



OPEN Flood susceptibility mapping utilizing the integration of geospatial and multivariate statistical analysis, Erbil area in Northern Iraq as a case study

Alaa Ahmed^{1,2,3✉}, Ali Al Maliki⁴, Bassim Hashim⁴, Dalal Alshamsi^{1,2}, Hasan Arman¹ & Ahmed Gad^{1,5}

Climate extreme events such as floods and droughts in any area have a significant impact on human life, infrastructure, agriculture, and the economy. In the last two years, flash floods caused by heavy rainstorms have become frequent and destructive in many catchments in Northern Iraq. The present study aims to examine flash floods in the Erbil region, Northern Iraq using Remote sensing (RS), Geographic Information System (GIS), and Principal Component Analysis (PCA) for geomorphic data. PCA results revealed that 12 geomorphic parameters exhibited a significant correlation with two different statistical components. To facilitate practical application, ranks are assigned based on the calculated parameters for flood susceptibility mapping. Out of the 24 basins in the current study, three basins (16, 3, and 14) have the highest geomorphometric values (36–39), indicating the zone most susceptible to flash floods and making up a maximum area of 38.58% of the studied region. Six basins (4, 8, 9, 10, 12, and 15), which have geomorphometric values between 30 and 35 and cover a land area of 27.86%, are the most moderately vulnerable to floods. The remaining basins, which make up 33.47% of the research, are occasionally subject to floods and have geomorphometric scores below 30. The precision of the flood susceptibility mapping was validated using the bifurcation ratio and drainage density relationship as well as past flood damages, such as economic losses and human casualties. Most of the recorded injuries and fatalities took place in areas that were particularly prone to severe past flooding. Additionally, the investigation revealed that 44.56% of all populated areas are located in extremely vulnerable basins. The findings demonstrate a notable correlation between the identified flood-susceptible areas and the occurrence of past flood damage.

In recent decades, the impact of climate change has become increasingly observable. Hydro-meteorological disasters, for example, have become increasingly frequent and destructive worldwide¹. One-third of all the global geophysical disasters in the world are floods, which have a significant negative impact on infrastructure, economy, and population^{2–4}. Location, population, socioeconomic and cultural circumstances, institutional and political measures, as well as coping mechanisms that distinguish the impacts on individuals and human systems, have all been demonstrated to affect how vulnerable societies are to the effects of these disasters⁵, other factors that affect the level and scope of damage caused by a disaster in a place, including its susceptibility profile and features (frequency, duration, depth, and spatial extent)⁶. In this situation, vulnerability is predicated on the interactions of socioeconomic and environmental physical elements.

The Middle East and Northern Africa (MENA) region is among the most vulnerable to the potential impacts of climate change⁷. It is expected that the region will suffer from higher temperatures and intense heat waves affecting inhabitants and crop yields and will also affect marine ecosystems and fisheries⁸. In the MENA region,

¹Geosciences Department, United Arab Emirates University, 15551 Al Ain, United Arab Emirates. ²National Water and Energy Center, United Arab Emirates University, 15551 Al Ain, United Arab Emirates. ³Geology Department, Division of Water Resource, Desert Research Center, Mathaf El Matariya Street, Cairo 11753, Egypt. ⁴Ministry of Science and Technology, Environment, Water and Renewable Energy Directorate, Baghdad 765, Iraq. ⁵Geology Department, Faculty of Science, Ain Shams University, Cairo 11566, Egypt. ✉email: aahmed83@uaeu.ac.ae

less but more intense rainfall, coupled with higher temperatures will likely cause more droughts and greater flooding, sea level rise, more intense cyclones, and new areas exposed to dengue, malaria, and other vector and waterborne diseases. Therefore, determining the risks, vulnerabilities, and hazards associated with flooding has become critical. The increase in flooding incidents can be primarily attributed not only to climatological factors associated with heavy rainfall events, but also to the geomorphic features, land use changes, population expansion, and urbanization^{9,10}. Moreover, geological and structural features considerably influence the predictability of floods^{11–13}. The analytical hierarchy process, fuzzy logic, and genetic algorithms, as well as decision tree models, multivariate statistics, and the hydrological forecasting system, have all been developed as tools for analyzing flood risk. Thus, analyzing geomorphic parameters could aid in predicting floods, sediment yields, and erosion rates^{14–17}. The morphometric parameters have been introduced for measuring and mathematically analyzing the characteristics associated with the drainage basin¹⁸. Recently, the assessment of basin morphometry has improved in terms of accuracy, speed, and cost owing to the introduction of geographic information systems (GISs), remote sensing (RS), and high-resolution digital elevation models (DEMs)^{19–21}. Several studies have demonstrated the potential of integrating RS and GIS for geomorphometric analysis of drainage basins^{14,22,23}. This method is being used to develop useful, precise, and affordable techniques for defining flood susceptibility zones^{14,24–26}. This method aids in determining the overall characteristics of the drainage basins, including their terrain and elevation, as well as the relationship between morphological factors and flood susceptibility within the drainage basin. When determining this interrelationship, it is crucial to use multivariate statistical analysis, such as principal component analysis (PCA), to reduce dimensionality in large data sets. Multivariate methods are commonly used in hydrological research to provide solutions and options for a specific problem²⁷. It has been used to develop hydrological prediction equations involving geomorphic parameters, identifying the interrelationship between the different parameters and their effect on flood susceptibility^{28–31}. The novelty of this study is a multi-disciplinary evaluation approach that combines geospatial data, statistical analysis and knowledge of flood hazards to improve the future adaptation plans. Similar physical and socio-economic characteristics, as well as flood-prone areas worldwide, could lead to the replication of such a hydro-social analytic approach.

Iraq is classified as one of the most environmentally vulnerable and fragile countries in the world³². The country is placed at the heart of the violent effects of climate extremes. Flash floods induced by heavy rainfall events and urbanization have become common and destructive, especially in the northern areas of the country.

Recently, the failure to respond to devastating flash floods in Erbil, Northern Iraq, was mainly attributed to the lack of available information about the flood-vulnerable areas. This caused huge economic damage and public outrage in the region. Following the coalition invasion of the country in 2003, northern areas serve as home to a significant population and host several resources that significantly contribute to the economy³³. Consequently, dramatic urbanization, landcover change, and a significant increase in surface runoff become evident. The runoff in built-up areas increased from approximately 4 million cubic meters in 1990 to approximately 14 million cubic meters in 2016 and is expected to have increased by 2020³⁴. Therefore, identifying and assessing flood susceptibility can effectively help raise awareness among the people living in flood-prone areas to reduce damage, fatalities, and losses. The objectives of the present study are as follows: (1) identify and evaluate geomorphological parameters; (2) investigate the multicollinearity of relationships between variables to eliminate less controlling factors from the analysis; (3) combine the remaining data using PCA for easier interpretation; and (4) estimate and evaluate flash flood risk in the northern Iraq region.

Methods

Study area. The Kurdistan province, located in northern Iraq, was selected for this study (Fig. 1). The province, which is ~ 350 km from Baghdad, covers a total area of 1131.44 km². It comprises five sub-districts: Erbil, Koysinjaq, Soran, Shaqlawa, and Choman. Topographically, the central regions are mostly flat, whereas the northeast and eastern regions are semi-mountainous and hilly. Erbil province has a semi-arid climate, with hot and dry summers (June to October) and cold and wet winters (December to March), with heavy rain from October to May. The annual rainfall fluctuates within the study area because of topographic variation, indicating a high potential for flash flood events. Annual rainfall in the northern regions is 1275 mm, while it is 612 mm in the southern regions. In general, the average temperature ranges from 6 to 16 °C in the winter and from 25 to 40 °C in the summer.

