



Newspapers as a validation proxy for GIS modeling in Fujairah, United Arab Emirates: identifying flood-prone areas

M. M. Yagoub¹ · Aishah A. Alsereidi¹ · Elfadil A. Mohamed² · Punitha Periyasamy³ · Reem Alameri¹ · Salama Aldarmaki¹ · Yaqein Alhashmi¹

Received: 22 November 2019 / Accepted: 4 July 2020
© The Author(s) 2020

Abstract

The UN Office for Disaster Risk Reduction listed 10 reasons businesses should reduce their disaster exposure, including risk factoring, which cannot be achieved without historical data about hazards, their locations, magnitudes, and frequencies. Substantial hazard data are reported by newspapers, which could add value to disaster management decision making. In this study, a text-mining program extracted keywords related to floods' geographic location, date, and damages from newspaper analyses of flash floods in Fujairah, UAE, from 2000–2018. The paper describes extracting such information as well as geocoding and validating flood-prone areas generated through geographic information system (GIS) modeling. The generation of flood-prone areas was based on elevation, slope, land use, soil, and geology coupled with topographic wetness index, topographic position index, and curve number. Analytical Hierarchy Process (AHP) produced relative weight for each factor, and GIS map algebra generated flood-prone areas. AHP inclusion helped minimize weight subjectivity among various experts. Of all areas, 85% are considered medium and low flood-prone zones, mainly mountainous areas. However, the 15% that are high/very high are dominated by urban areas in low coastal plains, predisposing them to flash floods. Eighty-four percent of flood events reported by newspapers were in areas rated as high/very high flood-prone zones. In the absence of flood records, newspapers reports can be used as a reference. Policymakers should assess whether flood-prone area models offer accurate analyses. These findings are useful for organizations related to disaster management, urban planning, insurance, archiving, and documentation.

Keywords Flash floods · Newspapers · Text mining · GIS · Flood-prone area map · UAE

✉ M. M. Yagoub
myagoub@uaeu.ac.ae

Extended author information available on the last page of the article

1 Introduction

Decision makers everywhere are concerned about the safety of their people before, during, and after disasters. Floods top these disasters, and their impact on human lives and economy is well documented in the literature (Alaghmand and Abdullah 2010; Duan et al. 2016; Luino et al. 2016). For example, a recent report by the United Nations Office for Disaster Risk Reduction (UNDRR; formerly UNISDR) revealed that floods affected 2.3 billion and killed 157,000 people between 1995 and 2015 (Flood List 2018). Due to the flood impact on humans and the economy, many flood tools have been developed around the world to aid in prediction and assessment such as the Global Flood Awareness System (GloFAS) of the European Commission Copernicus Emergency Management Service and in the USA, the Hazus program. GloFAS provides flood predictions based on weather forecasts and hydrological modeling (GloFAS 2020). It contains a global dataset of daily river discharges, a land surface runoff model (HTESSEL), and a river-routing model (LISFLOOD) (van der Kniff et al. 2010). For more details about GloFAS, the reader is referred to Harrigan et al. (2020). The United States Federal Emergency Management Agency (FEMA) developed the Hazards United States (Hazus) program to aid in estimating potential losses from earthquakes, floods, and hurricanes. The program utilizes GIS to estimate physical, economic, and social impacts of disasters (FEMA 2020). The FEMA Flood Map Service Center (MSC) provides interactive maps regarding flood hazard/risk to support the National Flood Insurance Program (NFIP). Details about how flood maps are generated can be found in Wright (2014). In addition to this, the UK Environment Agency provides interactive maps that show flood risks from rivers, surface water, and reservoirs (UK Environment Agency 2020). Alexander et al. (2011) developed a geographic information system (GIS) tool that helps end users to be actively engaged in the generation of flood maps and assessment processes. The tool facilitates discussion between end users/professionals and developers. It includes options for flood hazard assessment, vulnerability assessment, and risk assessment. Other initiatives on flood hazard mapping can be found in Prinos (2008) and Van Alphen and Passchier (2007).

Flood impact is evident in hazard areas that sometimes experience heavy rains. In arid regions, such as the United Arab Emirates (UAE), the risk is not that high, but there has been an expansion of such risk into new areas in cities that have become impermeable due to paving (less infiltration, poor absorption, less vegetation, more runoff). Moreover, cities at foothills of mountains in the UAE such as Fujairah, Ras Al Khaimah, and Al Ain are more exposed to flash floods, for example, the Al Ain floods (1982, 1988, 1990, 1993), the Al Qurayah flood (1995), and the Sharm floods (1997, 2009). This has led many researchers to conduct studies related to flash floods in the UAE. Al-Shamsei (1993) conducted a study on flash floods in Al Ain and generated a flood-prone area map based on climate, geomorphology, and hydraulics. The results revealed major damage to in buildings and farms. Al Murshidi (2012) conducted a study on flash floods' assessment using a geographic information system (GIS) in Dibba Fujairah. His study utilized GIS multi-criteria decision analysis and used three parameters to generate a flood-prone area map: the distance to the discharge channels, elevation, and slope. Additional parameters such as land use, soil, and geology, coupled with topographic wetness index (TWI), topographic position index (TPI), and curve number (CN) could be incorporated. The inclusion of more factors strengthens the flood-prone area model. However, previous studies related to flood potential mapping conducted in the UAE (Al

Murshidi 2012; Al-Shamsei 1993) have not verified their output models. This creates a gap in the literature, which this study tries to fill.

A study conducted by Forkuo (2011) showed how flood risk index maps can be generated by integrating ASTER satellite image data with GIS. In the absence of maps on flooded areas, remote sensing is an effective tool to trace the occurrences of floods and identify hydrogeological settings. There is now a better chance of tracing back flooded areas because of the increase in the number of satellites. Quirós and Gagnon (2020) validated flood risk maps using open-source optical and radar satellite imagery. They used normalized difference vegetation index (NDVI) based on the principles that water absorbs the energy in the red band and that clear water has a negative NDVI value (-1) and thus appears black in the processed satellite image. However, in many cases, flooded areas may not be captured by satellites due to temporal resolution, nature of the flood (the deluge caused by flash floods lasts for a few hours), size of the flooded area, or cloud cover (Quirós and Gagnon 2020). Therefore, newspapers reports could be used to trace back flooded areas. Known flood locations have been used to validate the flood-prone area maps. For example, Azizat and Omar (2018) used six factors to model flood hazard including rainfall, elevation, distance from the drainage network, soil texture, geology, and erosion. They evaluated the flood map using frequency ratio, statistical index, and Poisson methods and verified the model results by using 23 known flood locations recorded in 2013, 2014, and 2015. The locations of the points were checked with high and very-high risk zones, and the percentage of points was calculated. They found that the areas with frequent flood occurrence were located near drainage networks.

Periyasamy et al. (2018) used eight parameters to model flood hazard including rainfall, slope, drainage, soil, geological structures, lithology, geomorphology, and land use. They verified their hazard model output based on recorded floods that affected villages from 1996 to 2014. The output revealed that the study area was flooded mainly due to two reasons: aftermath of meteorological events, such as intense and prolonged rainfall, and increase in the water level in coastal and estuarine waters. Getahun and Gebre (2015) used six parameters to model flood hazard in Awash River Basin in Ethiopia. They were slope, elevation, rainfall, drainage density, land use, and soil type. Samanta et al. (2016) used four parameters to generate flood hazard maps in the area around Markham River, Papua New Guinea. They were elevation, slope, distance from river, and land use. In general, there is no consensus on the number of parameters that are required to generate flood hazard maps, and this depends on many factors such as data availability, geographic location of the study area (climate), and importance of the factor. In this study, the most important factors identified were elevation, slope, land use, soil, and geology coupled with TWI, TPI, and CN. Weight for each factor was set using AHP.

AHP is a multi-criteria decision making (MCDM) method (Saaty 1980). It provides a framework that can handle different views on a complex decision problem (Carver 1991; Estoque 2012; Ho 2008; Malczewski 1999; Samanta et al. 2016). AHP consists of three main levels. Level 1 is the goal, while level 2 (criteria/sub-criteria) and level 3 (alternatives/options) contain six necessary steps to get to that goal (Papaioannou et al. 2015). In AHP, multiple pairwise comparisons are based on a standardized comparison scale of nine levels (Ouma and Tateishi 2014). AHP has the advantage of permitting a hierarchical structure of the criteria, which provides users with a better focus on specific criteria and sub-criteria when allocating the weights (Ishizaka and Labib 2011; Saaty 1980). This method helps to discover and correct logical inconsistencies and allows for the “translation” of subjective opinions, such as preferences or feelings, into measurable numeric relationships (Goepel 2018). Due to these advantages, AHP is used in GIS

to improve decision-making processes (Estoque 2012; Samanta et al. 2016). Substantial research has been conducted on the combination of AHP and GIS for flood studies (Dash and Sar 2020; Kazakis et al. 2015; Ouma and Tateishi 2014; Papaioannou et al. 2015; Stefanidis and Stathis 2013).

Despite remote sensing and GIS being sources of data and visualization, they lack detailed information about flood impact/damage on human lives and economy. Historically, people used various other sources such as newspapers reports, flood markers, eyewitness accounts, photographs, private diaries, private letters, boulder deposits, town records, historic compilations, courts records, paintings, and secondary sources (Benson 1950; Helley and Lamarche 1973; Tropeano and Turconi 2004). In Italy, Luino et al. (2016) highlighted that correct territorial planning for the prevention and mitigation of geo-hydrological risk cannot ignore the wealth of information that can be obtained from historical research. They collected historical records including thousands of articles from national and local newspapers from 1800 until 2016 and generated a GIS database. This study, too, uses newspapers to build a model. Previous studies in the UAE did not incorporate newspapers as a source of data to fill the gap in missing information on floods. Newspapers include valuable information about events (what, where, when, how, who are affected) on an often overlooked and relatively unknown time period in history. They disperse information to the general public and provide effective warnings as one of the crucial tools to protect the public from natural hazards (Luino et al. 2018; Trimble 2008; Zhang 2010).

Combining newspaper data with GIS is an effective tool. For example, Yagoub and Jalil (2014) used GIS to map 220 sampled fire incidents in Sharjah, UAE. The incidents were extracted from newspapers (2002 to 2012), geocoded and used with weighted overlay analysis to propose new suitable locations for fire stations. In the United Kingdom, Taylor et al. (2015) used newspapers to enrich the National Landslide Database with 111 landslide events. Abascal-Mena et al. (2013) used the Geoparsing Web Service (Yahoo! Placemaker Web Service) and eXtensible Markup Language (XML) to extract geographic coordinates from online unstructured documents and link them into a map-like image in order to visualize textual information. Geoparsing offers the ability to turn text documents into geospatial databases. Geocoding newspaper reports on flood events in a GIS environment can add additional information to rain gauge stations (Escobar and Demeritt 2014). The use of newspaper reports to enrich data about floods is evident in the literature. Tarhule (2005) used newspapers to explore the nature of flood events in Niger's Sahel region. Escobar et al. (2016) utilized the Nexis digital newspaper database (LexisNexis 2018) to build a 25-year archive of flood events in the United Kingdom. They searched for data using date, region, keywords, and type of publication. Search results could then be filtered and downloaded for analysis. Although the database has proved useful in similar projects (e.g., Taylor et al. 2015), it unfortunately does not include reports on flood events in the UAE.

