]:

$$y_i = \sum_{j=0}^{n} \beta_j(u_i, v_i) x_j + \varepsilon_i \tag{7}$$

Where y_i is the dependent variable (Flood risk), x_j is the independent variables (Spatial criteria affecting floods), n is the total random points, ε_i is the residual GWR model, (u_i,v_i) denotes the coordinates of the ith point in space and $\beta_j(u_i,v_i)$ is the regression coefficient for coordinates of the ith point. To calculate the spatial weight matrix, it is necessary to specify the desired kernel function. According to previous research, this study used two kernels including exponential and bi-square these two kernels which are calculated as Eq. 8 to Eq. 10, respectively [25],[26]:

$$W(u_i, v_i) = \exp(-(\frac{|d_{ij}|}{b})) \tag{8}$$

$$W(u_i, v_i) = \begin{cases} (1-(\frac{d_{ij}}{b})^2)^2 & |d_{ij}| < b \\ 0 \quad otherwise \end{cases} \tag{9}$$

$$d_{ij} = \sqrt{(\delta_{ij}^x)^2 + (\delta_{ij}^y)^2} \tag{10}$$

Where d_{ij} is the Euclidean distance value between two observations i and j, δ_{ij}^{x} and δ_{ij}^{y} are the distance between two observations i and j along the horizontal and vertical axis and b is the bandwidth value. The regression coefficients are different for each location, so in the GWR model, local variation of the regression coefficients can be obtained by the standard deviation function according to Eq. 11 [24]:

$$SE = \sqrt{\sum_{i=1}^{n} \frac{(\beta_{ij} - \beta_j)^2}{n}} \tag{11}$$

Where β_{ij} is the regression coefficient for the factor j in the observation i, β_j is the mean regression coefficient of factor j in the total observations and n is the total observations.
To evaluate the ANN and GWR models output the Coefficient of Determination (R^2) is usually used to measure the goodness of fit and the RMSE value measure the residuals distribution of the observation, which are obtained based on Eq. 12 and Eq. 13 [25]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{13}$$

Where n is the total observations, y_i is the value for observation i, $\hat{y}_i$ is the estimated value for observation i and $\overline{y}$ is the mean value for total observations.

2.5. Binary particle swarm optimization

The PSO algorithm is an optimization algorithm that makes it less likely to be captured at a local minimum and can search uncertain and complex areas based on probabilistic rules [27]. Also, in this algorithm, the solution of the proposed path is not dependent on the initial population and starting from each point in the search space, the solution converges to the optimal solution [28]. After a while, Kennedy and Eberhart [29] introduced the Binary PSO algorithm, which, unlike the continuous version of it, is limited to having zero and one (binary) variables and the velocity value can change a particle from zero to one. According to the purpose of this study, The BPSO algorithm has been used. In this algorithm, Eq. 14 and Eq. 15 are used to update the velocity and position of each particle [29]:

$$V_i(t+1) = w \times V_i(t) + c_1 \times r_1 \times (pbest - X_i(t)) + c_2 \times r_2 \times (gbest - X_i(t)) \tag{14}$$

$$X_i(t+1) = \begin{cases} 1 & \rho \le sig(v_i^{t+1}) \\ 0 & Otherwise \end{cases} \tag{15}$$

Where $V_i(t)$ is the velocity of the particle i, $X_i(t)$ is the position of the particle i, $V_i(t + 1)$ is the velocity of the particle i in the next position, $X_i(t + 1)$ is the position of the particle i in the next position, *pbest* is the best position of the experience for the particle i, *gbest* is the best position experienced in all particles, c_1 is the personal learning coefficient, c_2 is the collective learning coefficient, w is the inertia weight and r_1, r_1 and ρ are random numbers in the range [0.1]. In this study, the steps of the BPSO algorithm (In combination with the ANN and GWR models) are as follows which showed in Fig. 6:

1. Give the initial value to a population of particles with random positions and velocities.
2. Training ANN and GWR models and calculating the fitness function (R^2) of each particle in this population.
3. Stop the BPSO algorithm if it reaches the stop criterion (100 Iterations), otherwise go to step 4. If the algorithm reaches the condition of stopping, then the selected criteria are the same effective parameters in flood risk zoning.
4. Determine the pbest and gbest for each particle.
5. Calculate the velocity of each particle and move to the next position based on the relations (Go to step 2).

