A comparative assessment of fuzzy logic and evidential belief function models for mapping artesian zone boundary in an arid region, Iraq

Alaa M. Al-Abadi, Suaad A. Al-Bhadili and Maitham A. Al-Ghanimy

ABSTRACT

This paper discusses and compares the potential application of the evidential belief function model and fuzzy logic inference system technique for spatial delineation of a groundwater artesian zone boundary in an arid region of central Iraq. First, a flowing well inventory of a total of 93 perennial flowing wells was constructed and randomly partitioned into two data sets: 70% (65 wells) for training and 30% (28 wells) for validation. Twelve groundwater conditioning factors were considered in the geospatial analysis depending on data availability and literature review. The random forest (RF) algorithm was first applied to investigate the most important conditioning factors in groundwater potential analysis. The most important factors with training flowing wells were used to develop predictive models. The prediction accuracy of the developed models was checked using the area under the relative operating characteristic curve. Results showed that the best model with a higher prediction accuracy of 86% was a fuzzy AND model followed by the evidential model with 84%. The main conclusion of this study is that the integrated use of the adapted models with RF offer a rapid assessment tool in groundwater exploration and can be helpful in groundwater management.

Key words | evidential belief functions, fuzzy logic, Iraq, Karbala

Alaa M. Al-Abadi (corresponding author)
Suaad A. Al-Bhadili
Department of Geology, College of Science,
University of Basra,
Basra,
Iraq
E-mail: alaaatiaa@gmail.com

Maitham A. Al-Ghanimy
General Commission of Groundwater, Ministry of
    Water Resources,
Karbala,
Iraq

INTRODUCTION

Groundwater is an important water resource around the world. The broad geographical distribution, huge reserves, generally good quality, and the ability to cope with seasonal fluctuations and contamination, helps water bearing layers maintain the pledge to ensure current and future safe supply. Poor management of hydrogeological systems (aquifers) along with the impact of inadequate land-use practices has caused adverse effects such as groundwater depletion, water-quality deterioration, and the decimation of aquatic ecosystems. Sometimes, mining of aquifers produces additional adverse problems such as land subsidence and a drying of the wetland. It is anticipated that pressures on groundwater resources will be increased mainly as a result of population growth and growing competition for water. Thus, it is crucial to develop proper management plans to ensure the long life of aquifers in both quantity and quality aspects. In this context, spatial delineation of aquifer potentiality has become a necessary and easy to implement option for the achievement of successful protection and management groundwater plans (Ozdemir 2011). It is also useful to plan and engineer the implementation of successful resource exploration.

In the past few years, different modeling techniques integrated with geographical information system (GIS) have been used as a spatial tool for demarcating groundwater potentiality. GIS is an important system for integrating

and analyzing information from different sources and disciplines. The GIS-based models can be effectively connected in multisource data analysis with heterogeneous and uncertain data (Chacón et al. 2006). Although many GIS-based models have been applied previously for modeling groundwater potentiality, it is still early to distinguish which one is the best. Therefore, the comparative studies of using different methods are highly necessary (Bui et al. 2012; Rahmati et al. 2016). The evidential belief function (EBF) technique is a mathematical framework for describing quantified belief held by an agent (Reineking 2014). It is based on the theory of Bayesian probability and has a relative flexibility to accept uncertainty and the ability to combine beliefs from multiple sources of evidence (Thiam 2005). The application of the EBF data-driven model in groundwater potential studies is still limited and only a few studies exist (Table 1). On the other hand, fuzzy logic (FL) is an approach for computing based on 'degrees of truth' rather than the usual 'true or false' (1 or 0) Boolean logic on which the modern computer is based. FL has been widely used in many fields of science and engineering. The power of FL is that it is easy to implement, and the process of assigning weights for groundwater conditioning factors used in the analysis is totally determined by the experts. The use of this technique for spatially studying groundwater productivity is still limited as well (Rather & Andrabi 2012; Aouragh et al. 2016).

The main objective of this paper is to delineate the spatial boundary of a groundwater artesian zone at Karbala Governorate, central Iraq, using EBF and FL prediction models under a GIS platform. The study also involves a comprehensive comparison of these two models and identifies the best one for potential delineation of this zone. The artesian zone in the study area is a portion of a series of long springs and flowing wells that are distributed parallel to the Euphrates River from north (Al-Anbar) to south (As-Samawah) in Iraqi western and southern deserts. Despite the importance of this zone, no studies have been performed so far to spatially delineate its boundary. Spatial prediction of a groundwater artesian zone in the interested area will contribute to effective management of groundwater as groundwater could be extracted with minimal effort. In addition, obtaining a new flowing well will become an easier task.