Recent experiments have used crowd-sourced data (Brisaboa et al. 2010; Kutija et al. 2014) and post-event surveys (Borga et al. 2011) to gather evidence on flood impacts. However, crowd-sourced/social media data lack completeness, consistency, and authorized references. Therefore, newspapers are considered more reliable source albeit with some limitations (Newby and Hardy 2018; Porter and Evans 2020), including errors and discrepancies due to misreported events, unreliable sources, or political slants and biases. This can be minimized by double-checking with more than one newspaper and using quantitative data from other sources (e.g., recorded rainfall by meteorological stations). Other problems that arise while using newspapers include the availability of old archives in the hardcopy format. Manually searching newspapers with long records is a tedious and time-consuming process.

The main objective of this study was to review newspapers reports on flash flood events published in five UAE newspapers during the period 2000–2018. A key component of the objective was automating the process of extracting flood information from newspapers reports using the text-mining approach. The second objective was to identify potential flood-prone areas based on multi-criteria analysis (MCA) (Belton and Stewart 2002; Malczewski 2006) and use flood events reported by newspapers as validation for the flood-prone areas. This flood-prone areas map could help city planners in drafting rules to prevent development in flood-prone zones (UN 2010). The objective of the flood-prone area map is not to stop the disaster, but to highlight possible flood zones that may impact society and economy.

This article is structured as follows: Sect. 2 provides a background about the study area, data, and methodology; Sect. 3 addresses the results; and Sect. 4 provides a general conclusion and recommendations.

2 Materials and methods

2.1 Study area

The eastern region of the UAE is the most prone to natural hazards of the entire country (floods, earthquakes, tropical storms/cyclones) (Al Ghasyah 2010). Therefore, it was selected as the study area. The region is at a strategic international location along the Arabian Gulf from where large volumes of oil is exported (Fig. 1). The offshore zones of the region (Gulf of Oman) represent an important asset for the UAE. This includes water (desalination plants), food (fishing industry), mineral exploration, navigation, research, and the laying of cables and pipelines. These zones include the territorial sea, contiguous zone, and the exclusive economic zone (UNCLOS 2019). The region covers a large inland area that extends along the Gulf of Oman (Fig. 1) and is dominated by the Al Hajar Mountains—the highest range in the eastern Arabian Peninsula that runs through both the UAE and Oman. The highest point in the study area measures 1,511 m above mean sea level. Many valleys (Wadis) originate from this mountain range, such as Ham, Zikit, and Al Wurayah. Wadi Ham is the largest and longest valley in the UAE. It spans 30 km from Masafi to the dam near Fujairah city. Several studies related to Wadi Ham have been conducted (Sherif et al. 2010, 2013). Although the valleys remain dry for most of the year, when it rains, the dry valley beds are transformed into gushing albeit ephemeral spectacles.

Fujairah is the main emirate in the eastern region of the UAE, covering an area of 1,580 square kilometers (1.9% of the UAE) with an estimated population of 243,127 inhabitants, almost 77% of whom live in main Fujairah city (Fujairah Statistics Centre 2018). The average temperature in Fujairah is around 15 and 47 degrees Celsius during winter and summer, respectively, with mean daily maximum relative humidity of 73 and average annual rainfall of 47 mm (Fujairah Statistics Centre 2018). Generally, the area falls in the arid climate zone (100–250 mm rainfall; Ghazanfar and Fisher 1998).

The majority of urban areas in the region are located at low coastal plains, which makes them more prone to flash floods (Fig. 2). To reduce the impact of flash floods, the UAE Ministry of Energy and Industry (<https://www.moei.gov.ae>) in cooperation with the Ministry of Climate Change and Environment (<https://www.moccae.gov.ae>) has adopted guidelines for protection against floods. They include a number of standards to be met when

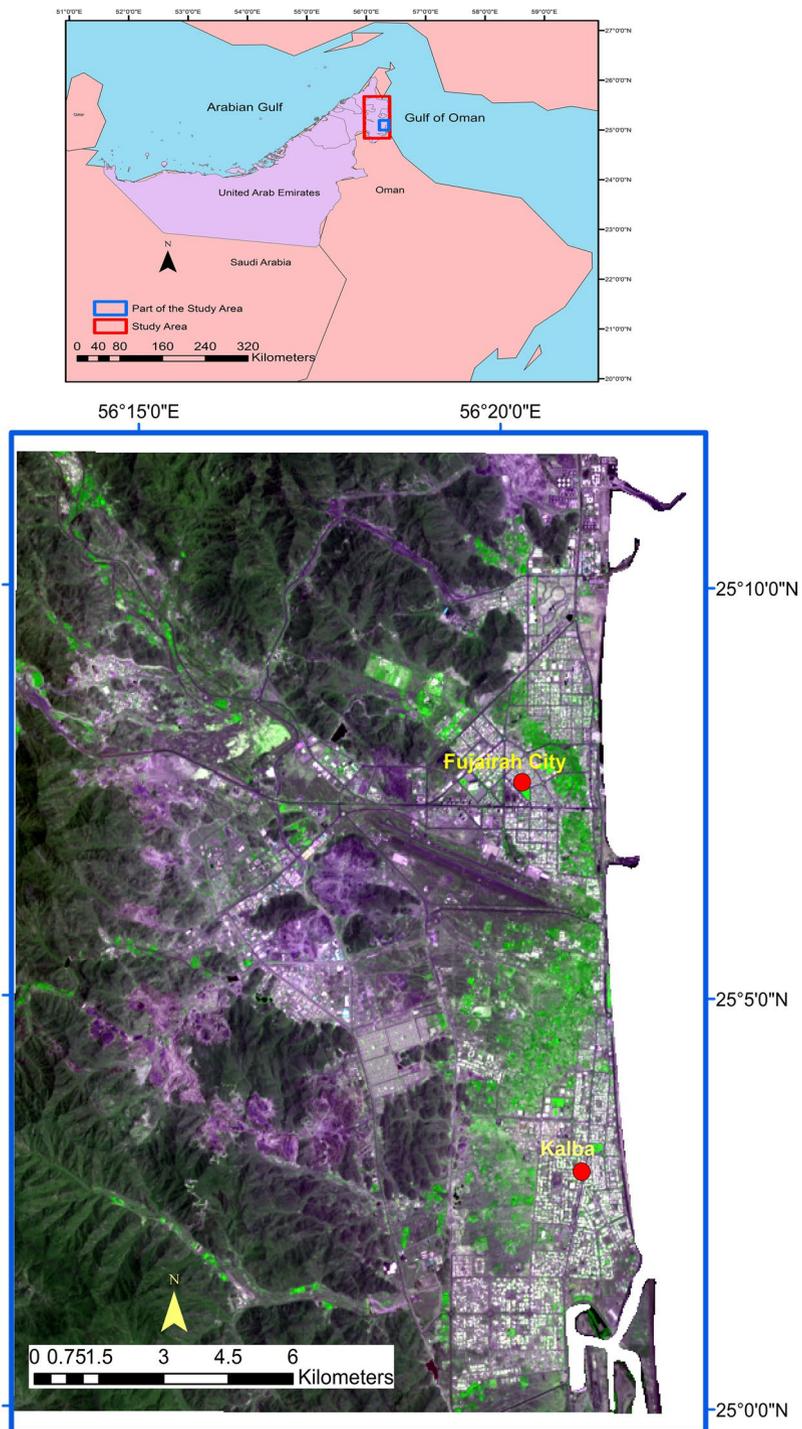


Fig. 1 Location of the study area



Fig. 2 Example of flash flood in the study area 16 Feb 2016. *Source:* https://twitter.com/Storm_centre

designing or developing any residential area close to valleys and estuaries. The identification of flood-prone zones from this study will support these guidelines.

2.2 Data

Data about flash-flood-related events were collected from local newspapers, including Albayan (Arabic; <https://www.albayan.ae>), Alittihad (Arabic; <https://www.alittihad.ae>), Al Khaleej (Arabic; <https://www.alkhaleej.ae>), Gulf News (English; <https://gulfnews.com/uae>), and Khaleej Times (English; <https://www.khaleejtimes.com>). Digital elevation model (DEM) and Landsat 8 satellite images for the study area were downloaded from the United States Geological Survey (USGS)—Earth Explorer website (<https://earthexplorer.usgs.gov/>). The land use map was produced from Landsat-8 images (2020). The soil map for the study area was extracted from the 1 km-resolution soil map of the Harmonized World Soil Database (HWSD 2009). Monthly rainfall data (2009–2018) were obtained from the Department of Civil Aviation and World Weather (2019). Table 1 shows the data collected to identify flood-prone areas. It should be noted that rainfall is treated as constant because

Table 1 Data and sources

	Data	Source	Format
1	DEM	USGS—Earth Explorer website	Raster
2	Valleys/Streams	Derived from DEM	Vector
3	Landsat 8 satellite image acquired on 25/4/2020 Path 159, Row 43	USGS—Earth Explorer website	Raster
4	Land use	Derived from satellite image	Vector
5	Soil	Harmonized World Soil Database	Raster
6	Geology	UAE Ministry of Environment and Water	Vector
7	Monthly rainfall data (2009–2018)	UAE Department of Civil Aviation and World Weather	Text
8	Historical flash flood events	Local newspapers	Text

the study area is small. The question here is that if the overall region is subjected to the same amount of rainfall, which areas are prone to flood? The monthly rainfall data collected were used to validate newspapers reports and not as a criterion for identifying potential flood-prone areas.

2.3 Method

2.3.1 Extraction of newspaper reports

Communication between various departments was established in order to obtain records of previous flood events, but without success. Therefore, Google's free search engine was used. The Google search results were saved in the comma-separated values (CSV) file format using the SEOquake extension (<https://www.seoquake.com/>), a free plug-in that helped save search results (websites) from search engines. This format was chosen because of its flexibility in importing and exporting data between various software. For example, the search for "UAE, Fujairah, flash floods" generated a listed of websites, which were downloaded in a CSV file. Normally, Google search lists 10 results for faster processing. However, saving every list of 10 websites in a file is not feasible; therefore, the Google search setting was changed to list 100 results per page. Google Search also provides an important setting for the region. The websites generated by the Google search were examined thoroughly, and the majority were found to be related to local newspapers. Therefore, the focus was directed toward local newspapers, which were found to have more detailed reports and coverage about flash floods in the UAE than other international sources, such as the FloodList web site (<https://floodlist.com>) and the Global Disaster Alert and Coordination System (GDACS) (<https://www.gdacs.org/>). Escobar et al. (2016) also revealed that regional newspapers in the United Kingdom tended to provide more detailed impact information than did the national press and wire services.