Data analysis. In this study, different data sources and methodologies are applied for identifying the most vulnerable flood areas (Fig. 2). The digital elevation model (DEM) from the Shuttle Radar Topographic Mission (SRTM) was used to extract streams and drainage basins from the study area. The SRTM is available for download from NASA's Earth Explorer website (<https://earthexplorer.usgs.gov/>). It is a single-pass synthetic aperture radar interferometry (InSAR) campaign that provides unique DEM data with a resolution of 90 m. In contrast to the 10-m mesh DEM developed by the Geographical Survey Institute of Japan, the elevation error of the DEM ranges from +7.4 m (forest areas) to 0.7 m (bare land), whereas the horizontal inaccuracy and shifting range from 0.13 arc-seconds for East/West to 0.19 arc-seconds for North/South. The high-resolution, open, accurate, comparable, and timely land use/land cover maps are derived from the ESRI 2020 Landcover database (<https://livingatlas.arcgis.com/landcover/>). The database is a global product for a user-specified area of interest and time period (2018–2022). The database was developed in conjunction with ESRI by Microsoft Planetary Computer and scaled using Microsoft Azure Batch using a deep learning model trained on 5 billion manually annotated Sentinel-2 pixels spread across 20,000 sample points worldwide. The final output comprises a composite of nine classes of LULC predictions over the required time period, with an overall accuracy of 86% worldwide, according to the validation points. In addition to the available geological map from The Iraq Geological Survey (<http://en.geosurviraq.iq/>), the soil map (1:50,000) was downloaded from the State Board for Land Reclamation, Minis-

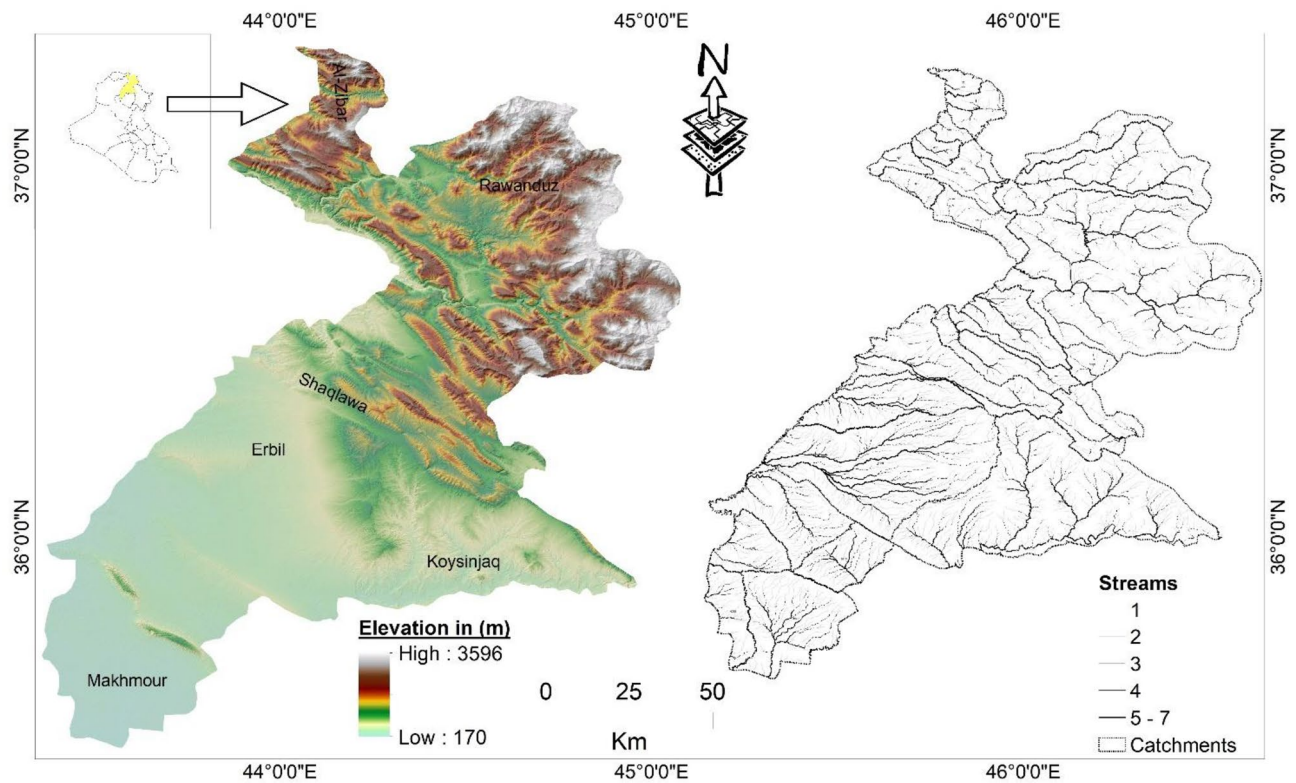


Figure 1. Location map of the study area.

try of Agriculture, Iraq (<http://zeraa.gov.iq/>). Soil classification has been developed using the FAO/UNEP international standard system based on the world soil map (FAO/UNESCO, 1974). Annual rainfall data from 1981 to 2022 are available from NASA's Prediction of Worldwide Energy Resources (<https://power.larc.nasa.gov>). The service offers precipitation data with a high resolution that was produced from the Integrated Multi-satellite Retrievals for the GPM component of NASA's Global Precipitation Measurement (GPM) program (IMERG). Global $0.1^\circ \times 0.1^\circ$ latitude/longitude grids are used to resolve IMERG precipitation data. Rainfall data had a higher resolution than previous rainfall grid maps; thus, it was utilized to create a spatial rainfall distribution map. Additionally, annual rainfall maxima collected from the RainSphere database of CHRS at University of California, Irvine (<https://rainsphere.eng.uci.edu/>) for the period from 1983–2021 has been used for the frequency analysis.

The Weibull distribution is commonly used for frequency analysis as well as risk and reliability analysis of the lifetimes of systems and their components. Its applications have been reported frequently in hydrology and meteorology. Weibull in 1939 introduced the equation for the analysis of food magnitude for the corresponding return periods. Plotting the distribution involves ranking the data. It is specifically used with the exponential distribution and the Rayleigh distribution. In addition to other forms of probability distributions, this technique's probability distribution is tied to the scale and shape parameters. The density function for this method is highly sensitive to changes in the value of the form parameter. Here, the rank of the food discharge for a particular distribution, the quantity of observations (N), and the frequency of occurrence (f), sometimes referred to as the probability percentage.

Moreover, historical data on the flood disaster was available through the OFDA/CRED International Disaster Database (EM-DAT), which provides crucial core information on the occurrence and consequences of 22,000 mass disasters that have occurred worldwide from 1900 to the present. The database is compiled from various sources, including non-governmental organizations, insurance providers, and academic institutions. EM-DAT is available to the general public at <https://www.emdat.be/>. For this study, historical losses including human and economic damages recorded in EM-DAT across all flood events from 2006 through 2021 in the study area to validate the resultant final map.