THE STUDY AREA

The study area lies about 100 km southwest of Baghdad, the capital of Iraq, between (44°25′00″–43°45′00″) longitude (E/W) and (32°40′00″–32°20′00″) latitude (N/E) (Figure 1) and covers an area of about 4,051 km². The study area is relatively flat, and the elevation ranges from 0 to 224 m (Figure 2). It is almost completely covered by pebbly or gypsiferous pebbly soil or gypcrete in addition to eolian, sheets, and shrub dunes. The climate is dry and relatively hot in summer, and cold with little rain in winter. It is believed to be influenced by the Mediterranean Sea climate. The monthly averages of climatic variables at Karbala station for the period 1990–2014 from the Iraqi Meteorological Organization is summarized in Table 2. The wind prevalent in the area is mostly northwest–southeast accompanied by sand storms in the summer and sometimes winds from the south and southwest. The study area is covered by gypcrete deposits except an area limited by Razzazah Lake and Tar-Al Sayed where Dib-dibba, Injana, and Nfayil formations are outcropped. The rock exposed in the study area dates back to the Miocene period (Figure 3(a) and 3(b)). A brief description of the exposed formations is summarized in Table 3. Tectonically, the study area is located within an unstable shelf which dates back to the cratonic era according to the bilateral division of Iraq. In general, the study area is located between two tectonic zones, the Mesopotamian zone and Al-Salman zone within a stable shelf. The Mesopotamian zone is relatively flat terrain with a gradient of less than 10 cm per kilometer extending from Baiji in the northwest to the Arabian Gulf in the southeast (Jassim & Goff 2006). The Salman zone comprises northeast–southwest and prominent northwest and southeast trending uplifts and depressions, bounded by faults. Two groups of faults exist within the area of interest. The first group trends northeast–southwest like the Khanaquin Baquba-Karbala fault and the second group trends northwest–southeast similar to the Abu Jir fault zone which is represented by the Heet-Abu Jir fault which appears at approximately the center of the study area (Al-Amiri 1979). The most apparent geomorphological features within the study area are: Najaf-Karbala plateau, Al-Razzazah depression, rock cliff, and mesas and buttes. The important lake in the study area is Milh Lake (also known as Razzaza Lake). It is located a few kilometers west of