Newspaper content was reviewed for news items covering flash flood events in Fujairah Emirate between 2000 and 2018. This period was chosen because of the availability of digital editions of the newspapers. Taxonomy ("a set of words that have been organized to control the use of terms") was used to facilitate retrieving flood items from newspapers reports (Escobar and Demeritt 2014). A different Boolean search was conducted (e.g., by inputting the keyword "flood" OR "rain" OR "flash flood") to extract event-related articles along with the date and geographical filters (geoparsing). The reports were saved in document files with a date stamp and included information about the flood date, area affected, and damage caused. The main issue was to determine how to extract meaningful information from the reports and put them in a structured (tabular) format that would be easy for decision makers. Manual data entry from a document file to a structured format (e.g., Excel table) is a tedious, error-prone, costly, and time-consuming process especially when the actual data are non-structural (text report). Moreover, reading many document files and scanning them for flood impact is another challenge. To automate this process, a Java program was developed to read the document file and extract important flood damage information (text mining) using text-processing functions like keyword extraction, segmentation, parsing, and identification of contexts and similar terms.

The program was run on sample English documents by feeding keywords such as disaster words (flood, flash flood, rain, heavy rain, etc.), name of the location, damage caused (e.g., road blocked, house damaged), and evacuation procedures. This was supported by a database of geographic names in the area and link to the data extracted from the

newspapers. However, it was difficult to find uniform newspapers reports on which to run the program. Therefore, data from some reports were extracted manually. The program's algorithm works as follows:

- Creating a text file (CSV) with the following headers “Disaster type,” “Date,” “Location,” and “Impact.”
- Checking if the title/headline of the newspaper report or the body text includes (flood “OR” flash flood “OR” rain “OR” heavy rain, etc.); if yes, moving on to the next step otherwise terminating the search.
- Checking the date and if the month falls during the rainy season (October–April) copying the date to a text file (CSV file); otherwise, writing a message for manual intervention.
- Searching for the name of the area/location based on names given in the database of geographic names. Copying the name to the text file.
- Searching for impact keywords (damage, houses, buildings, roads, farms, etc.)
- Extracting a complete sentence that includes the impact keywords (e.g., “400 houses were damaged”) and copying it to the text file under the header “Impact.”

“Appendix” shows the flowchart of the program and a sample of the Java code.

The keywords in flood reports were first extracted/summarized using Text Analytics software. There are many text analytics and natural language software available in the market such as Natural Language Toolkit (<https://www.nltk.org>), KH Coder (<https://kncoder.net/en/>), and RapidMiner (<https://rapidminer.com/get-started>). In this study, RapidMiner was selected because it is an open source data-mining framework that offers many operators that can be combined into a process. Text from different data sources can be loaded and transformed by different filtering techniques to analyze data. Text Extension includes all operators necessary for statistical text analysis and natural language processing (NLP). It supports several text formats and provides standard filters for tokenization, stemming, and stop-word filtering for preparing and analyzing texts. The collected newspapers report about flash floods were converted from Word files to plain text and the occurrence of keywords in all documents was analyzed.

2.3.2 Factors for identifying flood-prone areas

Factors for delineating flood-prone areas were identified based on a literature review, consultation with five experts, and availability and accessibility to data. The factors include elevation, slope, land use, soil, and geology coupled with TWI, TPI, and CN.

Elevation and slope: The probability of a flood increases with decreasing elevation and hence is a strong indicator for flood susceptibility. Slope is the rate of change of the surface in horizontal and vertical directions (Cabrera and Lee 2020). Areas with greater slope will have more runoff and thus a higher runoff co-efficient (Chang and Guo 2006). In areas with low slope (flatter terrain), there is a substantial decrease in the movement of water and an increase in the likelihood of ponding (Asare-Kyei et al. 2015; Ballerine 2017). Areas located close to valleys are more likely to flood (Cabrera and Lee 2020). The study area contains a series of valleys that are dried out most of the year. Valleys were overlaid on the final maps to improve visualization.

Land use: Land use determines how much rainfall infiltrates the soil and how much becomes runoff (Asare-Kyei et al. 2015). Land use for infrastructure such as buildings

and roads (impervious surfaces) decreases penetration capacity of the soil and increases the water runoff, while vegetation act as a water storer (Asare-Kyei et al. 2015). Rapid urbanization has contributed significantly to the frequency and size of flood events in arid and semi-arid regions such as the UAE (Al Alabdouli et al. 2019). The land use/land cover map was derived from a satellite image (April 25, 2020) using supervised classification. Accuracy of the classification was checked with 500 random ground truth samples, and a confusion matrix was created. The availability of satellite image helped in the generation of up-to-date land use data and is considered an important input parameter.

Soil: The soil type controls the amount of water that can infiltrate the ground and hence the amount of water that becomes flow (runoff). The dominant soil groups found in the area are Calcisols, Leptosols, and Solonchaks. Calcisols is soil with accumulation of secondary calcium carbonates (USDA Group A). Leptosols is very shallow soil over hard rock or in unconsolidated gravelly material, and Solonchaks is strongly saline soil (USDA Group B). The soil map was reclassified into two main soil hydrological groups (A and B) based on the United States Natural Resources Conservation Service classification method (USDA-NRCS 1986). Group A soils have low runoff potential and high infiltration rates even when thoroughly wetted. Group B soils have moderate infiltration rates when thoroughly wetted (USDA-NRCS 1986).

Geology: The lateral movement of storm water is restricted to the weathered formations of the top zones, whereas the downward movement of the same looks for the fissured and jointed features of the bottom zones (Cao et al. 2016; Periyasamy et al. 2018). Therefore, inclusion of geology in flood modeling, especially in areas where soil is in a development stage, is important. The majority of the area is covered by ophiolite mountain (igneous and metamorphic rocks). The main rock types in the study area include Gabbros, Ultrabasics, Fluvialite deposits, Limestone Facies, and Metamorphics (UAE Atlas 1993).

TWI: TWI, also known as the compound topographic index, is an indicator of the effect of local topography on runoff flow direction and accumulation (Beven et al. 1979; Wolock and McCabe 1995) (Eq. 1). It shows how water interacts with the topography. The result is a GIS data layer (raster) that depicts areas with drainage depressions where water is likely to accumulate (Ballerine 2017). The TWI provides important information at a very low cost compared to detailed hydrologic and hydraulic studies and is an excellent planning tool (Ballerine 2017).

$$TWI = \ln(a / \tan \beta) \quad (1)$$

where \ln is the Napierian logarithm, a is the upslope area per unit contour length (catchment area), and $\tan \beta$ is the slope gradient.

TWI was generated from flow direction, flow accumulation, and slope. The output was smoothed to reduce local variations and remove noise. The file was reclassified into five classes using the natural break breaks (Jenks) method.

TPI: TPI compares elevation of each cell to the average elevation around that cell within a predetermined radius (Weiss 2001). Positive TPI values indicate higher areas (ridges), lower values represent locations that are lower than the surrounding areas (valleys), and values near zero are flat areas (Jenness 2004). TPI is used for landform classification, and this helps in the identification of flood-prone areas (slope, drainage network) (Al-Husban 2019; Jenness 2011). TPI was produced from DEM to classify the landscape into slope positions (valleys, lower slope, flat slope, upper slope, ridge). The TPI tool developed by Jenness (2011) and modified by Dilts (2015) was used.

CN: CN is an empirical parameter used in hydrology for predicting direct runoff or infiltration from rainfall excess (Papaioannou et al. 2015; USDA-NRCS 1986). CN is based on hydrologic soil group, land use, and soil moisture conditions. CN values range between 0 and 100 (Table 2). A high curve number means high runoff and low infiltration (urban areas), whereas a low curve number means low runoff and high infiltration (dry soil) (Zhan and Huang 2004). The soil and land use layers were merged and attributes such as soil type (hydrological soil group), land use type, and CN for each land use type were added (Table 2). The CNs were extracted from USDA-NRCS (1986).

All the data were geo-referenced to a common coordinate system (WGS 84, UTM, Zone 40), edited for quality assurance and quality control, and organized in the proper format. ArcGIS version 10.5.1 software was used to process the data.

2.3.3 AHP: expert scores of criteria and weights

Five experts (one each in modeling, climate, urban planning, soil, and geology) were surveyed to give their judgments regarding flood causal factors. Based on discussions with experts and preliminary maps produced, some factors such as dams and valleys (streams) were excluded from the criteria because their effect is included in the slope and TWI. Therefore, the factors considered in the criteria were elevation, slope, TWI, TPI, CN (land use, soil), and geology. The experts were asked to make pairwise comparison among the factors and sub-factors (two hierarchical levels of factors) based on the fundamental scale of AHP (1–9). The number of comparisons was calculated using Eq. 2 (Teknomo 2020). Comparison matrices were created, and consistency ratio (CR) was calculated. The CR is defined as ratio between Consistency Index (CI) and Random Index (RI) (Saaty, 1980, 1990) (Eqs. 3, 4). The experts were allowed to change their scores until CR was below 10% as recommended by Saaty (1990), Papaioannou et al. (2015), and Wu et al. (2015).

$$\text{Number of comparisons} = \frac{n(n - 1)}{2} \tag{2}$$

where n is the total number of criteria.

$$\text{CR} = \frac{\text{CI}}{\text{RI}} \tag{3}$$

$$\text{CI} = \frac{\lambda_{\max} - n}{n - 1} \tag{4}$$

Table 2 CNs for the land use/cover in the study area

Land cover/use	CN for Hydrologic Soil Group A	CN for Hydrologic Soil Group B	Note
Built-up	77	85	Town houses
Green areas	39	61	Parks, greenways
Gravel/sand	63	77	Natural desert
Rocks	98	98	Impervious

Source USDA-NRCS (1986)

where λ_{\max} is the maximum eigenvalue of the comparison matrix (sum of the consistency measure), and n is the total number of criteria (Saaty 1980).

Pairwise comparison judgments (priorities) assigned by the five experts were combined using the geometric mean to form a judgment for the group (Saaty and Shang 2007). The basis for using the geometric rather than the arithmetic mean to combine judgments of different individuals has been justified mathematically by Saaty (1980). The same method was applied to find priorities of the sub-factors (Stefanidis and Stathis 2013). Examples of software that can be used to calculate the AHP parameters include Super Decision (<https://www.superdecisions.com/>), Expert choice (<https://www.expertchoice.com/2020>), and an extension developed by Oswald Marinoni (2004). In this study, the online free AHP tool developed by Goepel (2018) was used (<https://bpmg.com/ahp/ahp.php>).

The following weights and rates set by the experts (AHP) were assigned to the GIS layers: elevation, slope, geology, TWI, TPI, and CN. The GIS layers were then converted to integer type raster format and reclassified to a common scale (1 to 5) with the highest rating of 5 implying extreme probability to flood and the rating of 1 implying low probability. The final map was also created on a fixed 1 to 5 scale based on the Index for Risk Management (INFORM; Marin-Ferrer et al. 2017).

The newspapers' reports about previous floods were geocoded as point features (latitude, longitude) and used as a validation to the flood-prone areas map. This was done by overlaying previous flood events on the map and checking the frequency (number) and percentage of the events on each zone.