The ArcGIS hydrology spatial analyst tool extracts streams and delineates watersheds in a sequential manner during subsequent processing, beginning with filling sinks to remove any errors from the raw DEM, identifying the flow direction from higher to lower, assigning direction codes, and delineating the flow accumulation to determine the higher possibility flow accumulation pixels. The drainage network was built through a trial-and-error method, taking into account pixels with flow accumulation greater than a threshold of 900 m^2 . The user provided pour points, which were used to build the stream network and identify the sites with the highest flow accumulation. Pour locations were determined assuming that the calculated drainage basins encompassed the entire river basins. A stream order was determined using a stream ordering method^{35,36}. Strahler's ordering

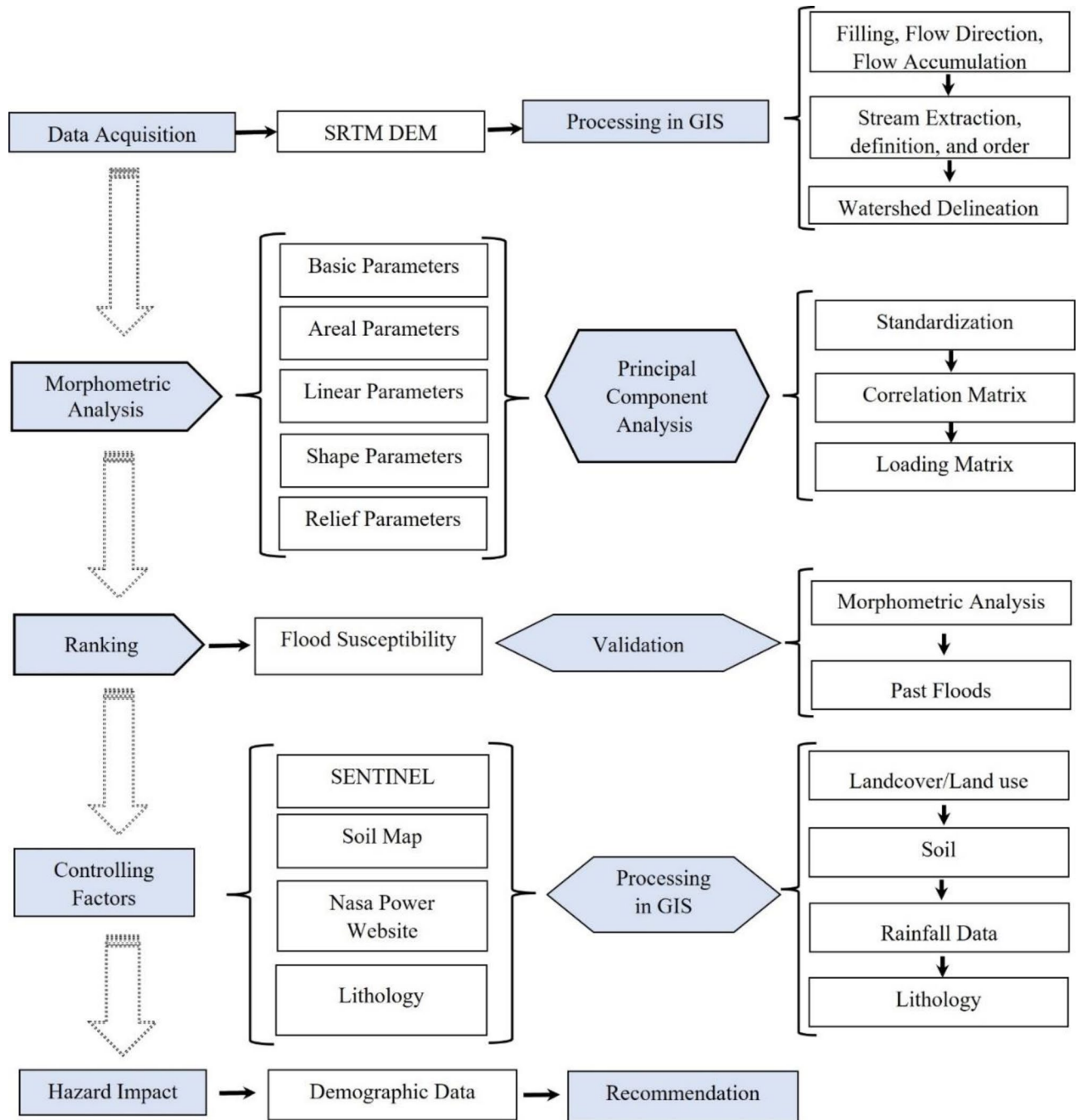


Figure 2. Flow chart showing the methodology applied in this study.

scheme places streams without tributaries in the first order, streams with two first-order tributaries in the second order, and so on. A well-drained basin comprises fifth-order stream channels³⁵.

To assess flood susceptibility, 24 morphometric variables (linear, areal, shape, and relief) were selected from the study area and computed employing an established formula (Table 1). Stream order (No), stream number (Nu), stream length (Lu), sinuosity ratio (Si), and length of overland flow (Lo) were calculated and ranked as linear drainage morphometric criteria. Drainage density (Dd), drainage frequency (Fs), stream texture (Dt), constant of channel maintenance (C), and infiltration ratio (If) were also extracted and included as areal drainage aspects. Shape parameters calculated include the elongation ratio (Re), circulatory ratio (Rc), compactness coefficient (Cc), form factor (Ff), and shape index (Sw), whereas relief parameters include the basin relief (Bh), relief ratio (Rr), and ruggedness number (Rn). PCA represents a multivariate statistical method commonly and efficiently used to reduce data size while preserving as many changes as possible in the dataset for faster data processing³⁷. Many variables can be reduced to significant factors using PCA, and the final factors summarize the original data. XLSTAT software was used to perform PCA for 24 morphometric variables in 24 watersheds. Because the approach is heavily reliant on the sum variance of the original variables, it is critical to represent

Parameter	Formula	Reference
Basin area (km)	A	Schumm (1956)
Basin perimeter (km)	P	Schumm (1956)
Basin length (km)	L_b	Gregory and Walling (1973)
Total stream length (km)	$L_u = L_1 + L_2 \dots L_u$	Strahler (1957, 1964)
Total stream number	$N_u = N_1 + N_2 \dots N_u$	Strahler (1957, 1964)
Drainage density	$Dd = L_u/A$	Horton (1932)
Frequency	$F_s = \sum N_n/A$	Horton (1932)
Elongation ratio	$E_b = \sqrt{2} Ab/lb$	Schumm (1956)
Circularity ratio	$C = 4\pi A/P^2$	Miller (1953)
Form factor	$F = A/L^2$	Horton (1932)
Compactness coefficient	$c = P/2\sqrt{pA}$	Gravel us (1914)
Sinuosity ratio	$SI = OL/EL$	Schumm (1956)
Longest channel	L_c	Horton (1932)
Slope	$S = B_h/L_b$	Schumm (1956)
Time of concentration	$T_c = 0.0078L_b^{0.77}S^{-0.38}$	Kirprich 1940
Length of overland flow	$L_o = 0.5 \times 1/Dd$	Schumm (1956)
Constant of channel maintenance	$C = 1/Dd$	Schumm (1956)
Shape index	L_b^2/A	Horton (1945)
Infiltration number	$F_s \times Dd$	Faniran 1986
Basin width	W_b	Horton (1932)
Drainage texture	$Td = \sum N_u/P$	Horton (1945)
Basin relief	$B_h = H - h$	Schumm (1956)
Relief ratio	$R_r = H_r/L$	Schumm (1956)
Ruggedness number	$R_n = H/Dd$	Strahler (1957, 1964)
Bifurcation ratio	$(R_b) R_b = N_n - 1/N_n$	Schumm (1956)

Table 1. List of parameters, formulas, and references used in the present study.

the variables in a standard form, implying the same unit of measurement for each variable, to ensure a sample variance equal to 1. The correlation matrix is then examined with the total variance set to n . The inter-correlation matrix is obtained by standardizing the factors based on the following formula:

$$X = (x_{ij} - x_j)/S_j \quad (1)$$

where X denotes the matrix of standardized parameters; x_{ij} denotes the i th observation on the j th parameters; $i = 1 \dots N$ (Number of observation); $j = 1 \dots P$ (Number of observations); x_j denotes the j th parameters; and S_j denotes the standard deviation of the j th parameters. The correlation matrix of parameters is the minor product moment of the standardized predictor measures divided by N and is given as follows:

$$R = (x' \times x) \quad (2)$$

where, x' denotes the transpose of the standardized matrix of predictor parameters. The principal component loading matrix, which indicates how much an individual parameter is correlated with different factors, is obtained by pre-multiplying the characteristics vector with the square root of the characteristic values of the correlation matrix. Thus,

$$A = Q \times D^{0.5} \quad (3)$$

where A denotes the principal component loading matrix; Q denotes the characteristics vector of the correlation matrix; and D denotes the characteristics value of the correlation matrix. A rank (y) was determined for each basin using a relative ranking method based on the degree of flood susceptibility, where 1 indicates a very low susceptibility and 5 is a very high susceptibility in proportion to the value of each morphometric parameter (x). A linear interpolation method described by³⁸ was used to calculate ranks. When the morphometric variable and the flood event have a positive correlation, Eq. (4) is used. Otherwise, Eq. (5) is used.

$$y \propto x, \quad y'_n = (y_2 + y_1)(x'_n - x_{\min})/(x_{\max} - x_{\min}) + y_1 \quad (4)$$

$$y \propto 1/x, \quad y'_n = (y_2 + y_1)(x'_n - x_{\max})/(x_{\min} - x_{\max}) + y_1 \quad (5)$$

The flood risk degree increases with the parameter value for group 1 (Eq. 4), whereas the flood risk degree decreases with the parameter value for group 2 (Eq. 5). Finally, the number of basins was determined by adding

the standardized scores for each attribute. The calculated data were used to classify the basins into groups based on the risk and degree of flood vulnerability.