Table 1 | Literature review of applying EBF and FL in groundwater studies

| Study	Study purpose	Thematic layers used	Relevant findings
Nampak et al. (2014)	Investigate the applicability of an EBF model for spatial delineation of groundwater productivity at Langat basin, Malaysia using GIS technique	12 groundwater conditioning factors including elevation, slope, curvature, SPI, TWI, drainage density, lithology, lineament density, land use, normalized difference vegetation index (NDVI), soil and rainfall	The output of the developed model proved the efficiency of EBF in groundwater potential mapping with success and prediction rates of 78% and 72%, respectively
Mogaji et al. (2014)	Explore the potential of a GIS-based EBF model as a spatial prediction model to groundwater productivity potential mapping in the southern part of Perak, Malaysia	7 groundwater factors including drainage density, lineament density, lineament intersection density, lithology, average annual rainfall, slope, and soil type	The obtained results indicate the usefulness of the EBF model in spatial mapping of groundwater potential zones and the capability of this model in managing uncertainty associated with the developed EBF model. The prediction accuracy of the developed model was about 85%
Pourghasemi & Beheshtirad (2015)	The objective was to produce groundwater spring potential mapping and its performance assessment using EBF model in Koohrang Watershed, Chaharmahal-e-Bakhtiari Province, Iran	12 factors including altitude, slope aspect, slope degree, slope length (LS), TWI, plan curvature, land use, lithology, distance from rivers, drainage density, distance from faults, and fault density	The prediction accuracy of the EBF model was 82% and thus regarded as a very good model to delineate groundwater potentiality
Tahmassebipoor et al. (2005)	The capability of using weights-of-evidence (WOE) and EBF models for groundwater potential mapping was tested and compared in the Ilam Plain, Iran	11 factors including lithology, land use, distance from river, soil texture, drainage density, altitude, curvature, TWI, slope percent, lineament density, and rainfall	The results showed the capability of WOE and EBF as effective prediction models for groundwater potential mapping. The prediction accuracy of the EBF model was 83.7% and better than of the WOE model with 78.2% prediction accuracy
Park et al. (2014)	The EBF model was applied and validated for analysis of groundwater-productivity potential in Boryeong and Pohang cities, an agriculture region in Korea, using GIS	Spatial database related to topography, lineament, geology, forest, soil and groundwater were constructed	Results confirmed the higher capability of EBF for delineating groundwater potential mapping with 83.41% and 77.53% accuracy in Boryeong and Pohang areas, respectively
Rather & Andrabi (2012)	Develop a FL-based model for groundwater potential in the Jhagrabaria Watershed of Allahabad District, Uttar Pradesh, India	Geomorphology, geology, physiography, lithology, lineament, contour, drainage, and water body	Findings showed the high capability of FL model to demarcate groundwater zones in the study area
Aouragh et al. (2016)	A GIS-based FL was developed in this study to delineate groundwater potential boundaries in the Middle Atlas plateaus, Morocco	Lithology, slope, karst degrees, land cover, lineament, and drainage density	Results confirmed the high efficacy of FL model to generate groundwater potential
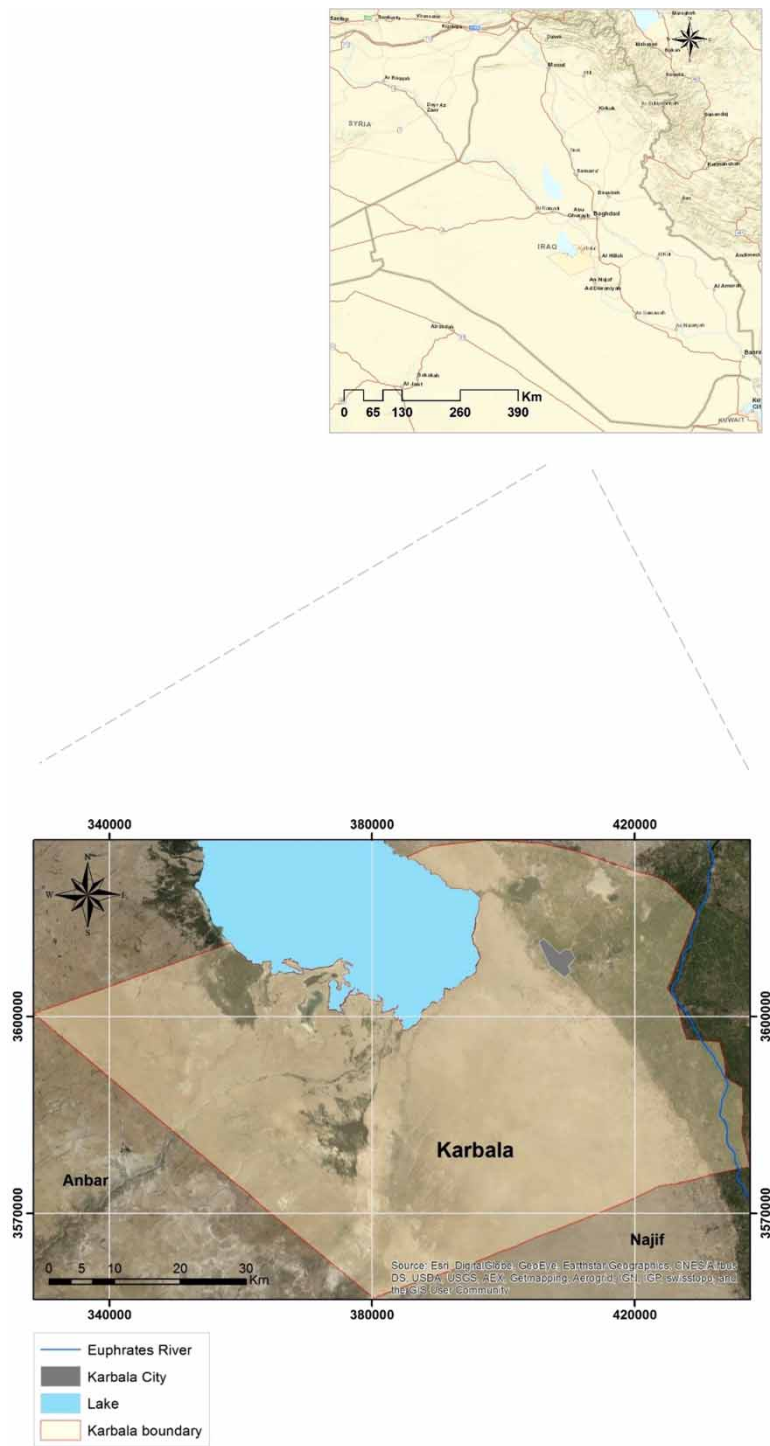
Figure 1 | Location of the study area.