3 Results and discussion

3.1 Extraction of flood events from newspaper reports

Search engines within newspaper websites were not effective despite relevant material in the archives. For example, Albayan and Alittihad did not show results when a search was conducted for floods in Fujairah while Google Search generated results and provided links directed to the newspapers themselves. Khaleej Times provides an advanced search option based on keywords and date from 2003 onward. It has more reports (192) than other newspapers about flood events in the UAE. Information extracted from the newspaper reports include date, area affected, and impact (damage). Results from the text-mining model in RapidMiner text analytics software showed the words with the number of appearances in the text. In 10 newspaper reports tested, the word "Fujairah" appeared 59 times and "flood" appeared 52 times (Table 3). The occurrences of the words provided a guide to the words that could be included in the flood dictionary.

Table 4 lists samples of flash flood events that were extracted from newspapers reports, organized in a structured format for easy review by decision makers. On average, flash floods were reported once a year in newspaper reports. Newspapers are key to understanding flood events and their impact, as they provide a clear picture about the damage caused. This includes number of people evacuated, accidents on roads, transport disruption, landslides, as well as damage to property, farms, and power lines (Table 4). These reports informed the readers about the events and their consequences and some of them provided references to previous floods and content to improve public awareness on how to minimize flood impact and adopt preventive measures. Newspapers also revealed damages not commonly reported in other sources: for instance, the destruction of the trade fair in Al Bedyah

Table 3 Sample of words with highest total occurrences in 10 newspapers reports

	Word	Total occurrences	Document occurrences
1	Fujairah	59	10
2	Flood	52	10
3	Rain	45	9
4	December	26	3
5	Police	21	6
6	Water	19	6
7	People	17	6
8	Road	17	5
9	Evacuate	14	4
10	Resident	14	6
11	Damage	13	6
12	House	13	6

popular market on April 18, 2003, resulting in losses of up to AED 150,000 and the destruction of communication towers (Al Khaleej 2013). Although the majority of newspapers referenced in this study are well-known, errors in reporting and estimating damage could not be ruled out. It was thus better to verify whether the event had been reported by more than one newspaper. Regardless of their accuracy, newspapers filled an important information gap in terms of the difficulty of obtaining official records. The organization of data in a tabular form (Table 4) showing locations/space (*where*), temporal info (*when*), *what* and *who* are affected provides better visualization about flood incidences, which may help policymakers in formulating flood-prevention strategies by taking decisions based on actual data.

Newspaper reports triggered policies that came into effect post the Al Qurayah flood (December 11, 1995; Alittihad 2012) based on a previous study conducted by Al Ghasyah (2010). Al Qurayah is small town located in the northern part of Fujairah Emirate with a population of 5026 and 414 houses. The flood occurred due to heavy rains over three days and the failure of a dam in Safad valley (second biggest valley in Fujairah). It damaged buildings, roads, and farms; resulted in the loss of poultry and livestock; and led to many people being evacuated from their houses, although no loss of life was reported (Al Ghasyah 2010). The UAE government allotted a budget of \$4 million to rebuild the dam, and three breakwaters are being built in the Safad and Thayb valleys. New houses for the Al Qurayah population were built and compensation was paid to those affected (Al Ghasyah 2010). This case provides proof of the flood damage and its economic consequence.

Fujairah main city showed a high frequency of flood occurrence (31.4%) (Table 5). This may be due to many reasons such as the low elevation and the steep slope of the city at the mouth of Ham Valley. Moreover, the large city area makes it more vulnerable to floods in comparison with other small towns. The frequency metric should be interpreted by taking into account the causes and impact or “damage.” For example, the flood frequency et al. Qurayah is only 5.7%, but its impact is the biggest (evacuation of 5,000 people and damage to houses) (Tables 4, 5). Of all the flood events reported, 80% of them occurred during the winter and rainy seasons (October–April); this is consistent with historical rainfall records (Fig. 3). There is no annual flood probability detected from the data.

Table 4 Sample of flash flood events extracted from newspaper reports

No	Date	Location/area affected	Impact	Source newspaper
1	April 18, 2003	Dibba + Sharm + AlAqah	Accidents on roads and damage to farms and some of the private property of residents, power outage, telephone lines were affected	Albayan
2	November 12, 2013	Alfasel Area + AlIawyeen	Houses were affected and trees fell	AlIitihad
3	March 3, 2016	Eastern region of Fujairah	Unstable soil condition, road slide on the Fujairah–Khor Fakkan road, damage to private and residential properties, car damage, wall collapse	Gulf News
4	December 18, 2017	Sakamakam- Fujairah	600 people evacuated to school in Sakamakam- Fujairah	The National

Table 5 Sample of spatiotemporal variation of flood frequency

	Location	Flood frequency at the location between 2000 and 2018	Occurrences (%)	Month	Number of flood occurrences	Flood occurrences (%)
1	Al Aqah	1	2.9	January	2	10
2	Al Bedyah	2	5.7	February	0	0
3	Al Qurayah	2	5.7	March	4	20
4	Al Tawyeen	2	5.7	April	2	10
5	Dhadna	1	2.9	May	0	0
6	Dibba	4	11.4	June	0	0
7	Fujairah	11	31.4	July	0	0
8	Masafi	5	14.3	August	1	5
9	Murbeh	1	2.9	September	1	5
10	Qadfa	2	5.7	October	2	10
11	Sharm	3	8.6	November	4	20
12	Port of Fujairah	1	2.9	December	4	20
		35	100		20	100

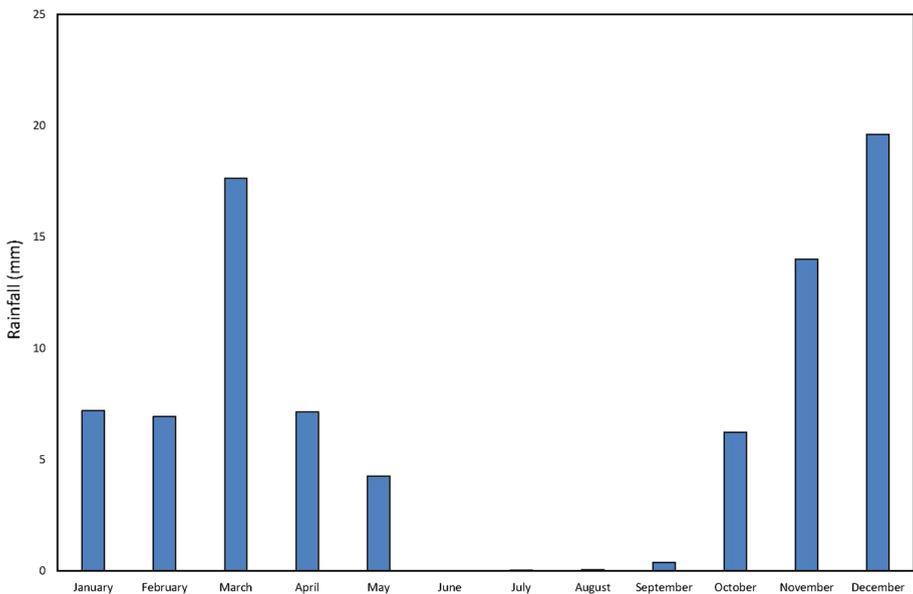


Fig. 3 Average rainfall 2009–2018 at the study area. *Source:* Department of Civil Aviation and World Weather

Misspelled town names (e.g., Al Qurayah and AlQaryah) reported by newspapers was one of the problems faced in this study. Awareness thus needs to be spread among newspapers organizations to use standardized geographical names adopted by the municipality in each emirate. Differences in the reported number of “affected people” were also noticed;

for example, the National newspaper reported that 600 people were affected in the flash flood on December 18, 2017, while FloodList.com reported that 188 people were displaced by the flood. Porter and Evans (2020) cited many cases where there was a difference in number of deaths reported by post-disaster news reporting and official figures.

It is very clear that newspapers have played an important role in documenting flash floods in Fujairah and could be used with other sources to reconstruct their history. They also reach a large audience (not only the scientific community) and attract the attention of decision makers who may act quickly to reduce the flood impact or manage the disaster and its magnitude.

3.2 Land use classification accuracy

Four land cover/use types were identified from the satellite image using supervised classification (Table 6). It was noted that similar rocks have different reflectance due to the effect of slope (shadow), and this was taken into account when delineating training signatures. The lowest user's accuracies were in built-up and green areas (Table 6). This may be due to mixed pixels where some buildings include trees and gardens and some green areas include buildings. The overall classification accuracy of the land use map is 89% with a kappa of 0.85 (Table 6). A study conducted by Yagoub and Kolan (2006) resulted in a classification accuracy of 88% and Thomlinson et al. (1999) set a target of an overall accuracy of 85% with no class less than 70% accurate. Therefore, the accuracy obtained is consistent with previous work and meets the typically suggested classifications accuracy of 85% (Foody 2002).

3.3 AHP results

Table 7 shows a sample of experts' pairwise comparisons of the decision matrix. The values close to one have minimum flood probability and values close to nine have maximum flood probability. The group is homogenous in the majority of the comparisons (Table 8). The CR for all experts ranges between 2.2% and 4.0%, which is acceptable because it is below 10% (Saaty 1990; Papaioannou et al. 2015; Wu et al. 2015). Overall, the CR indicates that the evaluations are sufficiently consistent. The derived weights for the factors (geometric mean) show that elevation, slope, and TWI have the highest weights (Table 8). Cabrera and Lee (2020) found that the weights for criteria in the AHP method are 42% for rainfall, 23% for slope, 15% for elevation, 10% for distance to the main channel, 6%

Table 6 Confusion/error matrix

Class	Green area	Built-up	Rocks	Gravel	Total	User's accuracy
Green area	107	7	5	6	125	0.856
Built-up	8	109	4	4	125	0.872
Rocks	3	5	114	3	125	0.912
Gravel	3	2	5	115	125	0.920
Total	121	123	128	128	500	0.00
Producer's accuracy	0.884	0.886	0.890	0.898	0.0	0.890
Kappa = 0.85						

The overall classification accuracy is given in bold

Table 7 Sample of pairwise comparisons of the decision matrix

Factor	Elevation	Slope	TPI	TWI	CN	Geology
Elevation	1	1	2	2	1	8
Slope	1	1	1	1	2	7
TPI	0.50	1	1	1	2	6
TWI	0.50	1	1	1	2	6
CN	1	0.50	0.50	0.50	1	2
Geology	0.13	0.14	0.17	0.17	0.50	1

Table 8 Priorities (weights) of factors/criteria

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Geometric mean (weight)
Factor	Priority	Priority	Priority	Priority	Priority	Priority
Elevation	0.260	0.257	0.204	0.197	0.190	0.220
Slope	0.213	0.216	0.220	0.213	0.208	0.214
TPI	0.187	0.168	0.192	0.223	0.195	0.192
TWI	0.187	0.189	0.214	0.195	0.240	0.204
CN	0.119	0.133	0.133	0.134	0.138	0.131
Geology	0.035	0.036	0.037	0.038	0.028	0.035
Total	1.0	1.0	1.0	1.0	1.0	1.0
CR	3.9%	4.0%	2.6%	3.6%	2.2%	

for drainage, and 4% for soil type. Results by Danumah et al. (2016) found that slope was the most significant causative factor in flood occurrence. This implies that local conditions play an important role in determining which areas are more vulnerable to flood, and, consequently, weight for the same flood factor could vary from one place to another (Getahun and Gebre 2015; Kazakis et al. 2015; Papaioannou et al. 2015). Geology has the lowest weight, and this may be due the steep nature of the area, where water drains very fast toward the sea and hence has a low chance of percolating down. A study by Dash and Sar (2020) also concluded that geology is the least significant factor in flood generation. The weights of the factors and sub-factors (Table 9) were used to generate the potential flood map.