Regardless of the method used for validation, it is essential to validate the flood susceptibility maps³⁹. Morphometric parameters such as Rb, Dd, and Fs were usually compared to the proposed model in order to validate the use of the flood susceptibility map^{16,40–42}. In this regard, basin divided into three zones of the vulnerable flood basins based on the Rb in conjunction with Dd and Fs⁴¹. Zone “A” represents low flood susceptibility, zone “B” represents a high-flood risk, and zone “C” represents an intermediate flood risk. Furthermore, known flood areas from past years were used to validate the final flood susceptibility map. In the present study, the same morphometric parameters (Rb, Dd, and Fs) of the drainage basins are computed and graphically presented for comparing and validating the obtained results. For additional validation, the past flood locations from the International Disaster Database (EM-DAT) are compared to the identified vulnerable areas.

Results

Morphometric analysis. The peak discharge of the river, which causes flash floods, is significantly influenced by the geomorphometric features of the basins. For the 24 sub-basins (see supplementary, Table S1), the basin area (A) ranges from 4618 km² in the central regions (basin no. 14) to 107.61 km² in the northern regions (basin no. 20). In addition, basin perimeter (P) ranged from 57.71 km² for basin no. 21 to 338.49 km² for basin no. 14. The number of streams is relatively high for the central basins, such as 3, 4, 9, 14, and 16, indicating a significant discharge capability and flooding susceptibility. Among the 24 watersheds, 11 are assigned as fourth-order basins, 9 are fifth-order, 4 are sixth order, and 1 represents seventh order (see supplementary, Table S1). Further, the stream length characteristics of the different watersheds exhibited a noticeable variation, with basin no. 14 having the highest (2048 km) and basin no. 21 having the lowest (89 km) total length (Fig. 3). The overland flow (L_o) significantly impacts the physiographical, topographical, and hydrological development of these streams in the drainage basins. The sinuosity index (Si) ranges from 0.54 to 1.41. risk of erosion⁴³.

Because Dd has a direct impact on the flood hydrograph, susceptibility modeling heavily relies on it as a geomorphometric factor. The results of this study showed a variation between the upper basins in the central (highest values) and southern areas (lowest values), owing to variations in the slope, soil infiltration, and rainfall throughout the study area (Fig. 3). The higher infiltration numbers in the central and southern basins indicate lower infiltration and higher drainage conditions. The reciprocal of drainage density was used⁴⁴ to define the morphological property of drainage basins regarding a constant of channel maintenance (C). Similarly, the Fs values of the studied basins range from 0.06 to 3.97 (see supplementary, Table S2). Such considerable values in the central areas imply moderate to high runoff and substantial peak discharge, posing flash flood risks compared to other basins within the study area (see supplementary, Figure S1). Furthermore, the Dt values vary from 0.38

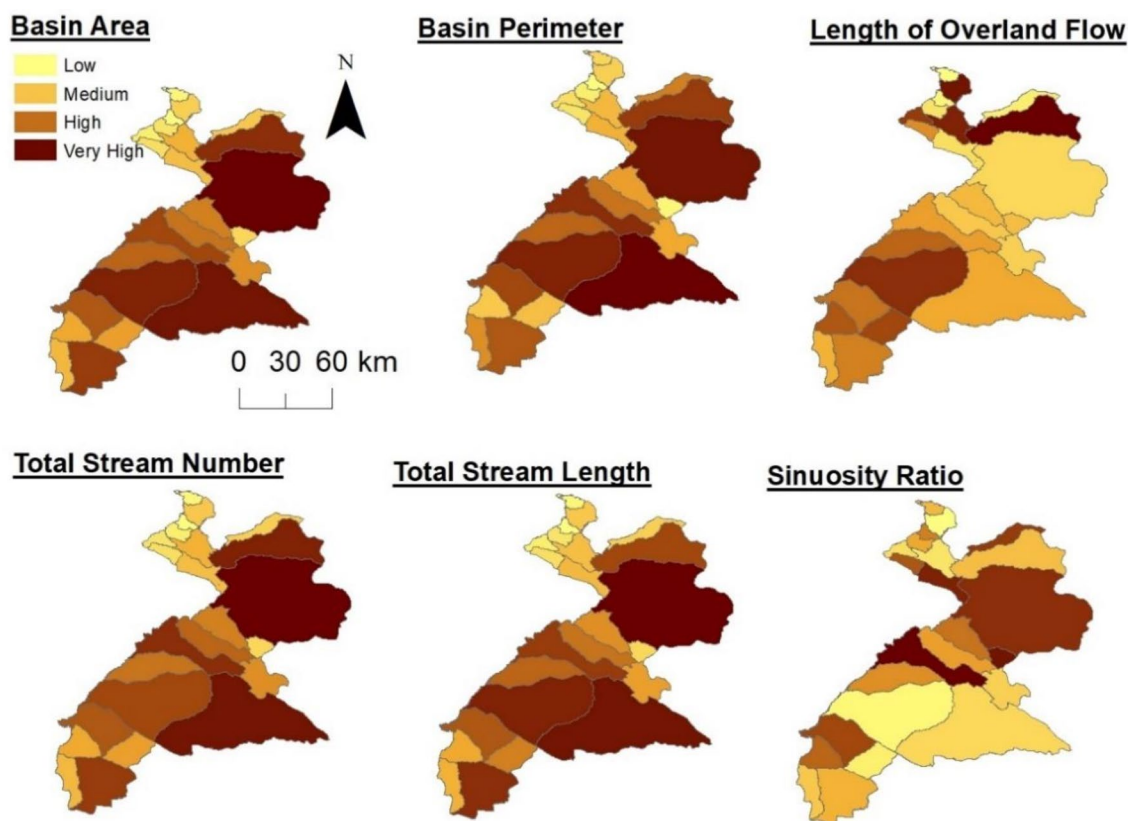


Figure 3. Spatial distribution of some morphometric parameters throughout the study area.

to 8.95, with high values indicating very coarse texture as well as the presence of unstable slope materials, soft rocks, and an increased the calculated circularity ratios (Rc) for basins are between 0.19 (basin no. 9) and 0.62 (basin no. 13). Moreover, the elongation ratio (Re) of the basins ranges from 0.27 to 1.0, indicating that most watersheds are circular, with a few exceptions being elongated. Additionally, the shape factors for the basins range from 0.25 to 15.74, confirming their circular nature and making them more susceptible to flooding and experiencing greater erosion and sediment transport capacities.