Karbala. The lake is listed as a wetland of international importance. The lake is rather shallow and the water level changes with the season. Two major aquifer groups exist within the study area (Figure 4(a)). These are (i) limestone of the Palaeogene Um-ErRdhuma, Jill and Dammam formation and (ii) sands of the Quaternary Mesopotamian flood plain. The
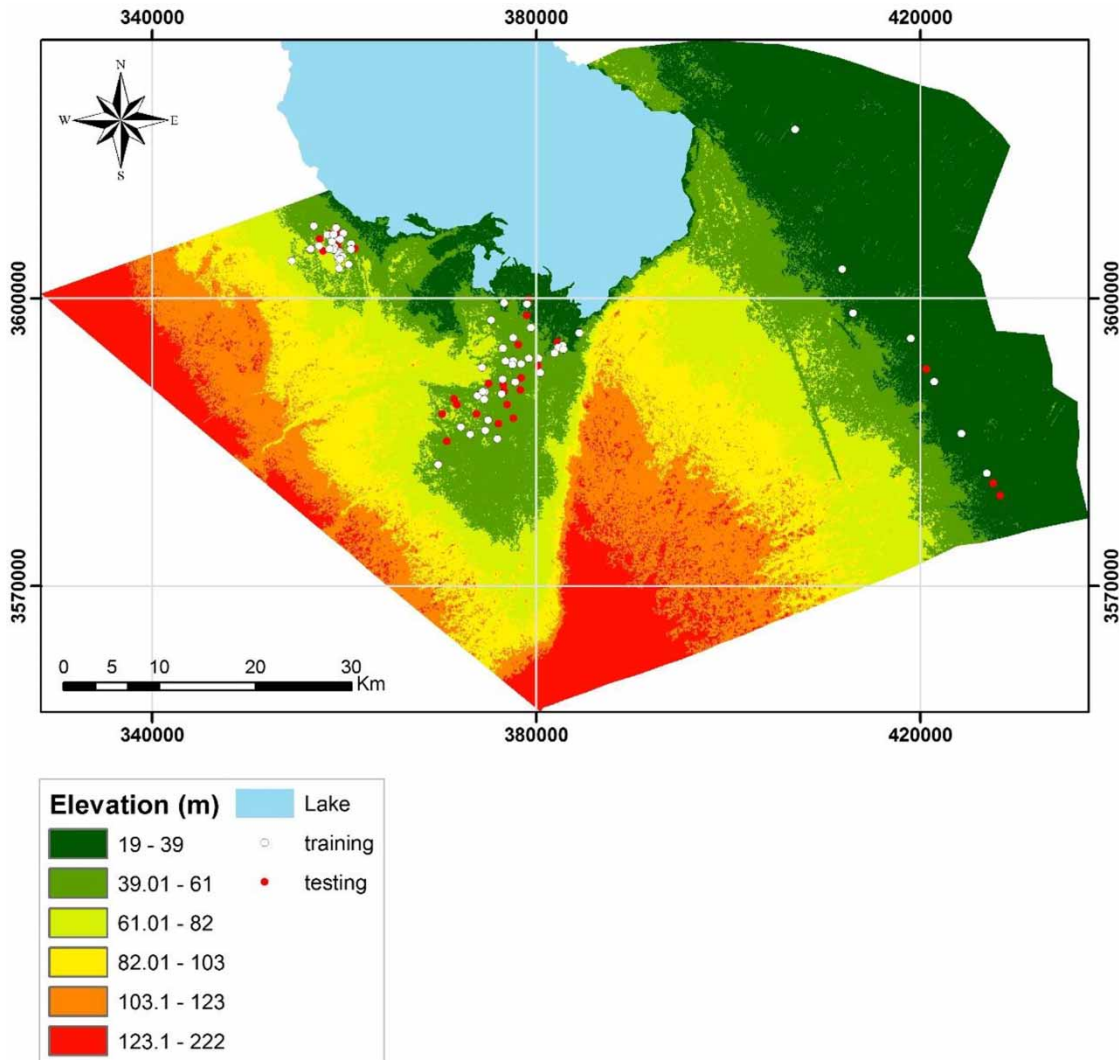
Figure 2 │ Elevation (m) in the study area with training and testing data sets.