3.4 Flood-prone areas map

Based on the criteria specified in Table 9, a flood-prone area map was generated (Fig. 4). Almost 85% of the area falls within medium, low, and very low zones (Table 10). This is because a large part of Fujairah (85% of the area) is covered by mountains with elevation greater than 80 m above mean sea level (Table 11, Fig. 5). The high elevation leads to a quick response time of the watershed (Al Alabdouli et al. 2019), and this makes flat areas more prone to flash floods. The rest of the area (15%) falls within high-prone zones (Table 10). This includes low areas facing the Gulf of Oman, where population centers are spread out, such as in Fujairah city and Khor Fakkan. The average elevation in these

Table 9 Weights of the factors considered in generating the flood-prone area map

Factor	Feature Category	Weight sub-factor	Rate	Weight	Factor	Feature category	Weight sub-factor	Rate	Weight
Elevation (meters)	0–20	0.311	5	0.220	TWI	3.2–5.6	0.125	1	0.204
	20–40	0.219	4			5.6–6.9	0.144	2	
	40–60	0.188	3			6.9–8.4	0.193	3	
	60–80	0.166	2			8.4–10.4	0.250	4	
	> 80	0.116	1			10.4–21.1	0.288	5	
Slope (degrees)	0–5	0.256	5	0.214		39	0.169	2	0.131
	5–10	0.222	4			61	0.219	3	
	10–15	0.222	3			77	0.220	4	
	15–20	0.172	2			85	0.275	5	
	> 20	0.128	1			98	0.117	1	
TPI	–209 to –45	0.216	3	0.192		GA*	0.205	4	0.035
	–45 to –15	0.250	4			UL	0.180	3	
	–15 to 13	0.310	5			FL	0.385	5	
	13 to 55	0.135	2			LM	0.127	2	
	55 to 214	0.089	1			MT	0.103	1	

*GA: Gabbros, UL: Ultrabasics, FL: Fluviatile, LM: Limestone, MT: Metamorphics

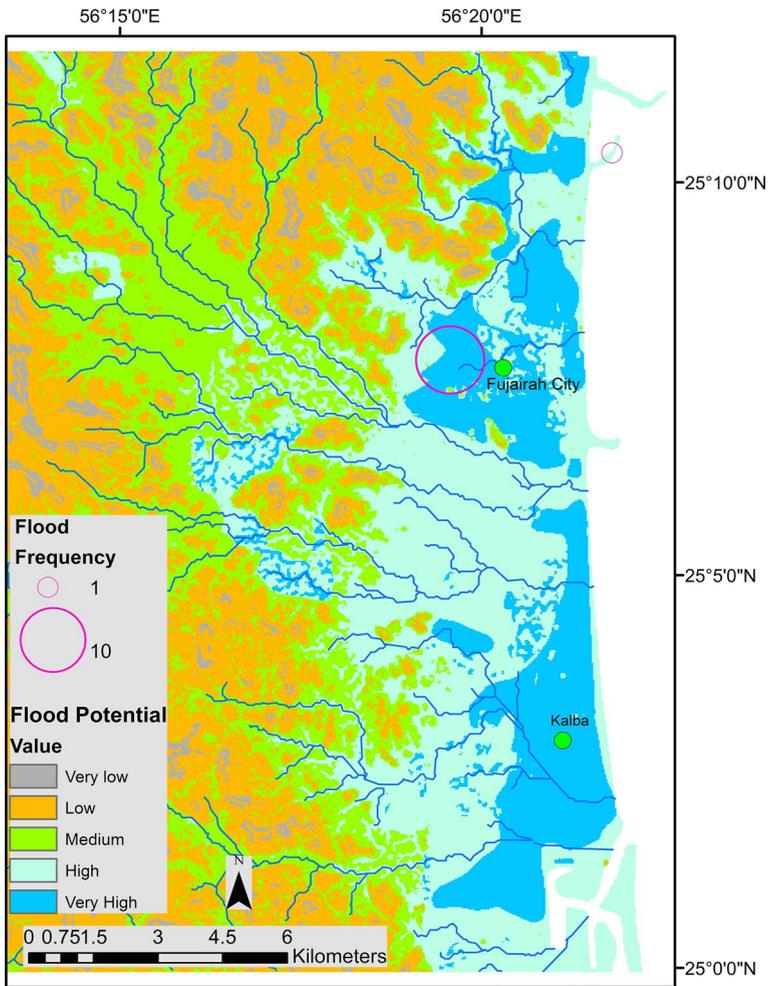


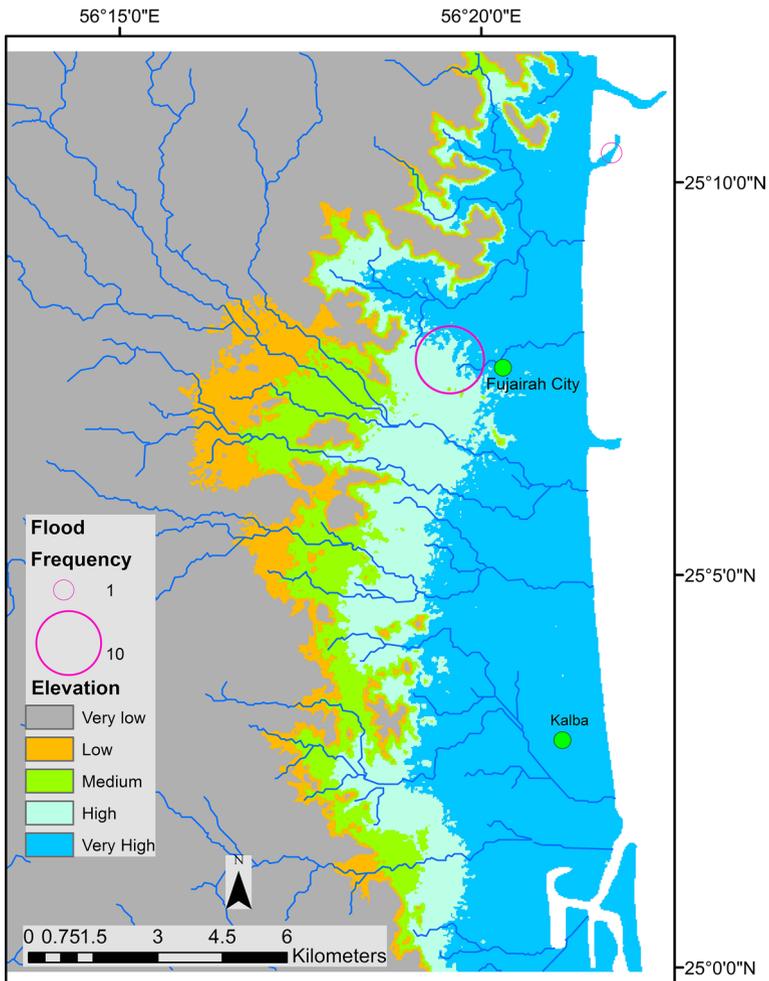
Fig. 4 Flood-prone area map for part of the study area

Table 10 Validation of potential flood-prone areas

Value	Flood potential	Area (km ²)	% of the area	Frequency of previous flood events in the zone	% of checked flood events
1	Very low	279	8.5	25	13.1
2	Low	1194	36.5	0	0
3	Medium	1299	39.7	5	2.6
4	High	421	12.8	35	18.3
5	Very high	81	2.5	126	66.0
	Total	3274	100	191	100

Table 11 Areas based on elevation

Elevation (meter)	Area (km ²)	Percentage of the total area (%)
0–20	218.09	6.66
20–40	123.73	3.78
40–60	86.35	2.64
60–80	72.10	2.20
> 80	2773.73	84.72
	3274	100

**Fig. 5** Elevation for part of the study area

population centers is 50 m. Slope is considered to be an important predictor of surface runoff (Vaezi et al. 2017). Based on slope, around 50% of this area has high and very high potential for flooding. The area includes regions in proximity of valleys and low elevation. Slope alone is not a good indicator for flash floods in the study area because it includes areas close to valleys at high elevation, and these areas are rarely flooded. Since slope is derived from elevation, most of the variance within the study area can be explained by the elevation (Fig. 4).

The high probability results matched the TWI results (Fig. 6). Therefore, TWI could be used alone to provide an indication about flood-prone areas (Ballerine 2017). This is because elevation, slope, flow direction, and flow accumulation are included in calculations of TWI. Papaioannou et al. (2015) found that the most important index/factor is the TWI, followed by the CN. In this study, the CN did not show a good match with the final flood-prone areas. This may be due to the broad classification of land use and the

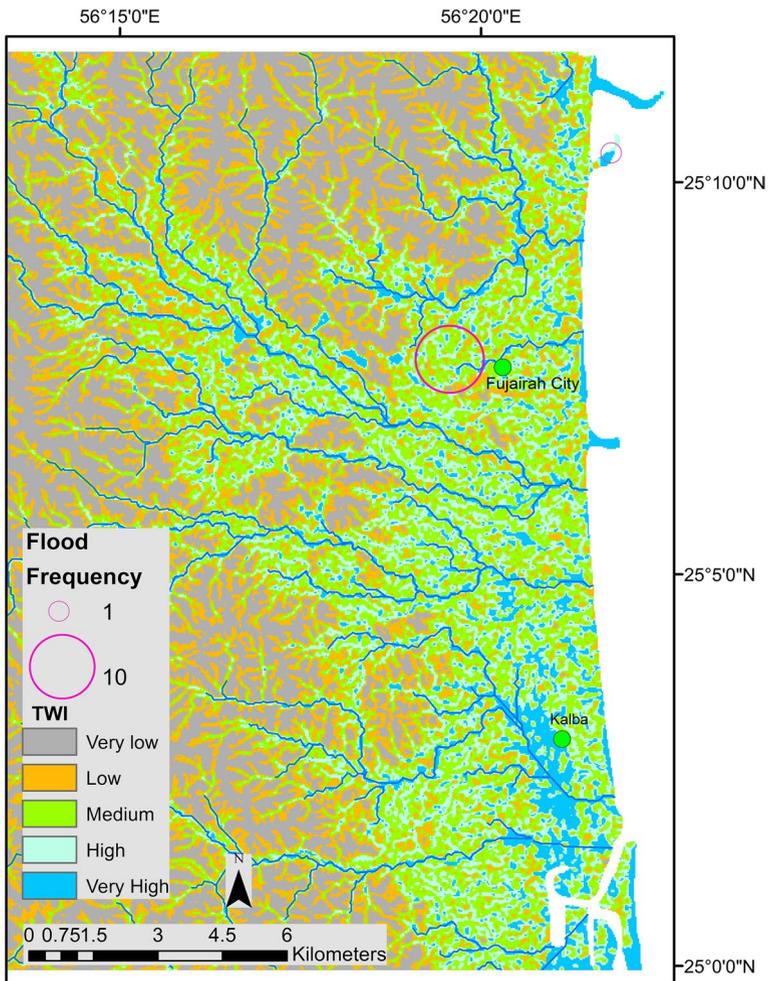


Fig. 6 TWI for part of the study area

alteration of soil characteristics in urban areas. A detailed breakdown of built-up infrastructure, such as commercial and residential areas, pavement, and so on, will enhance specification of CN for land use and consequently improve the output results.