Furthermore, the compactness coefficient (Cc) and form factor (Ff) are both used to determine the relationship between the hydrological conditions and the circular shape of the drainage basin. All watersheds have a Cc of ≥ 1 , indicating less elongation, high erosion, and high flood susceptibility⁴⁵. Horton (1945) introduced Ff to forecast the flow intensity of a drainage basin manipulating the direct relationship between peak discharge and the inverse relationship between the squares of the axial length of the drainage basin. Ff values for the basins range from 0.06 to 3.7. Watersheds with high Ff values experience high peak flows with short duration. In contrast, a long watershed with a low form factor experiences modest peak flows that last longer. The period of concentration (Tc), a measure of the physical properties of a basin, is often low in the central and southern basins (see Supplementary, Figure S2), supporting the latter's high vulnerability to flooding. Basin relief (Bh) influences stream gradients, flooding patterns, and the number of transported sediments. Basins 1 and 2 have the lowest basin relief (50 m), whereas basin no. 14 has the highest (2770 m). In the study area, steep slopes dominated the northern and central basins, while gentle slopes dominated the southern basins (see supplementary, Table S2).

Selection of the causative factors. The correlation matrix of 24 geomorphic parameters indicates four classes (see Supplementary, Table S3): First, excellent correlations were found between basin perimeter and basin area, total stream number and basin area, stream frequency and density, channel maintenance and stream frequency, infiltration number, drainage density and frequency, channel maintenance, Lo, and ruggedness number. Second, a significant correlation (correlation coefficient >0.75) was noted between total stream number and basin length, total stream number and total stream length, total stream length and perimeter of the basin, slope and form factor, ruggedness number and drainage density, stream frequency, as well as Lo and time of concentration. Third, a moderate correlation (correlation coefficient >0.60) existed between basin length and perimeter, total stream length and basin length, length of the main channel and total stream length, basin width and basin area, drainage texture and drainage density, length of overland flow, channel maintenance, relative relief ratio and compactness coefficient, as well as slope and elongation ratio. In addition, a weak correlation (correlation coefficient >0.7) existed between the time of concentration and total stream number, basin width and form factor, basin width and infiltration number, drainage texture and total stream number, and ruggedness number and basin width. At this stage, it is challenging to classify the parameters and determine the significance of the various aspects. Therefore, the PCA was used. The component loading matrix derived from the correlation matrix shows that the first four components, whose eigenvalues are greater than 1, account for about 85.3% of the total explained variance (Table 2). The rotated component matrix indicates that components are significant. Component 1 (areal parameters) is highly correlated with density, frequency, length of overland flow, constant of maintenance, infiltration number, and ruggedness number, whereas component 2 (F2) is strongly correlated with the basin area, perimeter, basin length, total stream length and number, drainage density, main channel length, and time of concentration (Table 3).

As previously stated, the most crucial morphometric characteristics exhibit high correlation. As a result, only twelve of the twenty-four parameters used in this study are used to map flood susceptibility. Thus, the PCA was useful in removing the variables with the least relevance and grouping the remaining variables into physically significant components.

Flood susceptibility mapping. The degree of flash flood susceptibility is determined for 24 basins using the geomorphometric ranking approach (see Supplementary, Table S4). The approach has been applied by many researchers to evaluate the impacts of the morphometric characteristics of the drainage basin on flood

	Eigen value	Initial		Extracted	
		Variability (%)	Cumulative (%)	Variability (%)	Cumulative (%)
F1	8.47	35.27	35.27	35.14	35.14
F2	7.69	32.02	67.29	32.12	67.26
F3	2.69	11.22	78.51		
F4	1.62	6.76	85.27		
F5	0.78	3.24	88.51		
F6	0.44	1.85	90.36		
F7	0.15	0.64	90.99		
F8	0.1	0.41	91.41		
F9	0.06	0.25	91.65		
F10	0.04	0.15	91.80		
F11	0.01	0.04	91.84		

Table 2. Total variance explained of the studied basins.

Parameter	D1	D2
A	-0.25	0.85
P	-0.24	0.91
L _b	0.31	0.85
Lu	0.01	0.9
Nu	0.27	0.92
Dd	0.95	0.08
Fs	0.92	-0.19
Bh	-0.07	-0.66
Er	-0.76	-0.05
Cr	-0.24	-0.18
Ff	-0.69	-0.9
Cc	0.5	-0.79
Sr	-0.22	0.18
Lc	0.34	0.85
S	-0.24	0.56
Rr	0.39	-0.66
Tc	0.1	0.78
Lo	0.95	0.08
Cc	0.95	0.08
Si	-0.59	0.02
If	0.88	-0.03
Bw	-0.61	0.55
Dr	0.79	0.49
Rn	0.86	-0.01

Table 3. Rotated component matrix of the geomorphic parameters.

susceptibility^{14,46}. The method was used to calculate the cumulative ranking score for various geomorphometric features divided into two groups: positively correlated features (total stream number and length, basin area, drainage density, frequency and texture, and ruggedness number) and negatively correlated features (basin length, length of overland flow, time of concentration, stream maintenance, and infiltration number). The cumulative ranking score was used to divide the study area into susceptibility zones using geomorphometric data. The geomorphometric number is inversely proportional to the flash flood susceptibility of basins. According to the investigation, all basins have a geomorphometric number between 26 and 39. Out of the 24 basins in the present study, basins 16, 3, and 14 have the highest geomorphometric values (36–39), signifying the zone most vulnerable to flash floods and making up a maximum land area of 38.58% of the study area. The moderately susceptible basins are basins 4, 8, 9, 10, 12, and 15, which cover a land area of 27.86% and have geomorphometric numbers between 30 and 35. The remaining basins comprise 33.47% of the study; they have limited exposure to floods and have geomorphometric numbers < 30 (Fig. 4).

It is critical to validate flood susceptibility to illustrate how effectively the map performed. Rb is considered to represent the most important parameter for determining the degree of risk and has a greater impact than the other parameters⁴¹. Rb versus Dd and Fs relationships were displayed in Fig. 5. This method differs from PCA and ranking methods, which equalize the impacts of all parameters and are screened to identify the most significant factors. This difference can be attributed to the various weights that are assigned to the various parameters.

Discussions

In arid regions, flash floods are among the most devastating natural disasters, responsible for loss of life and serious damage to vital infrastructure and buildings⁴⁷. Reliable and accurate data on flash flooding in arid environments are frequently lacking due to the remoteness and sparse habitation of such areas²¹. Delineation of flood susceptible areas is considered the basic action toward flood mitigation. In this study, the flood-prone areas are delineated through the integration of flood causative factors in a GIS environment, PCA, and ranking method. Results of the PCA indicate that the hydro-morphometric characteristics of the drainage basins can be combined into two main groups that account for 85% of the variation of the data and have clear physical meaning. The PC represents drainage density, stream frequency, length of overland flow, and constant maintenance as the most important predictors of flash flooding, which is consistent with what is commonly believed. Consequently, basins of the eastern areas are more susceptible to flooding when compared to other basins throughout the study area. Typically, basins in these areas are characterized by a high drainage density, high stream frequency, and high overland flow indicating high runoff potentiality and less infiltration^{48,49}. In addition to geomorphic features, the duration of rainfall, soil erosion, increased population, urbanization and settlements, increased infrastructure, the efficiency of the sewer system, and public awareness, could contribute to flash floods⁵⁰. In recent years, Climate

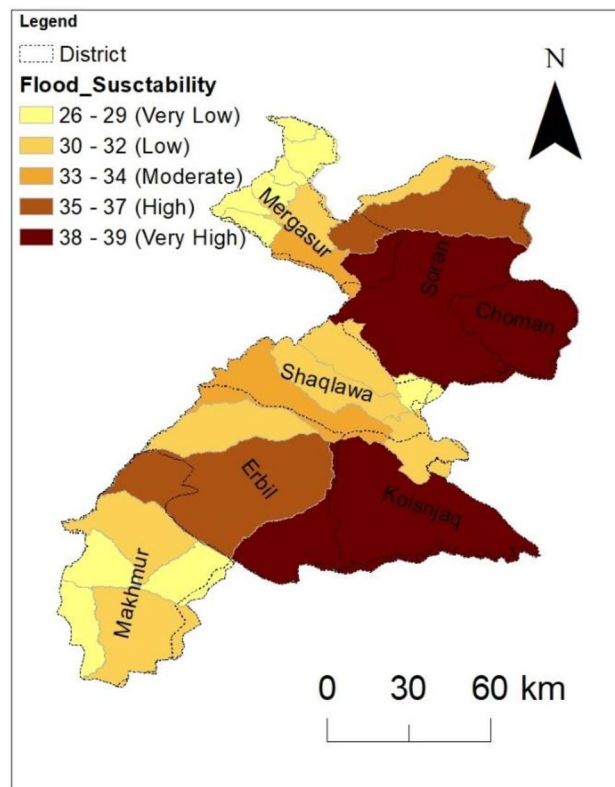


Figure 4. Flood susceptibility map and susceptible areas in the different districts.