study area was considered a discharge area for the regional aquifer according to the flow direction of the study area (Figure 4(b)). The general groundwater flow is from southwest to northeast and there are no approved studies to track the piezometric levels in the study area; only an approximate position is shown in Figure 4(b), after Sissakian (2000). There are some factors affecting the movement and flow system of groundwater in the study area: (i) the permeability and fracture density of the rock units that contain groundwater, (ii) the location of Abu Jir fault, (iii) the vertical movement of groundwater from the underlying aquifers, and (iv) the change in lateral facies and the thickness of water-bearing beds.

METHODOLOGY

A flow chart that describes the overall methodology in this paper is shown in Figure 5. Basically, a study of the potential of spatial groundwater of an area requires two main crucial steps: the preparing of a groundwater borehole inventory map and the identification of the influencing groundwater occurrence factors. The inventory of flowing wells in the study area was prepared through extensive field surveys in 2015. From these field surveys, 93 flowing wells were fixed and an inventory map was prepared. For modeling purposes, these wells were partitioned into two data sets using a random algorithm: training and testing. Out of 93,

Table 2 | Mean monthly record of climatic variables at Karbala station for the period 1990–2014

	Climatic variables				
Month	Temperature (°C)	Rainfall (mm)	Relative humidity (%)	Wind speed (m/s)	Evaporation (mm)
October	26.83	2.93	46.52	1.9	203.12
November	17.94	9.08	60.66	1.73	101.78
December	12.7	12.67	71.57	1.76	62.15
January	10.81	17.43	74.57	2.03	59.19
February	13.4	14.16	61.4	2.54	91.56
March	17.84	10.33	50.25	2.3	170.15
April	24.31	12.86	43.4	3.13	235.52
May	30.3	1.94	34.4	3.1	326.05
Jun	34.65	0	29.14	4.01	408.42
July	36.9	0	30.75	4.23	437.20
August	36.6	0	32.2	3.21	391.90
September	32.58	0.45	37.2	2.4	232.28
Average	24.56	6.82	47.67	2.70	232.28

approximately 65 (70%) wells were used for training, and the remaining 28 (30%) wells were used for testing.

The type and number of groundwater condition factors used in mapping groundwater potentiality differ in the literature and there is no standard way to select them. Local conditions and data availability are the main constraints in selecting groundwater factors from one study to another. The factors related to geology, structural setting, soil, topography related factors, and geomorphology are often considered to be the factors most influencing groundwater potentiality. In this study, and depending on the availability of data in the first place, a total of 12 factors were considered for the analysis. These factors are ground surface elevation, slope angle, aspect, profile curvature, topographic wetness index (TWI), stream power index (SPI), lithological units, fault density, distance to faults, distance to lake, aquifer major groups, and depth to groundwater. A brief description of these factors and their importance in groundwater potential is given in Table 4. Raster maps of elevation (m), slope angle in %, profile curvature, aspect, and secondary topographic indices such as TWI and SPI were generated from a digital elevation model (DEM) of type ASTER-GDEM with a spatial resolution of 1 arc-second after essential pre-processing. Elevation was classified into three categories (<30 m, 30–90 m, and >90 m; Figure 6(a)), slope angle

into four categories (<2%, 2–8%, 8–15%, and 15–30%; Figure 6(b)), curvature into three categories (<0 concave, 0 flat, >0 convex; Figure 6(c)), and finally, aspect into nine categories (flat, north, northeast, east, southeast, south, southwest, west, and northwest; Figure 6(d)). The TWI and SPI were constructed using DEM and ArcHydro tools in ArcGIS 10.2™ and classified into five categories for both factors after applying a focal statistic module (Figure 6(e) and 6(f)).