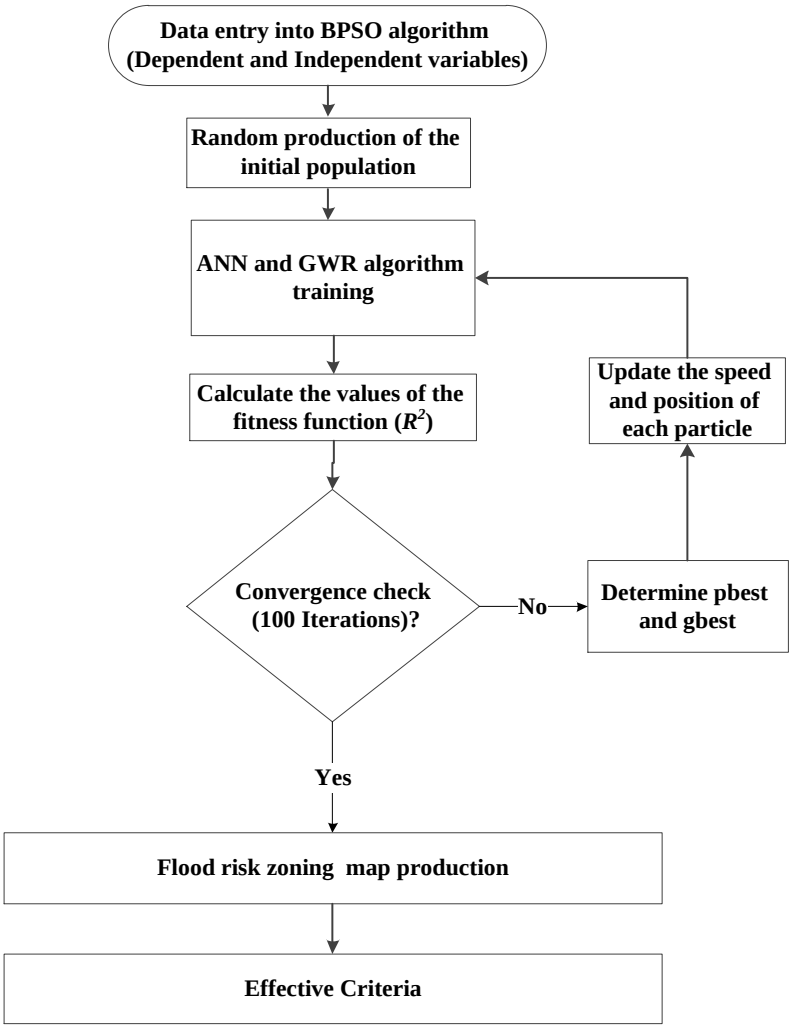


Figure 6. Calculation steps of the recommended models

3. Results and Discussion

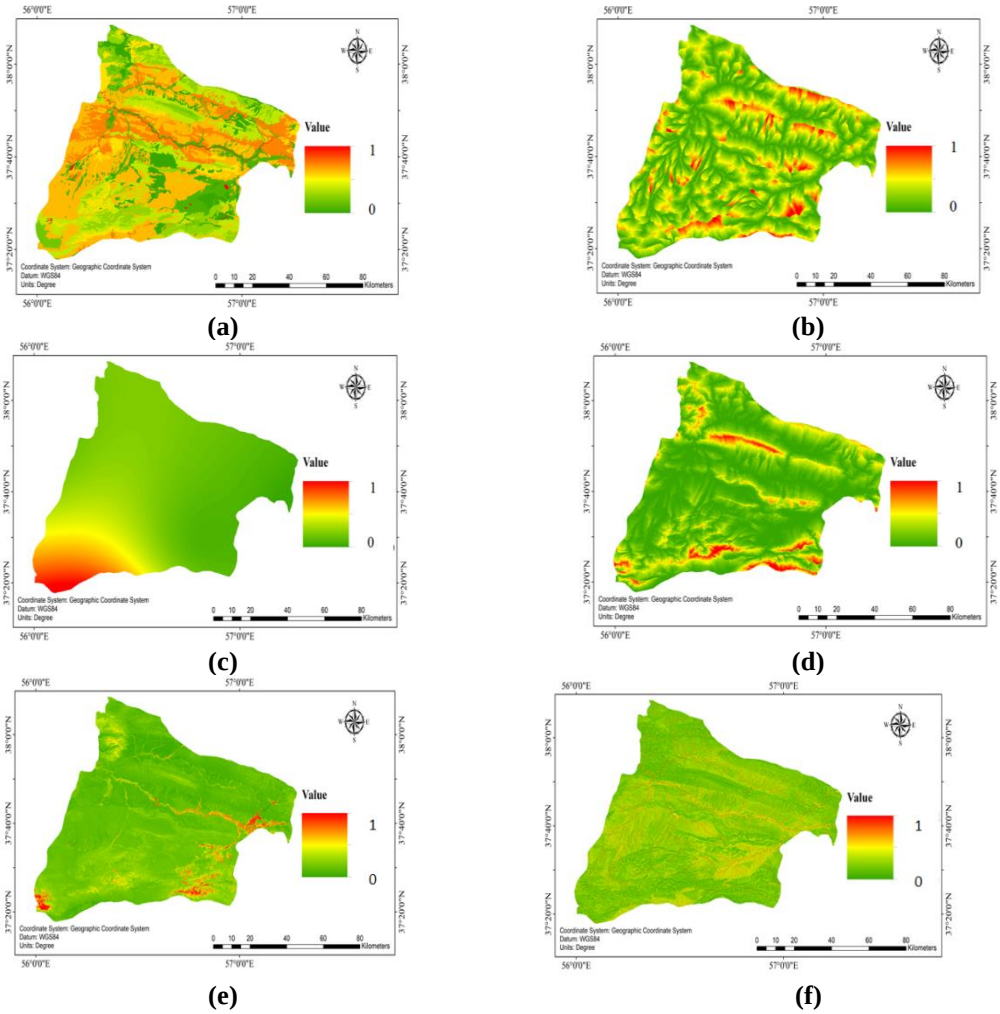
3.1. Define dependent and independent variables

According to Table 1, in this study, the spatial criteria affecting floods are considered as independent variables. These criteria, in the order mentioned, form the particle dimension of the BPSO algorithm.