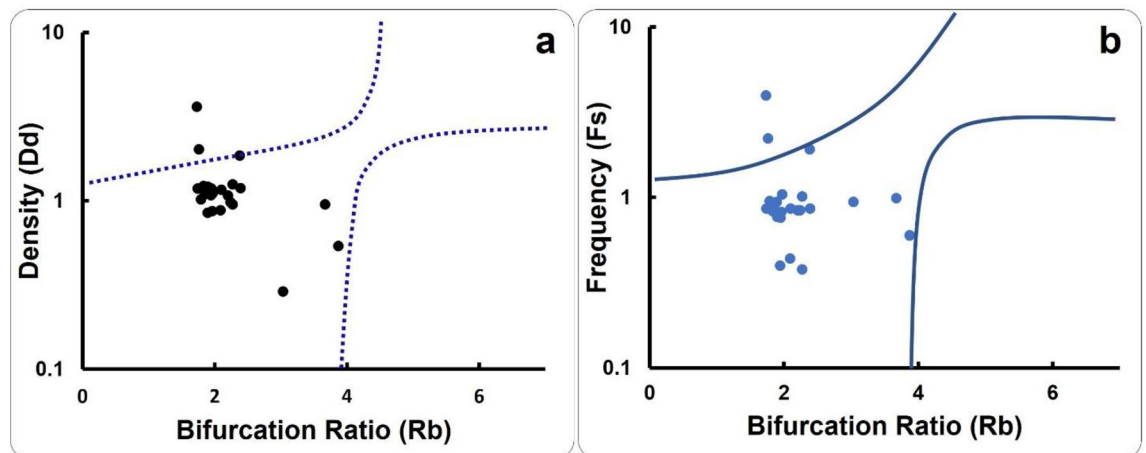


Figure 5. validating the flash flood hazard degree using (a) Rb versus Dd; (b) Rb versus Fs.

change including variability in temperature, rainfall patterns, and storm activity is increasing⁵¹. There is rising evidence that the global climate is changing in response to natural and human factors⁵². Spatial distribution of the average annual rainfall indicates a variation from the western to the eastern parts of the study area (Fig. 6). Six heavy rainfall events were recorded between 2008 and 2018⁵³.

Over the study area, intense rainfall has returned almost every two years, causing numerous flash floods with no warning. Results show that the total annual rainfall of the study area ranges from 244 to 633 mm, with an average of 436 mm (see Supplementary, Figure S3). The maximum annual daily rainfall varies from 12 to 45 mm with an average of 25 mm. In addition, the number of rainy days ranges between 111 to 234 mm, with an average of about 194 days/year. Based on the maximum daily rainfall data, statistical investigation is accomplished through the Weibull method and examined by probability distribution to fit the rainfall data to reach the best distribution curve, which matches the actual data of each rainfall station. From this graph, it is evident

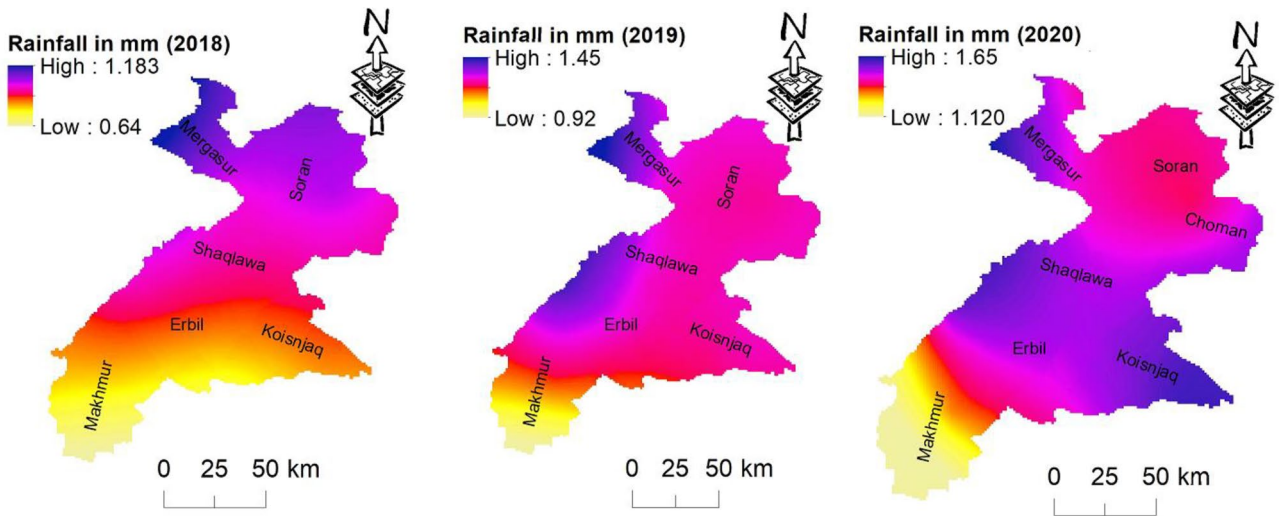


Figure 6. Annual rainfall distribution for the period 2018–2020.

that that in the following periods (5, 10, and a 100 years), an increase in the rainfall estimates is predicted (see Supplementary, Figure S3).

Moreover, changes in land use/cover significantly affect the occurrence of surface water, infiltration, evapotranspiration, and soil erosion^{54,55}. Evidently, the most common change that occurs in this area represents a transition from earlier natural to impervious surfaces. As cities expand due to population growth, fewer wetlands and more built-up regions exist. Infrastructure in populated regions is associated with an increase in impervious surface, which changes the hydrologic characteristics of the area (see supplementary, Figure S4). Therefore, soil infiltration decreases, surface runoff increases, and more sediment loads are transported into water bodies downstream, contributing to flooding. Thus, changes in land use and land cover cause soil compaction, reducing soil infiltration, and increasing runoff volume and speed⁵⁶. The sewer system in the study area is old, disconnected, and insufficient to reduce the effectiveness of runoff during flood events. During the most intense rainfall events, outflow occurred due to sewer system failure. Finally, poor sewer system management and a lack of public awareness comprise factors that contribute to flash floods in this area. For proper flood management, clean and efficient inlets and sewer lines are essential.