The lithological units and faults raster maps were derived from the 1.1,000,000 scale maps of the geological survey of Iraq, respectively. The two hard copies of these maps were first scanned, georeferenced, and manually digitized in an ArcGIS environment. Three lithological units were identified, namely, carbonate rocks of Euphrates formation (Middle Miocene) and Quaternary deposits (sedimentary rocks of alluvium and alluvial fans) (Figure 6(g)). The proximity to faults was constructed using the Euclidean distance module and further classified into five classes: (0–4.6 km, 4.6–9.3 km, 9.3–13.9 km, 13.9–18.6 km, and >18.6 km) (Figure 6(h)). Raster layer of fault density (km/km²) was generated using kernel density module and classified into five categories: (0–0.17, 0.17–0.34, 0.34–0.52, 0.52–0.70, and 0.67–0.87) (Figure 6(i)). In the case of distance to the lake (Rezzaza Lake), five buffer categories
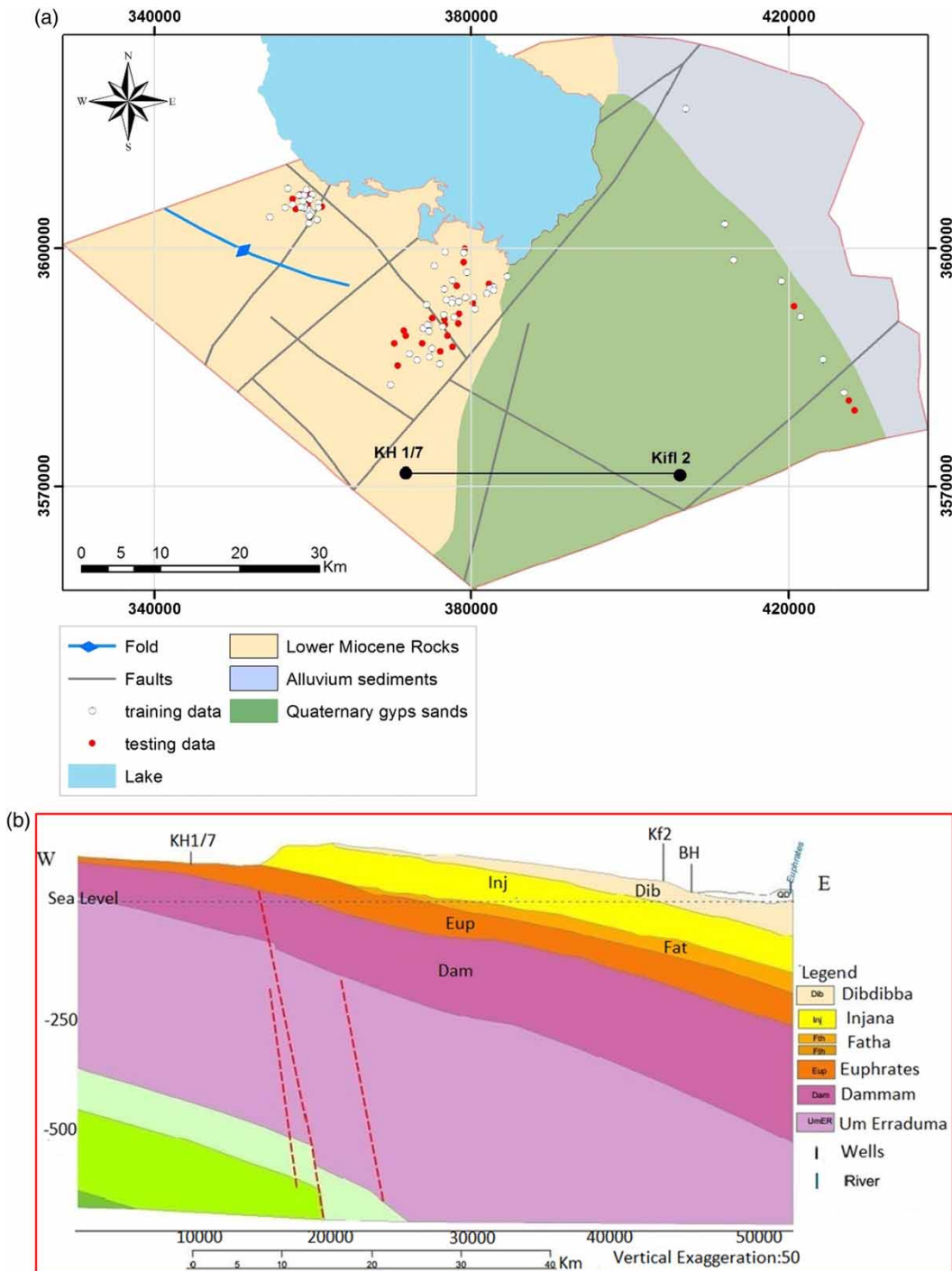
Figure 3 | (a) Geological map of the study area and (b) cross-section between borehole KH 1/7 and Kifl 2.