Table 1. Independent variables in this study

Order	Criteria	Order	Criteria
1	Flow Accumulation	6	Topographic Position Index
2	Horizontal Overland Flow Distance	7	Topographic Wetness Index
3	Modified Fournier Index	8	Vertical Overland Flow Distance
4	Normalized difference vegetation index	9	Digital Elevation Model
5	Slope	10	Curve Number

The spatial criteria affecting floods are related to June 2017, which was ob-tained from the Forests, Range and Watershed Management Organization of Iran. Therefore, SRTM (Shuttle Radar Topography Mission) images were used to prepare the DEM layer. Bands 4 and 5 of Landsat 8 images were also used to produce the NDVI layer. Six synoptic stations were used to calculate the MFI value. The CN layer was also prepared from the Forests, Range and Watershed Management Organization of Iran. Other layers were created using appropriate analysis in Arc Map and SAGA GIS software. For implementation, each of the criteria was produced in the form of a raster map with a pixel size of 30 meters. According to Fig. 7, the maps of these criteria are shown in a normalized way.



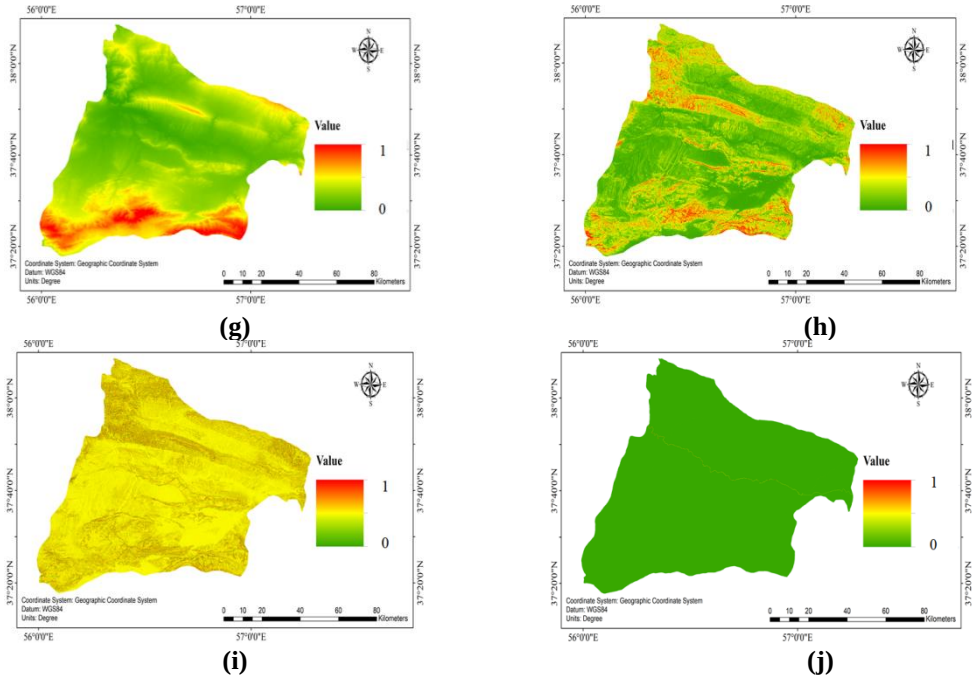
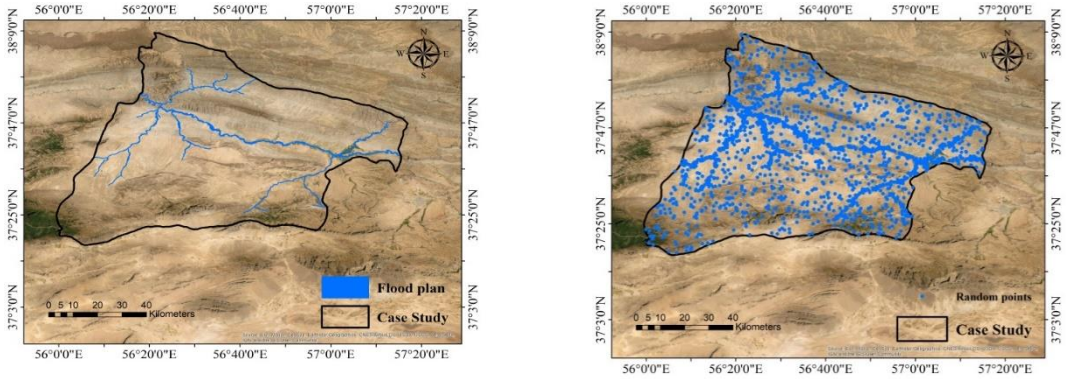


Figure 7. The maps of criteria (a) Curve number (b) Horizontal Overland Flow Distance (c) Modified Fournier Index (d) Vertical Overland Flow Distance (e) Normalized difference vegetation index (f) Topographic Wetness Index (g) Digital Elevation Model (h) Slope (i) Topographic Position Index (j) Flow Accumulation

In order to implement the proposed models, it is necessary to produce random points in the desired area. Therefore, based on the multiple ring buffer analysis, boundaries around the flood reference map (Fig. 8a) were defined. The flood is the result of the reference map related to June 2017, which has been obtained from the Forests, Range and Watershed Management Organization of the country. Then, according to Fig. 8b, 2000 points were generated randomly and uniformly in the study area. Among the 2000 random points created in the study area, 1000 points are outside the scope of the flood reference map and the defined areas around it (Table 2) and 1000 points are within the scope of the flood reference map and the defined areas around it. Then the values of all available information criteria for these points were calculated (in a normalized way).



(a) **(b)**

Figure 8. (a) The flood reference map (b) The random point created in the case study

Table 2. Define the dependent variable based on the defined multiple ring buffer analysis

Location of points	X ≤ 100 (m)	100 < X ≤ 200 (m)	200 < X ≤ 300 (m)	300 < X ≤ 400 (m)	400 < X ≤ 500 (m)	Out of buffers (500 < X) (m)
Value (No unit)	1	0.8	0.6	0.4	0.2	0

Note: X= The distance of random points from flood reference map.