A significant outcome of the present study is a regional flood susceptibility map of an arid and vulnerable climate area, which is one of the most flood-affected regions of the country. The value of the current work was expected to be limited by the large basin characteristics, which, in turn, could impact the expectation of the results to differ on a smaller scale. Interestingly, a comparison of the regional study results (here is the current study) with the results of local studies in the same area is beneficial. Therefore, this study shows that a large set of hydro-morphometric parameters can be reduced to a smaller set without loss of information, indicating redundancy in the data. The final map was certified for greater accuracy by comparing the flood hazard map with historical flood data from the International Disaster Database (EM-DAT), which highlights flood-prone locations (Table 4). Floods caused by heavy rain have become more common in recent decades. The findings of this investigation are in excellent agreement with the EM-DAT flood data. Most of the recorded injuries and fatalities in the terrible flash flood events of 2006, 2009, 2011, and 2021 occurred in extremely vulnerable regions (Table 4). Additionally, the investigation showed that 44.56% of all human settlements are in extremely sensitive basins (Choman, Mergasur, Soran, and Erbil). It is expected that the prepared map will help disaster managers,

	2006	2006	2008	2009	2011	2021
Location	Soran	Erbil	Choman	Mergasur	Erbil	Erbil
Total deaths	20.00		4.00	2.00	6.00	14.00
No injured	20.00				1.00	
No affected	18,000.00	41,890.00		3000.00		7500.00
No homeless			600.00		2000.00	
Total affected	18,020.00	41,890.00	600.00	3000.00	2001.00	7500.00
Total damages (*000 US\$)		1300.00				14,000.00
Total damages, adjusted (*000 US\$)		1747.00				14,000.00
CPI	74.40	74.40	79.46	79.17	83.01	100.00

Table 4. Past flood hazard level within the study area according to the EM-DAT Database.

planners, and local authorities to improve their understanding of flood vulnerability in the region. It can also be expected that the map will assist in guiding the operational responses of the various authorities in the region, especially in areas that have been identified as high-risk zones. The most effective strategy to lessen flooding is to implement a variety of carefully considered strategies that are suited to the specific catchment of interest⁵⁷ assert that no single mitigation approach has ever been able to effectively address the flood problem. Consequently, they claimed a better strategy that incorporates land use planning, engineering fixes, flood preparedness, and emergency management in the affected lowlands. This strategy would also consider the social and economic needs of communities in both the highland, which are frequently source areas and the lowland.

While there is typically nothing that can be done to reduce the rainfall amount that a region receives, there are effective steps that can be taken to reduce the amount of surface runoff, which is a source of flooding. To mitigate the effects of the increased impervious surface caused by urban dwelling developments, rainwater harvesting is considered as an effective method for achieving this. The technique has been used in many developing countries with inadequate drainage systems as a means of mitigating some anthropogenically induced ecological problems^{58–60}. Along with rainwater harvesting, there should be efficient land use/landcover management practices that could reduce the flood hydrograph peak which is usually of utmost concern. Reducing impermeable surfaces, unless essential, can also result in greater advantages. In order to effectively and sustainably reduce flooding, the use of materials and construction designs like permeable pavements for parking lots and sidewalks, green roofs, rain gardens, retention cisterns, and natural rehabilitation are recommended.

Conclusions

In this study, an effective mitigation technique for achieving early flood warning and effective flood crisis management by accurately mapping flood susceptibility in the Erbil area of northern Iraq was implemented through the integration of geospatial techniques, watershed characteristics, and PCA, flood risk zones were identified. The PCA analysis reduced the initial set of 24 morphometric parameters to two important statistical components, using the component loading matrix, which provided insights into the correlations between the morphometric variables and helped to identify the key factors influencing flood susceptibility. Based on the morphometric relative ranking, the study categorized the drainage basins into three classes: low, moderate, and high flood susceptibility. The analysis revealed that three basins have a considerable degree of flood susceptibility, six basins have a moderate degree, and the remaining basins have a low degree. The produced susceptibility mapping was validated by comparison with historical flood occurrences in the study area. The observed correlation between the identified flood-susceptible areas and past flood damages, such as economic losses and human casualties, further supports the accuracy of the flood susceptibility mapping approach. The integrated methodology of morphometric analysis, PCA, and GIS can be recommended as a valuable tool for efficient planning, management, and decision-making to detect flood hazards, create flood mitigation plans, and enhance the resilience of the region to future flood risks. In contrast to other methods routinely used for creating susceptibility maps, the present methodology requires considerably simpler input data, notably morphometric data, which can be readily retrieved from accessible DEMs with good resolution. The proposed method does not require other parameters, such as an inventory map, which is time-consuming and expensive to build. To fully curb the problem of flooding, it is highly recommended to include artificial intelligence in predicting the flood phenomenon in the region because of its importance in determining future plans and preventing flood dangers in the region.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 22 May 2023; Accepted: 22 July 2023

Published online: 24 July 2023

References

1. Llasat, M. C. *et al.* High-impact floods and flash floods in Mediterranean countries: The FLASH preliminary database. *Adv. Geosci.* **23**, 47–55 (2010).
2. Adhikari, P. *et al.* A digitized global flood inventory (1998–2008): Compilation and preliminary results. *Nat. Hazards* **55**, 405–422 (2010).
3. He, B. *et al.* Analysis of flash flood disaster characteristics in China from 2011 to 2015. *Nat. Hazards* **90**, 407–420 (2018).
4. Singh, O. & Kumar, M. Flood events, fatalities and damages in India from 1978 to 2006. *Nat. Hazards* **69**, 1815–1834 (2013).
5. Seneviratne, K., Baldry, D. & Pathirage, C. Disaster knowledge factors in managing disasters successfully. *Int. J. Strateg. Prop. Manag.* **14**, 376–390 (2010).
6. Abid, S., Sulaiman, N., Wei, C. & Nazir, U. In *IOP Conf. Series: Earth and Environmental Science*. 012059 (IOP Publishing).
7. Mohammed, S., Sayl, K. & Kamel, A. Ground water recharge mapping in Iraqi Western desert. *Int. J. Des. Nat. Ecodyn* **17**, 913–920 (2022).
8. Solomon, S. in *Agu Fall Meeting Abstracts*. U43D-01.
9. Kunkel, K. E. North American trends in extreme precipitation. *Nat. Hazards* **29**, 291–305 (2003).
10. Llasat, M. C., Marcos, R., Turco, M., Gilabert, J. & Llasat-Botija, M. Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia. *J. Hydrol.* **541**, 24–37 (2016).
11. Islam, A. R. M. T. *et al.* Flood susceptibility modelling using advanced ensemble machine learning models. *Geosci. Front.* **12**, 101075 (2021).
12. Kalantari, Z. *et al.* Assessing flood probability for transportation infrastructure based on catchment characteristics, sediment connectivity and remotely sensed soil moisture. *Sci. Total Environ.* **661**, 393–406 (2019).
13. Mahmood, S. & Rahman, A.-U. Flash flood susceptibility modelling using geomorphometric approach in the Ushairy Basin, eastern Hindu Kush. *J. Earth Syst. Sci.* **128**, 1–14 (2019).