Accuracy of the flood-prone area map (Fig. 4) was verified by previous flood events reported by the newspapers, 84.3% of which were found in areas with high potential for flooding (Table 10). This indicates that there is good agreement between the GIS model output and previous flood events. In the absence of a mechanism for checking GIS flood models, the results obtained are considered a step forward. However, it should be noted that the recorded events lack important information such as flood extent, water depth, and flow velocity (Aronica et al. 2002; Bates 2004; Molinari et al. 2017). Therefore, the output flood-prone area map should be combined with experts' opinions. It is recommended that crowd sourcing, security cameras, sensor network systems, and unmanned aerial vehicles/drones be used to record the extent of floods (Chang and Guo 2006; Feng et al. 2015).

The flood-prone area map provides useful information that can be effective in planning flood defenses and ensuring the safety of those living in high-prone areas. For instance, because it revealed that more prone areas are along the low coastal line, which unfortunately are areas with high population density, preventive measures—such as incorporating planning rules, avoiding development near high-prone zones, building of embankments, and the spreading of public awareness and early warning systems—can be taken. Building more storage dams in addition to the existing ones will help in saving water, recharging groundwater, and protecting the public from flash floods (Alabdouli et al. 2019; Rizk and Al Sharhan 2003). The Ministry of Environment and Water constructed 113 recharge and storage dams to utilize the estimated 150 million m³ per year from 15 main catchment areas (Ministry of Energy, United Arab Emirates, 2006). An early-warning flood system and interactive maps will be vital for reducing flood impact. Examples of these maps are those developed by the Federal Emergency Management Agency (FEMA)-USA Flood Map Service Center (MSC) (<https://msc.fema.gov/portal/advanceSearch#searchresultsanchor>) (FEMA 2019) and the Global Flood Awareness System (GloFAS) of the European Commission Copernicus Emergency Management Service (www.globalfloods.eu/). Moreover, remote sensing data could be utilized to complement rain gauges for precipitation estimation in real time (Isma'il and Saanyol 2013; Legates 2000; Schumann et al. 2018; Wehbe et al. 2017).

As in any GIS work, the possible inaccuracy of results is inherited from both attribute and positional errors, as a result of employing maps/images obtained in various times and with various scales. There are possible errors in land use classification as a result of generalization. For example, some built-up areas include green areas although they are classified as built-up using the law of majority. Elevation data are based on the Shuttle Radar Topography Mission (SRTM). The resolution of the cells of the source data is only three-arc second data (approximately 90 m). The SRTM data are generally not applicable for flood modeling (Finn 2008) and were only applied in this study because no other terrain data were available. Moreover, classification of elevation is based solely on bare ground, which is misleading. For example, low land is assumed to have a high probability of flooding, but this may not be true if the area has a good drainage system (especially in urban areas). The flood-prone area map will change over time due to the development of urban areas and roads or because of climate change (FEMA 2019). Therefore, the flood-prone area map produced here provides a broad and general guide of the current time and should not be used for any long-term decision making. Additional input parameters could be added to enhance the map such real-time precipitation data.

4 Conclusion

Newspapers contain a wealth of space and temporal information that support many research topics such as flood studies. In this research, newspapers were used to trace flash floods in the Eastern region of the UAE over a span of 18 years (2000–2018). Valuable information such as frequency of and damage caused by floods could be extracted from them. A flood-prone area map was created for the study area, and newspapers reports were used as a validation tool for GIS model output. A large portion of populated areas in the region were found to fall in medium- and high-prone zones. However, data in newspapers are “unstructured,” similar to information in mass media (Twitter, Facebook, WhatsApp, etc.), and extraction of/searching for data is not like searching using a standard database management system (DBMS) wherein data are organized in tables (structured data). This study thus needed smart text-mining software, which was developed to extract information from newspapers reports. The future plan is to develop an application that accelerates the search process and interacts with GIS software dynamically.

It is recommended that all newspaper archives in the UAE have not only a common database managed by the Center for Documentation but also a search engine similar to that developed by LexisNexis (2018). Moreover, it would be helpful if newspaper reports were complemented with accurate geographic coordinates of the events to improve location-aware services. Some of the newspapers are in Arabic and searching Arabic text is another area that requires the development of a search engine and linkage with a GIS database. Considering the necessity to improve the means and methods to assess and monitor floods, this paper presents the capabilities offered by GIS techniques to produce flood-prone area maps. In addition, it is evident that flood-prone area mapping at the pre-feasibility level could be carried out using secondary information from maps, satellite images, and published documents. The number of factors to be included in the preparation of a flood-prone area map could be increased based on data availability. This type of flood-prone area map in digital form may be used as a database that could be shared among various government and non-government agencies concerned about floods. There is a need for proper assessment and archiving of flood damage together with their geographic locations. Data about such damage could be used for compensation and to check the accuracy of flood-prone area maps.

Acknowledgements The authors acknowledged the financial support from UAE University Research Affairs Office under the Summer Undergraduate Research Experiences (SURE) Program 2019–2020 (Fund No.: G00003135). Thanks to the editorial board of *Natural Hazard* journal and the anonymous reviewers for their constructive suggestions and insights. Gratitude is extended to USGS, Fujairah Municipality, UAE Storm Center, Department of Civil Aviation, World Weather, FAO, Mr. Shehab Majud, Dr. Goepel KD, Dr. Tareefa Alsumaiti, Dr. Khalid Husein, Dr. Sharif H and all who provided support for this project.

Authors' contributions Conceptualization was contributed by Yagoub, Aishah, and Elfdel; data curation was contributed by Punitha Periyasamy, Reem Alameri, Salama Aldarmaki, Yaqein Alhashmi; formal analysis was contributed by Yagoub, Aishah, and Elfdel; funding acquisition was contributed by Yagoub; investigation was contributed by Yagoub, Aishah, and Elfdel; methodology was contributed by Yagoub; project administration was contributed by Yagoub; supervision was contributed by Yagoub; visualization was contributed by Yagoub; writing—original draft was contributed by Yagoub; Writing—review and editing was contributed by Elfdel.

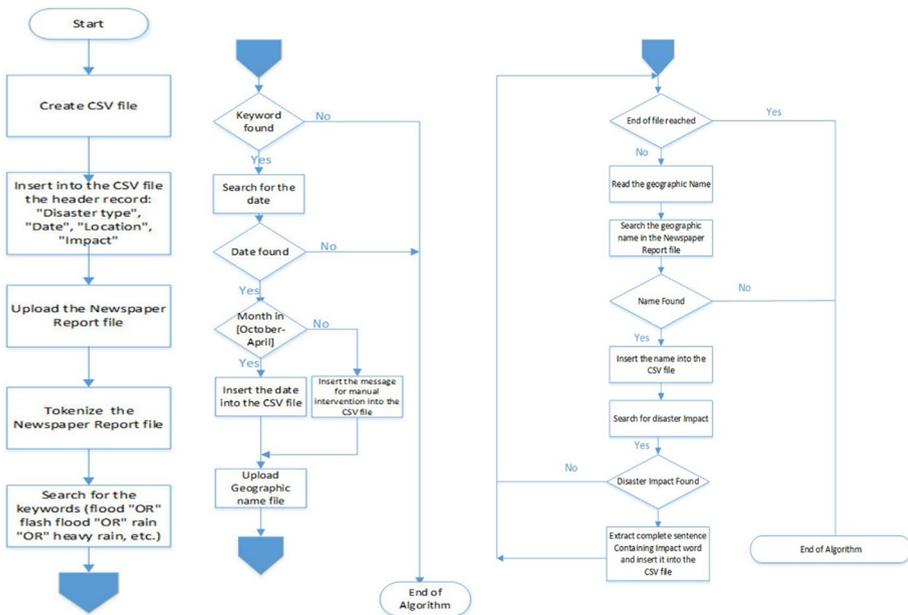
Funding This research was funded by the United Arab Emirate University-Research Affairs—SURE Plus 2019/2020 (Project No. G00003135). The views and conclusions are those of the authors and should not be taken as those of the sponsor.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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/>.

Appendix: Flowchart and sample of Java code for extracting flash flood data from newspapers



Sample of Java code for extracting flash flood data from newspapers

Sample of Java code for extracting flash flood data from newspapers

```
import java.io.*;
import java.util.*;
public class DisasterAlgorithm
{
    static int x;
    static String disaster;
    static String date;
    public static void main(String[] args) throws
    FileNotFoundException
    {
        String locationsFile = "Locations.txt";
        Scanner inFileLocations = new Scanner(new FileReader(locationsFile));
        String newsPaperReportFile;
        String locationName;
        String location;
        String month;
        String impact;
        boolean found = false;
        boolean foundLocation = false;
        boolean rainySeason;
        newsPaperReportFile = "Dec_18_2017.txt";

        try{

            FileWriter csvWriter = new FileWriter("disaster.csv");
            csvWriter.append("Disaster type");
            csvWriter.append(",");
            csvWriter.append("Date");
            csvWriter.append(",");
            csvWriter.append("Location");
            csvWriter.append(",");
            csvWriter.append("Impact");

            found = checkDisaster(newsPaperReportFile);
            if (found)
            {
                month = findMonth(newsPaperReportFile);
                rainySeason = checkRainySeasonMonth(month);
                if (!rainySeason)
                    date = "for manual intervention";
                while (inFileLocations.hasNextLine())
                {
                    location = inFileLocations.nextLine();
                    foundLocation = checkLocation(newsPaperReportFile, location);
                    if (foundLocation)
                    {
                        impact = findImpact(newsPaperReportFile, location);
                        csvWriter.append("\n");
                        csvWriter.append(disaster);
                        csvWriter.append(",");
                        csvWriter.append(date);
                        csvWriter.append(",");
                        csvWriter.append(location);
                        csvWriter.append(",");
```

```

        csvWriter.append(impact);
    } //end of if (foundLocation)

    // System.out.println(dataItems[0]);
} //end of while (inFileLocations.hasNextLine())

    csvWriter.flush();
    csvWriter.close();
} // end of if (found)
else // no disaster found
    System.out.println("No disaster data found");

} catch (Exception e) { System.out.println(e); }

} //end of method main

static boolean checkDisaster(String newsPaperFile) throws
FileNotFoundException