According to Fig. 8b, 2000 points created in the study area by a random sampling method to extract values and use in the data-driven models [30]. Then the values of all available information criteria (independent and dependent var-iables) for these points were calculated (in a normalized way). The correlation between the criteria from Eq. 16 was examined [31]:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n.\sigma_x.\sigma_y} \tag{16}$$

Where $\bar{x}$ and $\bar{y}$ are the mean of data x and y, n is the total data, σ_x and σ_y are the deviation of data x and y. As shown in Fig. 9, the correlation between criteria is near 0. Therefore, all criteria entered the algorithms.

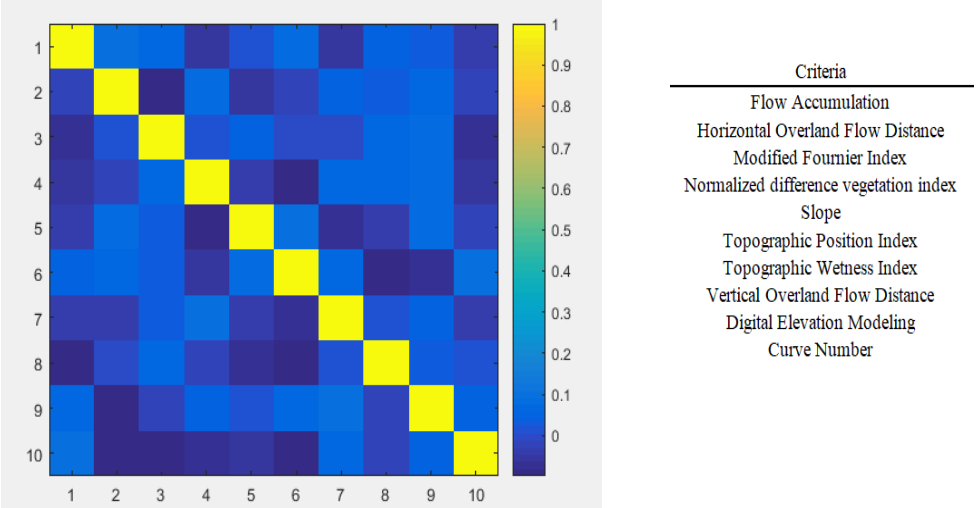


Figure 9. Correlation matrix between criteria

3.2. Implement data-driven models

For the implementation of the ANN and GWR models, 70% of the total data was used for training and 30% of the total data was used for testing, and all data were normalized before entering the algorithms. Based on previous research and trial and error method, a ratio of 70:30 was selected [32]. In article this ratio gives the best performance result. Due to the fact that one of the most important parameters for evaluating data-driven models (Model compatibility with data) is the Coefficient of Determination parameter (R^2), therefore, the BPSO algorithm fitness function has been selected to minimize the value of 1-R^2 [25]. The optimal values of the initial parameters of the BPSO algorithm were selected based on the experiments obtained from different iterations and through trial and error according to Table 3. The condition for stopping to simplify the implementation process is the number of specific executions.

Table 3. Set Parameters in the BPSO algorithm

Parameters	Value	Parameters	Value
Swarm size	30	C_2	2
Total iterations	100	W	1
C_1	2	Minimum and maximum velocity (m/s)	[-4,4]

Fig. 10 shows the swarm structure of the BPSO algorithm in this study, which the criteria mentioned in Table 1 form the particle dimension of the BPSO algorithm.

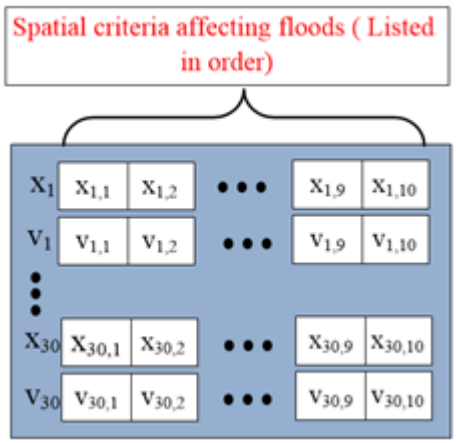
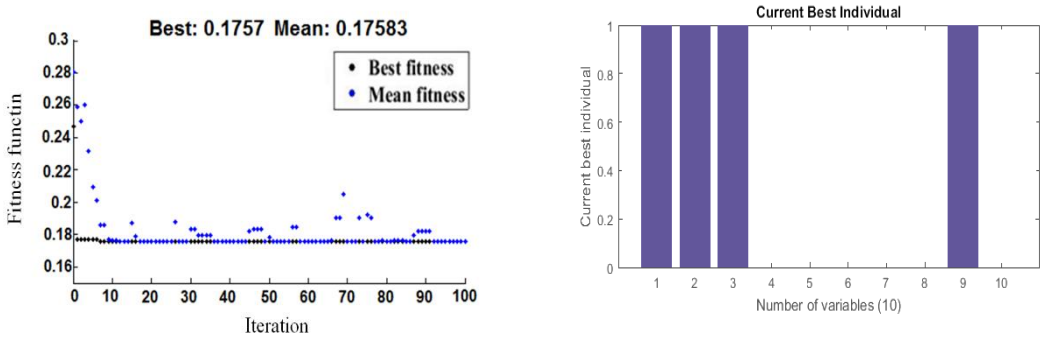


Figure 10. Swarm structure of the BPSO algorithm

Due to the random nature of the BPSO algorithm and based on previous research, this algorithm with the desired number of iterations was repeated 10 executions and the Best of these 10 executions was considered as the final output [33]. According to Fig. 11a, by performing the combination of the ANN model and the BPSO algorithm, the best value of fitness function (1-R^2) was obtained 0.1757 (the best of 10 executions). Also, According to Fig. 11b for the ANN model, four criteria of Flow Accumulation, Horizontal Overland Flow Distance, Modified Fournier Index and Digital elevation model were determined as effective criteria in estimating flood risk. In fact, Fig. 11b shows the best particle in terms of fitness function (R^2) among all particles (30 particles) in the 100th iteration of the BPSO algorithm.