(0–10.3 km, 10.3–20.5 km, 20.5–30.7 km, 30.7–40.9 km, >40.9 km) were created using the Euclidian Distance module in ArcGIS 10.2™ (Figure 6(j)). For the major aquifer

groups, a hard copy of this map was acquired from the General Commission of Groundwater office, scanned, georeferenced, and digitized using the Editor tool in ArcGIS 10.2

Table 3 │ Geological description of the formations in the study area (summarized after Jassim & Goff (2006))

Formation	Age	Deposition environment	Description
Euphrates	Middle Miocene	Carbonate inner shelf	Recrystallized and siliceous limestones with texture ranging from oolitic to chalky, locally containing rocks and shale coquinas
Nfayil	Middle Miocene		Marl and limestone, claystone and limestone
Fatha	Middle Miocene	Shallow marine	Anhydrite, mudstone, and thin limestone
Injana	Upper Miocene	Sub-marine	Red or gray colored silty marl or clay stones and purple silt stones
Dibdibba	Pliocene–Pleistocene	Alluvial fans of the stable shelf	Sand and gravel containing pebbles of igneous rocks (including pink granite) and white quartz, often cemented into a hard grit
Zahra	Pliocene–Pleistocene	Fluvio-Lacustrine and karstfill facies	Consists of 30 m of limestones (subsequently found to be reed-bearing fresh water limestone), marls, and sandy marls. Locally, sandstone occurs at the base of the formation
Quaternary	Pleistocene–Holocene	Continental	Mixture of gravel, sand, silt, and clay

(Figure 4). Finally, the groundwater depth data were compiled from the work of Al-Jiburi & Al-Basrawi (2015). The groundwater depth values are classified into four categories (Figure 6(k)): (<10 m, 10–20 m, 20–30 m, and 30–40 m).

Before using the EBF and FL models to demarcate groundwater potentiality in the study area, random forest (RF) algorithm was implemented in the free R package in feature selection process. The grid cells of the flowing well were assigned a value of 1, while the non-flowing well grid cells were assigned a 0 code value. To get optimal modeling results (Carranza & Laborte 2014), a total of 130 points (65 training flowing well locations and 65 non-flowing well locations) were selected for applying the RF algorithm.
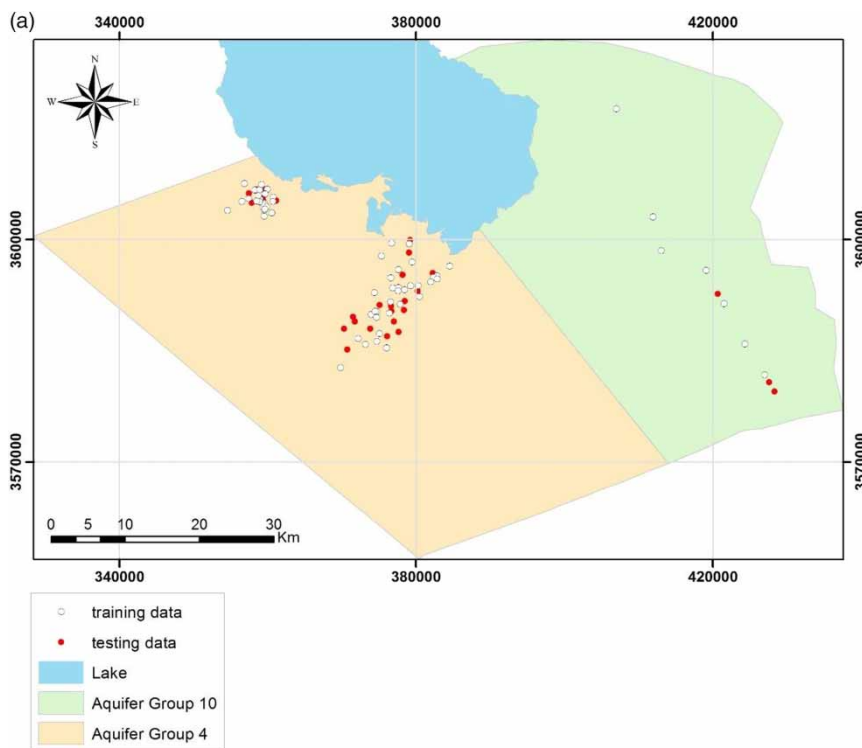