14. Adnan, M. S. G., Dewan, A., Zannat, K. E. & Abdullah, A. Y. M. The use of watershed geomorphic data in flash flood susceptibility zoning: A case study of the Karnaphuli and Sangu river basins of Bangladesh. *Nat. Hazards* **99**, 425–448 (2019).
15. Li, Z., Xu, X., Zhu, J., Xu, C. & Wang, K. Effects of lithology and geomorphology on sediment yield in karst mountainous catchments. *Geomorphology* **343**, 119–128 (2019).
16. Youssef, A. M., Pradhan, B. & Hassan, A. M. Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery. *Environ. Earth Sci.* **62**, 611–623 (2011).
17. Zhang, H., Shi, Z., Fang, N. & Guo, M. Linking watershed geomorphic characteristics to sediment yield: Evidence from the Loess Plateau of China. *Geomorphology* **234**, 19–27 (2015).
18. Mark, D. M. Geomorphometric parameters: A review and evaluation. *Geogr. Ann. Ser. B* **57**, 165–177 (1975).
19. Dano Umar, L. *et al.* Geographic information system and remote sensing applications in flood hazards management: A review. *Res. J. Appl. Sci. Eng. Technol.* **3**, 933–947 (2011).
20. Forkuo, E. K. & Tsawo, V. A. The use of digital elevation models for water-shed and flood hazard mapping. (2013).
21. Assaf, A. T., Sayl, K. N. & Adham, A. In *Journal of Physics: Conf. Series*. 012149 (IOP Publishing).
22. Ahmed, A., Hewa, G. & Alrajhi, A. Flood susceptibility mapping using a geomorphometric approach in South Australian basins. *Nat. Hazards* **106**, 629–653 (2021).
23. Mishra, K. & Sinha, R. Flood risk assessment in the Kosi megafan using multi-criteria decision analysis: A hydro-geomorphic approach. *Geomorphology* **350**, 106861 (2020).
24. Sarkar, D. & Mondal, P. Flood vulnerability mapping using frequency ratio (FR) model: A case study on Kulik river basin, Indo-Bangladesh, Barind region. *Appl. Water Sci.* **10**, 1–13 (2020).
25. Hussain, M. *et al.* GIS-based multi-criteria approach for flood vulnerability assessment and mapping in district Shangla: Khyber Pakhtunkhwa, Pakistan. *Sustain.* **13**, 3126 (2021).
26. Percival, S. & Teeuw, R. A methodology for urban micro-scale coastal flood vulnerability and risk assessment and mapping. *Nat. Hazards* **97**, 355–377 (2019).
27. Nandi, A., Mandal, A., Wilson, M. & Smith, D. Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environ. Earth Sci.* **75**, 1–16 (2016).
28. Chang, L.-C., Liou, J.-Y. & Chang, F.-J. Spatial-temporal flood inundation nowcasts by fusing machine learning methods and principal component analysis. *J. Hydrol.* **612**, 128086 (2022).
29. El-Rawy, M., Elsadek, W. M. & De Smedt, F. Flash flood susceptibility mapping in Sinai, Egypt using hydromorphic data, principal component analysis and logistic regression. *Water* **14**, 2434 (2022).
30. Gómez-Palacios, D., Torres, M. A. & Reinoso, E. Flood mapping through principal component analysis of multitemporal satellite imagery considering the alteration of water spectral properties due to turbidity conditions. *Geomat. Nat. Haz. Risk* **8**, 607–623 (2017).
31. Rahman, A. S. & Rahman, A. Application of principal component analysis and cluster analysis in regional flood frequency analysis: A case study in New South Wales, Australia. *Water* **12**, 781 (2020).
32. McMullen, C. Global Environment Outlook 6. (2019).
33. Ritchie, H. & Roser, M. Natural disasters. *Our World in Data* (2014).
34. Al-Quraishi, A. M. F. & Negm, A. M. *Environmental Remote Sensing and GIS in Iraq* (Springer, 2019).
35. Horton, R. E. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geol. Soc. Am. Bull.* **56**, 275–370 (1945).
36. Strahler, A. N. Quantitative analysis of watershed geomorphology. *EOS Trans. Am. Geophys. Union* **38**, 913–920 (1957).
37. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
38. Davis, J. C. & Sampson, R. J. *Statistics and Data Analysis in Geology* Vol. 646 (Wiley, 1986).
39. Hong, H. *et al.* Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. *Sci. Total Environ.* **625**, 575–588 (2018).
40. Taha, M. M., Elbarbary, S. M., Naguib, D. M. & El-Shamy, I. Flash flood hazard zonation based on basin morphometry using remote sensing and GIS techniques: A case study of Wadi Qena basin, Eastern Desert. *Egypt. Rem. Sens. Appl. Soc. Environ.* **8**, 157–167 (2017).
41. Shamy, E. Recent recharge and flash flooding opportunities in the Eastern Desert, Egypt: Possibilités de recharge de nappes et de crues subites dans le Désert Oriental, Egypte. *Ann. Geol. Surv. Egypt* **18**, 323 (1992).
42. Youssef, A., Pradhan, B., Gaber, A. & Buchroithner, M. Geomorphological hazard analysis along the Egyptian Red Sea coast between Safaga and Quseir. *Nat. Hazard.* **9**, 751–766 (2009).
43. Smith, K. G. Standards for grading texture of erosional topography. *Am. J. Sci.* **248**, 655–668 (1950).
44. Schumm, S. A. Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Geol. Soc. Am. Bull.* **67**, 597–646 (1956).
45. Zavoianu, I. *Morphometry of Drainage Basins* (Elsevier, 2011).
46. Bhat, M. S., Alam, A., Ahmad, S., Farooq, H. & Ahmad, B. Flood hazard assessment of upper Jhelum basin using morphometric parameters. *Environ. Earth Sci.* **78**, 1–17 (2019).
47. Prama, M., Omran, A., Schröder, D. & Abouelmagd, A. Vulnerability assessment of flash floods in Wadi Dahab Basin. *Egypt. Environ. Earth Sci.* **79**, 1–17 (2020).
48. Prasad, R., Mondal, N., Banerjee, P., Nandakumar, M. & Singh, V. Deciphering potential groundwater zone in hard rock through the application of GIS. *Environ. Geol.* **55**, 467–475 (2008).
49. Muneer, A. S., Sayl, K. N. & Kamal, A. H. Modeling of spatially distributed infiltration in the Iraqi Western desert. *Appl. Geomatics* **13**, 467–479 (2021).
50. De Roo, A., Odijk, M., Schmuck, G., Koster, E. & Lucier, A. Assessing the effects of land use changes on floods in the Meuse and Oder catchment. *Phys. Chem. Earth Part B* **26**, 593–599 (2001).
51. Zittis, G. *et al.* Climate change and weather extremes in the Eastern Mediterranean and Middle East. *Rev. Geophys.* **60**, 2021000762 (2022).
52. Pachauri, R. K. *et al.* *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, 2014).
53. Mustafa, A. M., Muhammed, H. & Szydłowski, M. Extreme rainfalls as a cause of urban flash floods; a case study of the Erbil-Kurdistan region of Iraq. *Acta Sci. Pol. Form. Circumictus* **18**, 113–132 (2019).
54. Warburton, M. L., Schulze, R. E. & Jewitt, G. P. Hydrological impacts of land use change in three diverse South African catchments. *J. Hydrol.* **414**, 118–135 (2012).
55. Yan, B., Fang, N., Zhang, P. & Shi, Z. Impacts of land use change on watershed streamflow and sediment yield: An assessment using hydrologic modelling and partial least squares regression. *J. Hydrol.* **484**, 26–37 (2013).
56. Kim, Y., Engel, B. A., Lim, K. J., Larson, V. & Duncan, B. Runoff impacts of land-use change in Indian River Lagoon watershed. *J. Hydrol. Eng.* **7**, 245–251 (2002).
57. Calder, I. R. & Aylward, B. Forest and floods: Moving to an evidence-based approach to watershed and integrated flood management. *Water Intern.* **31**, 87–99 (2006).

58. Gado, T. A. & El-Agha, D. E. Feasibility of rainwater harvesting for sustainable water management in urban areas of Egypt. *Environ. Sci. Pollut. Res.* **27**, 32304–32317 (2020).
59. Ezzeldin, M., Konstantinovich, S. E. & Igorevich, G. I. Determining the suitability of rainwater harvesting for the achievement of sustainable development goals in Wadi Watir, Egypt using GIS techniques. *J. Environ. Manage.* **313**, 114990 (2022).
60. Almazroui, M., Islam, M. N., Balkhair, K. S., Şen, Z. & Masood, A. Rainwater harvesting possibility under climate change: A basin-scale case study over western province of Saudi Arabia. *Atmos. Res.* **189**, 11–23 (2017).

Author contributions

A.A.: Conceptualization, methodology, formal analysis and investigation, software, data curation, validation, supervision, writing—original draft, writing—review & editing. A.A.: Conceptualization, methodology, formal analysis and investigation, data acquisition, validation, writing—original draft, writing—review & editing. B.H.: Conceptualization, formal analysis and investigation, data acquisition, validation, writing—original draft, writing—review & editing. D.A.: Conceptualization, methodology, data curation, writing—review & editing. H.A.: Formal analysis and investigation, data curation, supervision, writing—review & editing. A.G.: Formal analysis and investigation, data curation, validation, writing—review & editing. All authors contributed to the article and approved the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-39290-4>.

Correspondence and requests for materials should be addressed to A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023