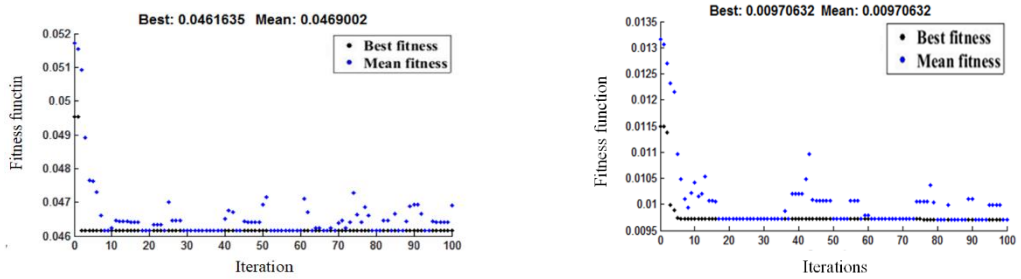


(a) **(b)**

Figure 11. (a) The best value of fitness function by combining of ANN model and BPSO algorithm (b) Effective criteria in estimating flood risk by combining of ANN model and BPSO algorithm

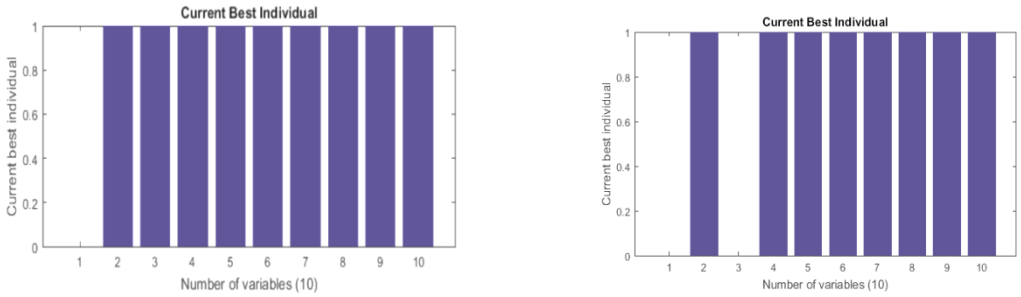
Then, the GWR model with two exponential and bi-square kernels and BPSO algorithm was combined to determine the effective criteria in estimating flood risk. To implement the GWR model, the random points' coordinates were used as inputs in its weight matrix. According to Fig. 12, the best value of fitness function (1-R2) for combination the GWR model with two exponential and bi-square kernels and the BPSO algorithm, was obtained 0.0461 and 0.0097 (the best of 10 executions), respectively. Also, According to Fig. 13 for the GWR model with the

exponential kernel, nine criteria of HOFD, NDVI, Slope, TPI, TWI, VOFD, DEM and CN and for the bi-square kernel, eight criteria of HOFD, MFI, NDVI, Slope, TPI, TWI, VOFD, DEM and CN were determined as effective criteria in estimating flood risk.



(a) **(b)**

Figure 12. The best value of fitness function by combining of GWR model and BPSO algorithm (a) Exponential kernel (b) Bi-square kernel



(a) **(b)**

Figure 13. Effective criteria in estimating flood risk by combining of GWR model and BPSO algorithm (a) Exponential kernel (b) Bi-square kernel

According to Fig. 14, the maps of estimated flood risk (based on effective criteria) by combining ANN and GWR models with BPSO algorithm showed in the range [0,1]. The estimated flood risk is classified into five output classes according to the Equal Interval classification method which is shown qualitatively from very low flood risk to very high flood risk. According to the results obtained from R^2 value (goodness of fit) and the RMSE value (Residuals distribution of the observation and accuracy of the models), the combination of GWR model with bi-square kernel and BPSO algorithm has a higher ability to estimate flood risk, which showed in Fig. 14c.