```

References

- Abascal-Mena R, López-Ornelas E, Zepeda JS (2013) Geographic information retrieval and visualization of online unstructured documents. *International Journal of Computer Information Systems and Industrial Management Applications* 5:089–097. https://www.mirlabs.org/ijcisim/regular_papers_2013/Paper84.pdf. Accessed 11 August 2018
- Alabdouli K, Hussein K, Ghebreyesus D, Sharif H (2019) Coastal runoff in the United Arab Emirates-The hazard and opportunity. *Sustainability* 11 (5406):1–19. <https://www.mdpi.com/2071-1050/11/19/5406>. Accessed 19 August 2018
- Alaghamand S, Abdullah RB (2010) GIS-based river flood hazard mapping in urban area: A case study in KayuAra river basin, Malaysia. *International Journal of Engineering and Technology* 2(6):488–500. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.442.5344&rep=rep1&type=pdf>. Accessed 19 August 2018
- Al-Husban YA (2019) Landforms classification of Wadi Al-Mujib Basin in Jordan, based on topographic position index (TPI) and the production of a flood forecasting map. *Dirasat Hum Soc Sci* 46:44–56
- Alittihad (2012) <https://www.alittihad.ae>. Accessed 20 August 2019
- Alkhaleej (2013) <https://www.alkhaleej.ae>. Accessed 13 August 2019
- Al Murshidi AH (2012) Flash flooding risk assessment using GIS in Dibba Fujairah, UAE. M.Sc. in Remote Sensing and Geographic Information Systems, UAE University
- Al Ghasyah Dhanhani H (2010) Evaluation of the response capability of the United Arab Emirates (UAE) to the impact of natural hazards. PhD Thesis, University of Bedfordshire, UK
- Alexander M, Faulkner H, Viavattene C, Priest S (2011) A GIS-based flood risk assessment tool; supporting flood incident management at the local scale. Proof of concept report to the Flood Risk Management Research Consortium (FRMRC2). July 2011. www.floodrisk.org.uk. Accessed 14 June 2020
- Al-Shamsei MH (1993) Drainage basins and flash flood hazards in Al Ain area, United Arab Emirates. M.Sc. Thesis, UAE University, https://scholarworks.uaeu.ac.ae/all_theses/71. Accessed 20 August 2019

- Aronica GT, Bates PD, Horritt MS (2002) Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrol Process* 16:2001–2016
- Asare-Kyei D, Forkuor G, Venus V (2015) Modeling flood hazard zones at the sub-district level with the rational model integrated with GIS and remote sensing approaches. *Water* 7:3531–3564
- Azizat N, Omar W (2018) Assessment of three flood hazard mapping methods: a case study of Perlis. *E3S Web of Conferences* 34, 02028. https://www.e3s-conferences.org/articles/e3sconf/abs/2018/09/e3sconf_cenviron2018_02028/e3sconf_cenviron2018_02028.html. Accessed 19 August 2018
- Ballerine C (2017) Topographic Wetness Index, Urban Flooding Awareness Act, Action Support: Will and DuPage Counties, Illinois. Contract report 2017–02, April 2017, Illinois state. <https://www.isws.illinois.edu/pubdoc/CR/ISWSCR2017-02.pdf>. Accessed 27 May 2020
- Bates PD (2004) Remote sensing and flood inundation modeling. *Hydrol Process* 18(13):2593–2597
- Belton V, Stewart T (2002) Multiple criteria decision analysis: an integrated approach. Kluwer Academic, Dordrecht
- Benson MA (1950) Use of historical data in flood-frequency analysis. *EOS Am Geophys Union Trans* 31:419–424
- Beven KJ, Kirkby MJ, Seibert J (1979) A physically based, variable contributing area model of basin hydrology. *Hydrol Sci Bull* 24:43–69
- Borga M, Anagnostou E, Blöschl G, Creutin J (2011) Flash flood forecasting, warning and risk management: The HYDRATE project. *Environ Sci Policy* 14(7):834–844
- Brisaboa NR, Luaces MR, Places AS, Seco D (2010) Exploiting geographic references of documents in a geographical information retrieval system using an ontology-based index. *Geoinformatica* 14(3):307–331
- Cabrera JS, Lee HS (2020) Flood risk assessment for Davao Oriental in the Philippines using geographic information system-based multi-criteria analysis and the maximum entropy model. *J Flood Risk Manag* 13:1–17
- Cao C, Peihua X, Yihong W, Jianping C, Lianjing Z, Cencen N (2016) Flash flood hazard susceptibility mapping using frequency ratio and statistical index methods in coalmine subsidence areas. *Sustainability* 8(948):1–18
- Carver SJ (1991) Integrating multicriteria evaluation with geographical information systems. *Int J Geogr Inf Syst* 5(3):321–339. <https://doi.org/10.1080/02693799108927858>
- Chang NB, Guo DH (2006). Urban flash flood monitoring, mapping and forecasting via tailored sensor network system. In: Proceedings of the IEEE international conference on networking, sensing and control, (ICNSC'06), Ft. Lauderdale, FL, USA, 23–25 April 2006
- Danumah JH, Odai SN, Saley BM, Szarzynski J, Thiel M, Kwaku A, Kouame FK, Akpa LY (2016) Flood risk assessment and mapping in Abidjan district using multi-criteria analysis (AHP) model and geo-information techniques (cote d'ivoire). *Geoenviron Disas* 3(10):1. <https://doi.org/10.1186/s40677-016-0044-y>
- Dash P, Sar J (2020) Identification and validation of potential flood hazard area using GIS-based multi-criteria analysis and satellite data-derived water index. *Journal of Flood Risk Management* 2020:1–14. <https://onlinelibrary.wiley.com/doi/full/10.1111/jfr3.12620>. Accessed 20 May 2020
- Dilts TE (2015) Topography Tools for ArcGIS 10.1. University of Nevada Reno. <https://www.arcgis.com/home/item.html?id=b13b3b40fa3c43d4a23a1a09c5fe96b9>. Accessed 13 May 2020
- Duan W, He B, Nover D, Fan J, Yang G, Chen W, Meng H, Liu C (2016) Floods and associated socioeconomic damages in China over the last century. *Nat Hazards* 82(1):401–413
- Escobar MP, Demeritt D (2014) Flooding and the framing of risk in British broadsheets, 1985–2010. *Publ Understand Sci* 23(4):454–471
- Escobar MP, Vinogradova M, Demeritt D (2016) Using newspapers as a source of data to assess flood impacts: Methodology note. <https://www.naturalhazardspartnership.org.uk/wp-content/uploads/2016/10/Methodology-note-on-impact-scoring-final.pdf>. Accessed 12 August 2019
- Estoque RC (2012) Analytic hierarchy process in geospatial analysis. In: *Progress in geospatial analysis*; Murayama, Y., Ed.; Springer: Berlin, Germany. 157–181
- FEMA (2019) Risk Mapping, Assessment and Planning (Risk MAP). <https://www.fema.gov/risk-mapping-assessment-and-planning-risk-map>. Accessed 30 July 2019
- FEMA (2020) Hazus. <https://www.fema.gov/hazus>. Accessed 13 June 2020
- Feng Q, Liu, J, Gong J (2015) Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China. *Water* (7):1437–1455. <https://www.mdpi.com/2073-4441/7/4/1437>. Accessed 10 May 2020
- Finn H (2008) Dam failure and inundation modeling: Test case for Ham Dam. Summary Report, Project conducted by “DHI Gulf” for UAE Ministry of Environment & Water

- Flood List (2018) UN—1995 to 2015, flood disasters affected 2.3 billion and killed 157,000. <https://floodlist.com/dealing-with-floods/flood-disaster-figures-1995-2015>. Accessed 3 October 2019
- Foody GM (2002) Status of land cover classification accuracy assessment. *Remote Sens Environ* 80(2002):185–201
- Forkuo EK (2011) Flood hazard mapping using Aster image data with GIS. *International Journal of Geomatics and Geosciences* 1: 932–950. <https://www.ipublishing.co.in/jggsvol1no12010/EIJGG S2051.pdf>. Accessed 20 August 2019
- Fujairah Statistics Centre (2018) Statistical year book. <https://www.fscfuj.gov.ae/books/book2018.pdf>. Accessed 15 September 2019
- Getahun YS, Gebre SL (2015) Flood hazard assessment and mapping of flood inundation area of the Awash River Basin in Ethiopia using GIS and HEC-GeoRAS/HEC-RAS Model. *J Civil Environ Eng.* 5:179. <https://www.omicsonline.org/open-access/flood-hazard-assessment-and-mapping-of-flood-inundation-area-of-the-awash-river-basin-in-ethiopia-using-gis-and-hecgeorashecra-model-2165-784X-1000179.php?aid=57721>. Accessed 19 November 2019.
- Ghazanfar SA, Fisher M (eds) (1998) *Vegetation of the Arabian Peninsula*. Springer, Berlin. <https://doi.org/10.1007/978-94-017-3637-4>
- GloFAS (2020) Global Flood Awareness System (GloFAS). www.globalfloods.eu. Accessed 15 June 2020
- Goepel KD (2018) Implementation of an online software tool for the analytic hierarchy process (AHP-OS). *Int J Anal Hierarchy Process* 10(3):469–487
- Harrigan S, Zsoter E, Alfieri L, Prudhomme C, Salamon P, Wetterhall F, Barnard C, Cloke H, Pappenberger F (2020) GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst Sci Data Discuss.* <https://doi.org/10.5194/essd-2019-232, inreview>
- Helley EJ, Lamarche VC (1973) Historic flood information for Northern California streams from geophysical and botanical evidence. U.S. Geological Survey Professional Paper. No. 485 - E.
- Ho W (2008) Integrated analytic hierarchy process and its applications: a literature review. *Eur J Oper Res* 18:211–228
- HWSD (2009) Harmonized World Soil Database. FAO, IIASA, ISRIC, ISSCAS, JRC. <https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/ar/>. Accessed 13 May 2020
- Ishizaka A, Labib A (2011) Review of the main developments in the analytic hierarchy process. *Expert Syst Appl* 38:14336–14345
- Isma'il M, Saanyol IO (2013) Application of Remote Sensing (RS) and Geographic Information Systems (GIS) in flood vulnerability mapping: Case study of River Kaduna. *International Journal of Geomatics and Geosciences* 3:618–627. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.21.8578&rep=rep1&type=pdf>. Accessed 19 November 2019
- Jenness J (2004) Calculating landscape surface area from digital elevation models. *Wildl Soc Bull* 32(3):829–839
- Jenness J (2011) Topographic Position Index extension for ArcGIS. Jenness Enterprises. https://www.jennessent.com/arcgis/arcgis_extensions.htm. Accessed 11 May 2020
- Kazakis N, Ioannis K, Thomas P (2015) Assessment of flood hazard areas at a regional scale using an index-based approach and analytical hierarchy process: application in Rhodope-Evros region, Greece. *Sci Total Environ* 538:555–563
- Kutija VA, Bertsch R, Glenis V, Alderson D, Parkin G, Walsh C, Robinson J, Kilsby C (2014) Model validation using crowd-sourced data from a large pluvial flood. In the 11th International Conference on Hydroinformatics, City University of New York (CUNY), 8 January. CUNY Academic Works https://academicworks.cuny.edu/cc_conf_hic/415. Accessed 13 July 2019
- Legates DR (2000) Real-time calibration of radar precipitation estimates. *Prof Geogr* 52(2):235–246
- LexisNexis (2018) Searchable database about newspapers. <https://internationalsales.lexisnexis.com>. Accessed 12 August 2019
- Luino F, Nigrelli G, Turconi L, Faccini F, Agnese C, Casillo F (2016) A proper land-use planning through the use of historical research. *Disas Adv* 9(1):8–18
- Luino F, Belloni A, Turconi L, Faccini F, Mantovani A, Fassi P, Marincioni F, Caldiroli G (2018) A historical geomorphological approach to flood hazard management along the shores of an alpine lake (Northern Italy). *Nat Hazards* 94:471–488. <https://doi.org/10.1007/s11069-018-3398-5>
- Malczewski J (1999) *GIS and multi-criteria decision analysis*. Wiley, New York, NY, USA
- Malczewski J (2006) GIS-based multicriteria decision analysis: a survey of the literature. *Int J Geogr Inf Sci* 20(7):703–726

- Marin-Ferrer M, Vernaccini L, Poljansek K (2017) Index for risk management INFORM concept and methodology report - version 2017, EUR 28655 EN, 10.2760/094023. <https://drmkc.jrc.ec.europa.eu/inform-index> Accessed 5 October 2019
- Marinoni O (2004) Implementation of the analytical hierarchy process with VBA in ArcGIS. *Comp Geosci* 30(6):637–646
- Molinari D, De Bruijn K, Castillo J, Aronica GT, Bouwer LM (2017) Validation of flood risk models: current practice and innovations. <https://www.nat-hazards-earth-syst-sci-discuss.net/nhess-2017-303/nhess-2017-303.pdf>. Accessed 13 November 2019
- Ministry of Energy, United Arab Emirates (2006) The United Arab Emirates: Initial national communication to the United Nations Framework Convention on Climate Change. Ministry of Energy, United Arab Emirates. <https://unfccc.int/resource/docs/natc/arenc1.pdf>. Accessed 5 October 2019
- Newby JM, Hardy SB (2018) Using newspapers as primary source. https://centuryofaction.org/index.php/main_site/document_project/using_newspapers_as_primary_sources. Accessed 11 August 2019
- Ouma YO, Tateishi R (2014) Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: Methodological overview and case study assessment. *Water* 6:1515–1545
- Papaioannou G, Vasilades L, Loukas A (2015) Multi-criteria analysis framework for potential flood prone areas mapping. *Water Resour Manag* 29(2):399–418
- Periyasamy P, Yagoub MM, Sudalaimuthu M (2018) Flood vulnerable zones in the rural blocks of Thiruvallur district. *South India Geoenviron Disas* 5:21. <https://doi.org/10.1186/s40677-018-0113-5>
- Porter JJ, Evans G (2020) Unreported world: a critical analysis of UK newspaper coverage of post-disaster events. *Geogr J*. <https://doi.org/10.1111/GEOJ.12353>
- Prinos P (2008) Review of flood hazard mapping. European Community Sixth Framework Programme for European Research and Technological Development. FLOODsite. www.floodsite.net. Accessed 14 June 2020
- Quirós E, Gagnon AS (2020) Validation of flood risk maps using open source optical and radar satellite imagery. *Trans GIS* 2020(00):1–19. <https://doi.org/10.1111/tgis.12637>. Accessed 5 May
- Rizk ZS, Al Sharhan AS (2003) Water resources perspectives: evaluation, management and policy. In: Alsharhan AS and Wood WW (eds.) Elsevier Science, Amsterdam, p 245–264
- Saaty TL (1980) The analytic hierarchy process. McGraw Hill, New York
- Saaty TL (1990) How to make a decision: the analytic hierarchy process. *Eur J Oper Res* 48(1):9–26
- Saaty TL, Shang JS (2007) Group decision-making: head-count versus intensity of preference. *Socio-Econ Plan Sci* 41:22–37
- Samanta S, Koloa C, Pal DK, Palsamanta B (2016) Flood risk analysis in lower part of Markham River based on multi-criteria decision approach (MCDA). *Hydrology* 3(3):29. <https://doi.org/10.3390/hydrology3030029>
- Schumann GJ-P, Bates PD, Apel H, Aronica GT (Eds.) (2018) Global Flood Hazard: Applications in Modeling, Mapping, and Forecasting. American Geophysical Union- Standards Information Network, Washington, DC <https://www.wiley.com/>
- Sherif MM, Mohamed MM, Shetty A, Almulla M (2010) Rainfall-runoff modeling of three wadis in the northern area of UAE. *J Hydrol Eng* 16(1):10–20
- Sherif M, Sefelnasr A, Ebraheem AM, Javadi A (2013) Quantitative and qualitative assessment of seawater intrusion in Wadi Ham under different pumping scenarios. *J Hydrol Eng* 19(5):855–866. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000907](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000907)
- Stefanidis S, Stathis D (2013) Assessment of flood hazard based on natural and anthropogenic factors using analytic hierarchy process (AHP). *Nat Hazards* 68:569–585
- Tarhule A (2005) Damaging rainfall and flooding: the other Sahel hazards. *Clim Change* 72(3):355–377
- Taylor FE, Malamud BD, Freeborough K, Demeritt D (2015) Enriching Great Britain's National Landslide Database by searching newspaper archives. *Geomorphology*. <https://www.sciencedirect.com/science/article/pii/S0169555X15300039?via%3Dihub>. Accessed 9 July 2019
- Teknomo K (2020) Analytic Hierarchy Process (AHP) tutorial. <https://people.revoledu.com/kardi/tutorial/>. Accessed 13 May 2020
- Thomlinson JR, Bolstad PV, Cohen WB (1999) Coordinating methodologies for scaling land cover classifications from site-specific to global: steps toward validating global map products. *Remote Sens Environ* 70:16–28
- Tropeano D, Turconi L (2004) Using historical documents for landslide, debris flow and stream flood prevention. *Appl North Italy Nat Hazards* 31:663–679
- Trimble SW (2008) The use of historical data and artifacts in geomorphology. *Prog Phys Geogr* 32(1):3–29
- Atlas UAE (1993) UAE atlas. UAE University Publication, Al Ain, UAE
- UK Environment Agency (2020) Check your long term flood risk. <https://flood-warning-information.service.gov.uk/long-term-flood-risk/map>. Accessed 4 May 2020

- UN (2010) Land and natural disasters guidance for practitioners. United Nations Human Settlements Programme (UN-HABITAT), 2010, Nairobi, Kenya. https://postconflict.unep.ch/humanitarianaction/documents/02_03-04_03-02.pdf. Accessed 13 November 2019
- United Nations Convention on the Law of the Sea (UNCLOS) (2019). Territorial sea and contiguous zone. https://www.un.org/depts/los/convention_agreements/texts/unclos/part2.htm. Accessed 19 June 2019
- USDA-NRCS (1986) Urban hydrology for small watersheds. https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1044171.pdf. Accessed 4 May 2020
- Vaezi AR, Zarrinabadi E, Auerswald K (2017) Interaction of land use, slope gradient and rain sequence on runoff and soil loss from weakly aggregated semi-arid soils. *Soil Tillage Res* 172:22–31
- Van Alphen J, Passchier R (2007) Atlas of flood maps. Examples from 19 European countries, USA and Japan. EXCIMAP, Ministry of Transport, Public Works and Water Management, The Netherlands
- van der Knijff JM, Younis J, de Roo APJ (2010) LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *Int J Geogr Info Sci* 24:189–212
- Wehbe Y, Ghebreyesus D, Temimia M, Milewskib A, Al Mandousc A (2017) Assessment of the consistency among global precipitation products over the United Arab Emirates. *J Hydrol Reg Stud* 12:122–135
- Weiss A (2001) Topographic position and landforms analysis. Poster presentation, ESRI User Conference, San Diego, CA, USA
- Wolock DM, McCabe GJ (1995) Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resour Res* 31(5):1315–1324
- World Weather (2019) Average rainfall amount, Fujairah 2019. <https://www.worldweatheronline.com/lang/ar/fujairah-weather-averages/fujairah/ae.aspx>. Accessed 26 July 2019
- Wright R (2014) What goes into a flood map: Infographic. <https://www.fema.gov/blog/2014-02-21/what-goes-flood-map-infographic>. Accessed 13 June 2020
- Wu Y, Zhong PA, Zhang Y, Xu B, Ma B, Yan K (2015) Integrated flood risk assessment and zonation method: a case study in Huaihe River basin. *China Nat Hazards* 78(1):635–651
- Yagoub MM, Jalil AM (2014) Urban fire risk assessment using GIS: Case study on Sharjah, UAE. *Int Geoinformatics Res Dev J (IGRDJ)* 5(3):1–8. https://www.widm.ca/wp-content/uploads/2017/04/Yagoub-et-al-Full-text_IGRDG202011.pdf. Accessed 19 November 2019
- Yagoub MM, Kolan GR (2006) Monitoring coastal zone land use and land cover changes of Abu Dhabi using remote sensing. *J Indian Soc Remote Sens* 34(1):57–68
- Zhang X (2010) Discourse analysis of China daily's disaster news reports: an appraisal approach. Master's thesis, Chengdu University of Electronic Science and Technology, Chengdu, China. Retrieved from <https://www.doc88.com/p-03344099865.html>. Accessed 5 October 2019
- Zhan X, Huang M (2004) ArcCN-Runoff: an ArcGIS tool for generating curve number and runoff maps. *Env Model Softw* 19:875–879

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

Affiliations

M. M. Yagoub¹  · Aishah A. Alseraidi¹ · Elfadil A. Mohamed² · Punitha Periyasamy³ · Reem Alameri¹ · Salama Aldarmaki¹ · Yaqein Alhashmi¹

Elfadil A. Mohamed
elfadil.abdalla@ajman.ac.ae

Punitha Periyasamy
punithaperiyasamy@gmail.com

Reem Alameri
201406301@uaeu.ac.ae

Salama Aldarmaki
201604639@uaeu.ac.ae

Yaqein Alhashmi
201410328@uaeu.ac.ae

- ¹ Department of Geography and Urban Sustainability, College of Humanities and Social Sciences, UAE University, Al Ain, UAE
- ² Information Technology Department, College of Engineering and Information Technology, Ajman University, Ajman, UAE
- ³ Department of Civil Engineering, SRM University, KattankulathurKancheepuram, Tamil Nadu 603203, India