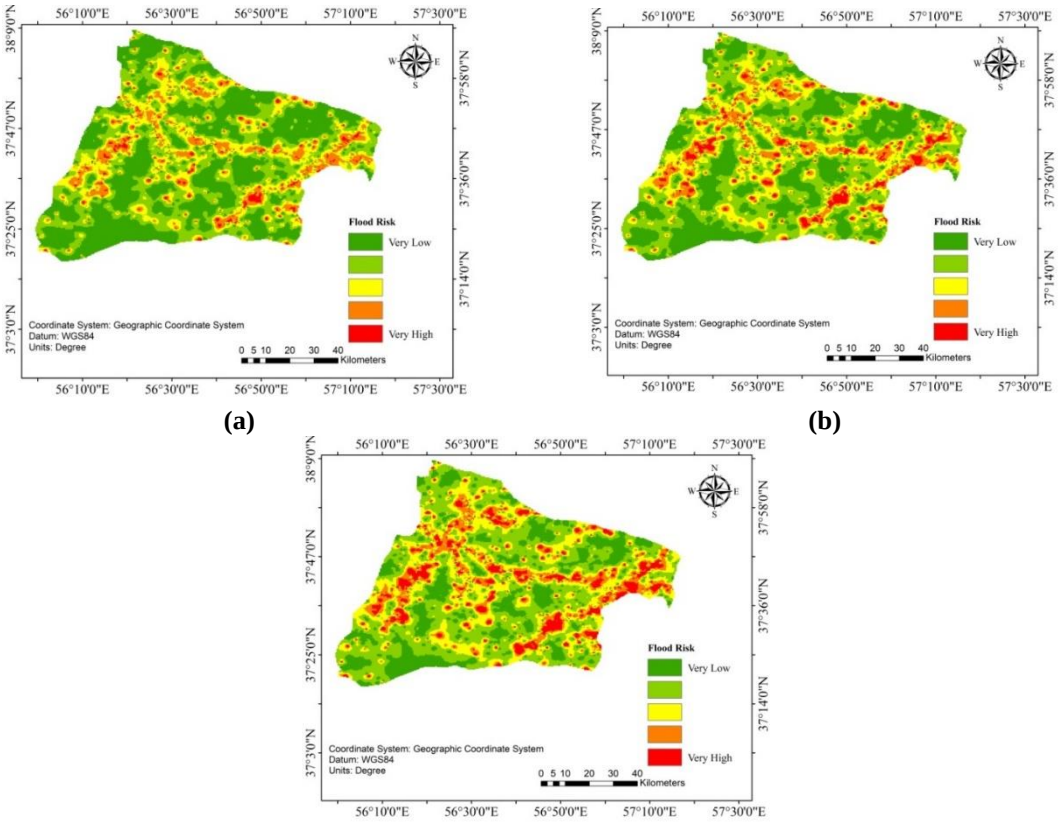


Figure 14. The map of estimated flood risk (a) ANN + BPSO (b) GWR (Exponential) + BPSO (c) GWR (Bi-square) + BPSO

In Fig. 15, the values of R2 and RMSE for the ANN and GWR models are shown. Accordingly, the bi-square kernel has higher accuracy in estimating flood risk based on effective criteria.

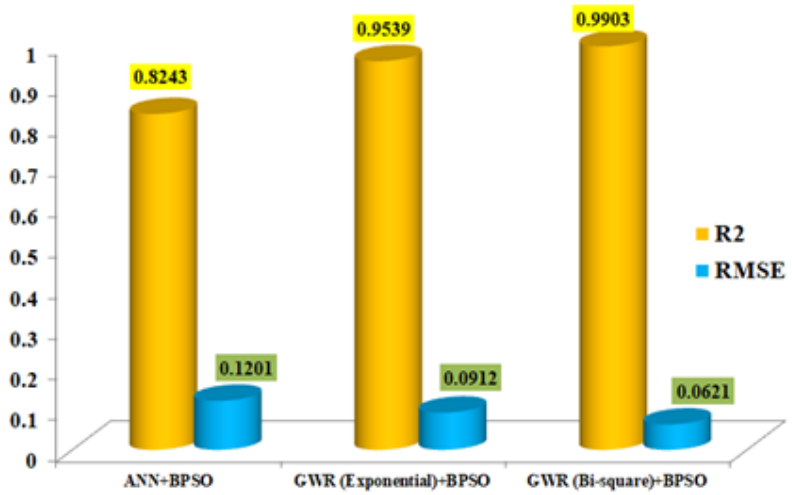


Figure 15. Comparison of ANN and GWR models in terms of R^2 and RMSE

As mentioned, since the regression coefficients are different for each location in the GWR model, local variation and spatial non-stationarity of the regression coefficients can be obtained by the standard deviation function. Fig. 16 shows the standard deviation of regression coefficients GWR model (with two exponential and bi-square kernels) for calculating the rate of local variation and spatial non-stationarity.

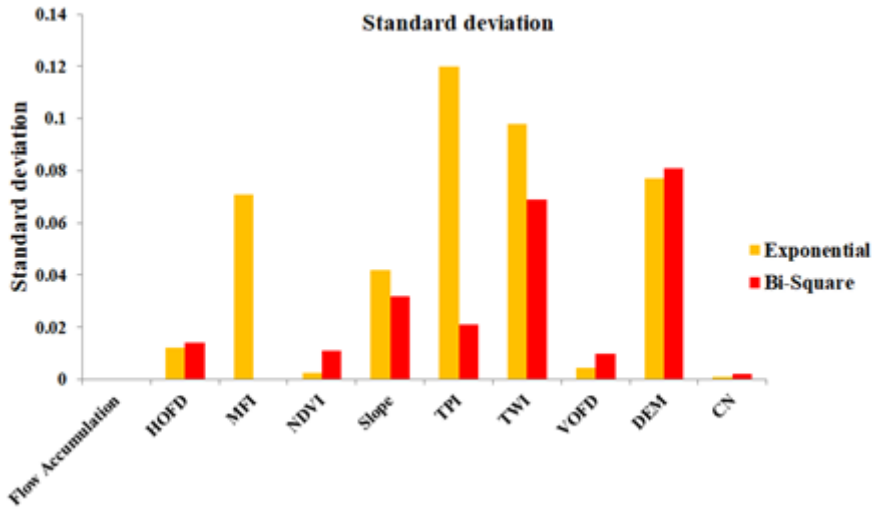


Figure 16. The standard deviation of regression coefficients GWR model with exponential and bi-square kernels

According to Fig. 16, for the GWR model with the exponential kernel, the relationship between TPI and flood risk with displacement has the most variation and the relationship between CN and flood risk has the least variation. Also, in the GWR model with the bi-square kernel, the relationship between DEM and flood risk with displacement has the most variation and the relationship between CN and flood risk has the least variation. Finally, global Moran's index was used to determine the spatial autocorrelation of GWR model residuals, which is calculated from Eq. 17 [34]:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i - \overline{X})(x_j - \overline{X})}{S_0 \sum_{i=1}^{n}(x_i - \overline{X})^2} \quad (17)$$

Where x_i and x_j are the estimated flood risk for random points i and j, W_{ij} is the spatial weight matrix between random points i and j, S_0 is the total of all weights, $\overline{X}$is the mean estimated flood risk for random points and n is the total random points. Table 4, shows the values of global Moran's index for GWR model residuals with two exponential and bi-square kernels.

Table 4. The values of Moran's index for GWR model residuals with two exponential and bi-square kernels

Parameters	Kernel type	
	Exponential	Bi-square
Moran's index	0.153	0.112
Expected index	0.000632	0.000632
Z-Score	11.72	12.03
P-value	0.000	0.000

4. Conclusions

Due to the increase in floods, especially in cities and the emergence of human, financial and environmental risks, the identification of criteria affecting the occurrence of floods is of great importance. Therefore, by identifying these criteria, we can prevent this phenomenon as much as possible by using public education to the people, enacting effective management laws and policies, and more oversight in order to deal with the stimulus criteria for increasing flood rates. Therefore, in this study, we tried to analyze the role of spatial criteria affecting floods in estimating flood risk, which has been neglected in many previous studies. The models used in previous research were not very suitable for spatial data and in most cases the spatial correlation and non-stationarity of the data were ignored. To achieve the main purpose of this study, the spatial and non-spatial data-driven models including GWR and ANN model were used to estimate the flood risk based on the effective criteria. The results showed that the GWR model used, taking into account the characteristics of spatial autocorrelation and spatial non-stationarity, has higher accuracy in estimating flood risk based on effective criteria. In this study, an attempt was also made to determine the effective criteria in estimating flood risk in the form of another study purpose. Therefore, the Binary particle swarm optimization algorithm was used in combination with the ANN and GWR models, which showed that the criteria have a significant effect in estimating flood risk (study area). The important point is that the mentioned method is not limited to this case study and can be used to estimate the flood risk in various types of regions.

Due to the success of the spatial data-driven model used in this research, it is suggested for future research other spatial data-driven models such as Generalized Method of Moments Estimation for Spatial Autoregressive (GMM-SAR), Matrix Exponential Spatial Specification (MESS) and combination of GWR with ANN were used.

References

1. Papaioannou, G., Vasiliades, L., Loukas, A. (2015). Multi-Criteria Analysis Framework for Potential Flood Prone Areas Mapping. Water Resources Management 29:399–418. https://doi.org/10.1007/s11269-014-0817-6
2. Lai, C., Shao, Q., Chen, X., et al. (2016). Flood risk zoning using a rule mining based on ant colony algorithm. Journal of Hydrology 542:268–280. https://doi.org/10.1016/j.jhydrol.2016.09.003
3. Hudson, P., Botzen, W.J.W. (2019). Cost–benefit analysis of flood-zoning policies: A review of current practice. WIREs Water 6:. https://doi.org/10.1002/wat2.1387

4. Pourghasemi, H.R., Razavi-Termeh, S.V., Kariminejad, N., et al. (2020). An assessment of metaheuristic approaches for flood assessment. Journal of Hydrology 582:124536. https://doi.org/10.1016/j.jhydrol.2019.124536
5. Rahmati, O., Zeinivand, H., Besharat, M. (2016). Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. Geomatics, Natural Hazards and Risk 7:1000–1017. https://doi.org/10.1080/19475705.2015.1045043
6. Kanani-Sadat, Y., Arabsheibani, R., Karimipour, F., Nasseri, M. (2019). A new approach to flood susceptibility assessment in data-scarce and ungauged regions based on GIS-based hybrid multi criteria decision-making method. Journal of Hydrology 572:17–31. https://doi.org/10.1016/j.jhydrol.2019.02.034
7. Sadeghi-Pouya, A., Nouri, J., Mansouri, N., Kia-Lashaki, A. (2017). Developing an index model for flood risk assessment in the western coastal region of Mazandaran, Iran. Journal of Hydrology and Hydromechanics 65:134–145. https://doi.org/10.1515/johh-2017-0007
8. Guevara, J., Zadrozny, B., Buoro, A., et al. (2018). A hybrid data-driven and knowledge-driven methodology for estimating the effect of completion parameters on the cumulative production of horizontal wells. In: Proceedings - SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers (SPE). DOI: 10.2118/191446-ms.
9. Wang, X., Liu, H. (2019). A Knowledge-and Data-Driven Soft Sensor Based on Deep Learning for Predicting the Deformation of an Air Preheater Rotor. IEEE Access 7:159651–159660. https://doi.org/10.1109/ACCESS.2019.2950661
10. Elsheikh, R.F.A., Ouerghi, S., Elhag, A.R. (2015). Flood Risk Map Based on GIS, and Multi Criteria Techniques (Case Study Terengganu Malaysia). Journal of Geographic Information System 07:348–357. https://doi.org/10.4236/jgis.2015.74027
11. Xiao, Y., Yi, S., Tang, Z. (2017). Integrated flood hazard assessment based on spatial ordered weighted averaging method considering spatial heterogeneity of risk preference. Science of the Total Environment 599–600:1034–1046. https://doi.org/10.1016/j.scitotenv.2017.04.218
12. Al-Juaidi, A.E.M., Nassar, A.M., Al-Juaidi, O.E.M. (2018). Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. Arabian Journal of Geosciences 11:1–10. https://doi.org/10.1007/s12517-018-4095-0
13. Ardiansyah, A., Sumunar, D.R.S. (2020). Flood Vulnerability Mapping Using Geographic Information System (GIS) in Gajah Wong Sub Watershed, Yogyakarta County Province. Geosfera Indonesia 5:47. https://doi.org/10.19184/geosi.v5i1.9959
14. Woznicki, S.A., Baynes, J., Panlasigui, S., et al. (2019). Development of a spatially complete floodplain map of the conterminous United States using random forest. Science of the Total Environment 647:942–953. https://doi.org/10.1016/j.scitotenv.2018.07.353
15. Khosravi. K., Nohani. E., Maroufinia. E., Pourghasemi, H.R. (2016). A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Natural Hazards 83:947–987. https://doi.org/10.1007/s11069-016-2357-2

16. Jancewicz, K., Migoń, P., Kasprzak, M. (2019). Connectivity patterns in contrasting types of tableland sandstone relief revealed by Topographic Wetness Index. Science of the Total Environment 656:1046–1062. https://doi.org/10.1016/j.scitotenv.2018.11.467
17. Vojtek, M., Vojteková. J, (2019). Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. Water (Switzerland) 11:. https://doi.org/10.3390/w11020364
18. Eini, M., Kaboli, H.S., Rashidian, M., Hedayat, H. (2020). Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. International Journal of Disaster Risk Reduction 50:101687. https://doi.org/10.1016/j.ijdrr.2020.101687
19. Alam, A., Ahmed, B., Sammonds, P. (2020). Flash flood susceptibility assessment using the parameters of drainage basin morphometry in SE Bangladesh. Quaternary International. https://doi.org/10.1016/j.quaint.2020.04.047
20. Saa-Requejo, A., Martin-Sotoca, J.J., Luis Valenciam, J. et al. (2019). Modified Fournier index as a new metric of integrated degradability index
21. El_Jerjawi, N.S., Abu-Naser, S.S. (2018). Diabetes Prediction Using Artificial Neural Network. International Journal of Advanced Science and Technology 121:55–64. https://doi.org/10.14257/ijast.2018.121.05
22. Lee, S., Hong, S.M., Jung, H.S. (2018). GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. Geocarto International 33:847–861. https://doi.org/10.1080/10106049.2017.1303091
23. Murray, A.T., Xu, J., Baik, J., et al. (2020). Overview of Contributions in Geographical Analysis: Waldo Tobler. In: Geographical Analysis. DOI: 10.1111/gean.12257.
24. Wu, D. (2020). Spatially and temporally varying relationships between ecological footprint and influencing factors in China's provinces Using Geographically Weighted Regression (GWR). Journal of Cleaner Production 261: https://doi.org/10.1016/j.jclepro.2020.121089
25. Fotheringham, A.S., Oshan, T.M. (2016). Geographically weighted regression and multicollinearity: dispelling the myth. Journal of Geographical Systems 18: https://doi.org/10.1007/s10109-016-0239-5
26. Oshan, T.M., Li, Z., Kang, W., et al. (2019). MGWR: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS International Journal of Geo-Information 8: https://doi.org/10.3390/ijgi8060269
27. Aghbashlo, M., Tabatabaei, M., Nadian, M.H., et al. (2019). Prognostication of lignocellulosic biomass pyrolysis behavior using ANFIS model tuned by PSO algorithm. Fuel 253: https://doi.org/10.1016/j.fuel.2019.04.169
28. Abed, K.A., Ahmad, A.A. (2020). The best parameters selection using pso algorithm to solving for ito system by new iterative technique. Indonesian Journal of Electrical Engineering and Computer Science 18: https://doi.org/10.11591/ijeecs.v18.i3.pp1638-1645

29. Kennedy, J., Eberhart, R.C. (1997). Discrete binary version of the particle swarm algorithm. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. DOI: 10.1109/icsmc.1997.637339.
30. Elfil, M., Negida, A. (2019). Sampling methods in clinical research; an educational review. Archives of Academic Emergency Medicine 7:52. https://doi.org/10.22037/emergency.v5i1.15215
31. Aad, G., Abbott, B., Abdallah, J., et al. (2014). Measurements of spin correlation in top-antitop quark events from proton-proton collisions at √s =7 TeV using the ATLAS detector. Physical Review D - Particles, Fields, Gravitation and Cosmology 90: https://doi.org/10.1103/PhysRevD.90.112016
32. Paulino, Â., Guimarães, L.N.F., Shiguemori, E.H. (2019). Hybrid adaptive computational intelligence-based multisensor data fusion applied to real-time UAV autonomous navigation. Inteligencia Artificial 22:162–195. https://doi.org/10.4114/intartif.vol22iss63pp162-195
33. Saeidian, B., Mesgari, M.S., Pradhan, B., Ghodousi, M. (2018). Optimized location-allocation of earthquake relief centers using PSO and ACO, complemented by GIS, clustering, and TOPSIS. ISPRS International Journal of Geo-Information 7: https://doi.org/10.3390/ijgi7080292
34. Zemestani, A., Soori, H. (2019). Relationship between fatal road traffic injury rates and Human Development Index in Iran. Journal of Injury and Violence Research 11. https://doi.org/10.5249/jivr.v11i2.1435