Figure 4 │ (a) Major aquifer group map of the study area. (b) General groundwater flow and approximate position of piezometric level in selected sections in the study area (Al-Jiburi & Al-Basrawi 2015). (Continued.)
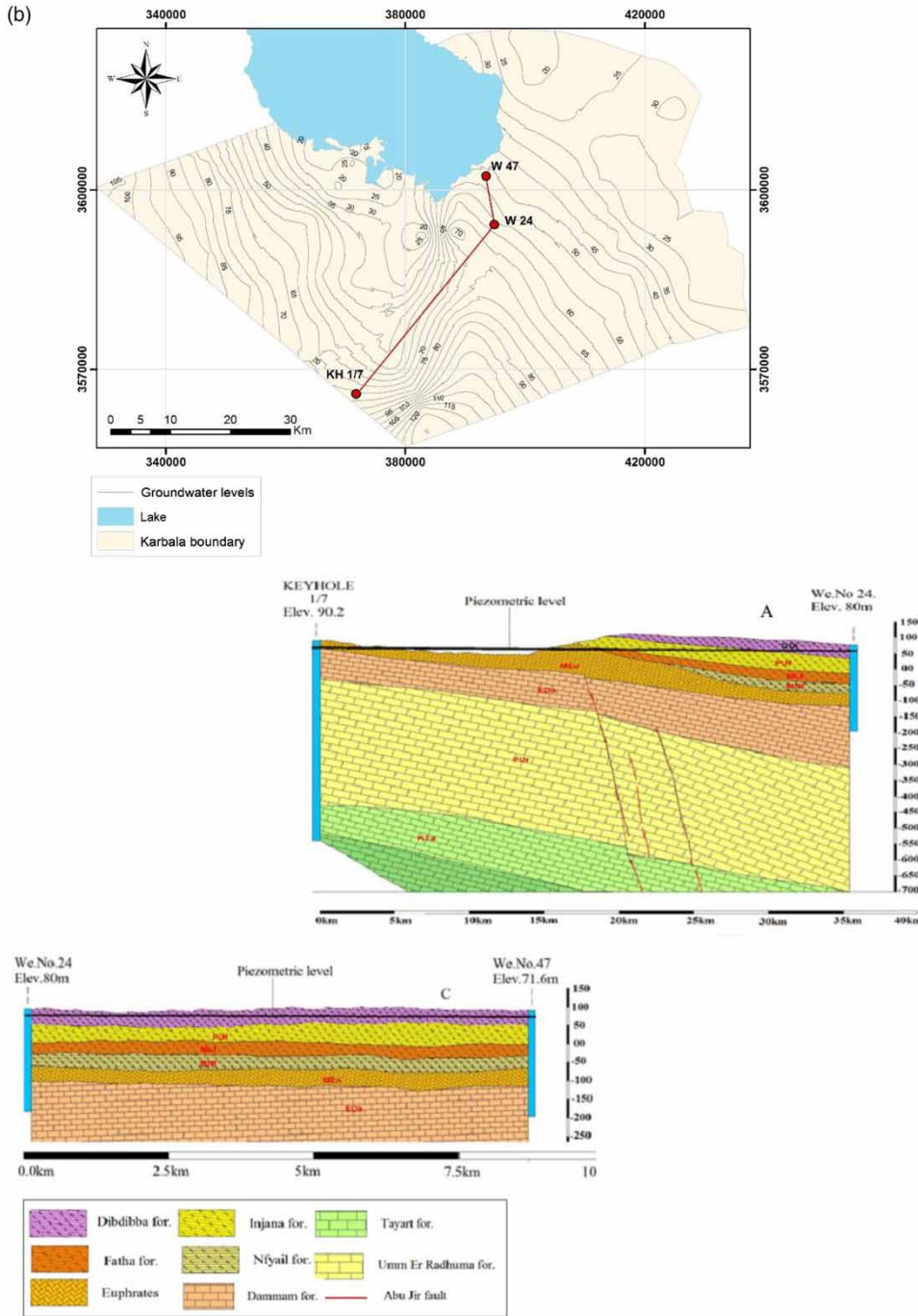
Figure 4 | Continued.

The average nearest neighbor (ANN) spatial pattern analysis was first applied to find the distance from any flowing well points and estimated corresponding probability that there was no flowing well location next to it (Carranza & Laborte

2014). The results proved that the expected mean distance was about 6 km. This distance represents the distance separating the non-random and random field in the area being studied. Therefore, all training points with non-artesian
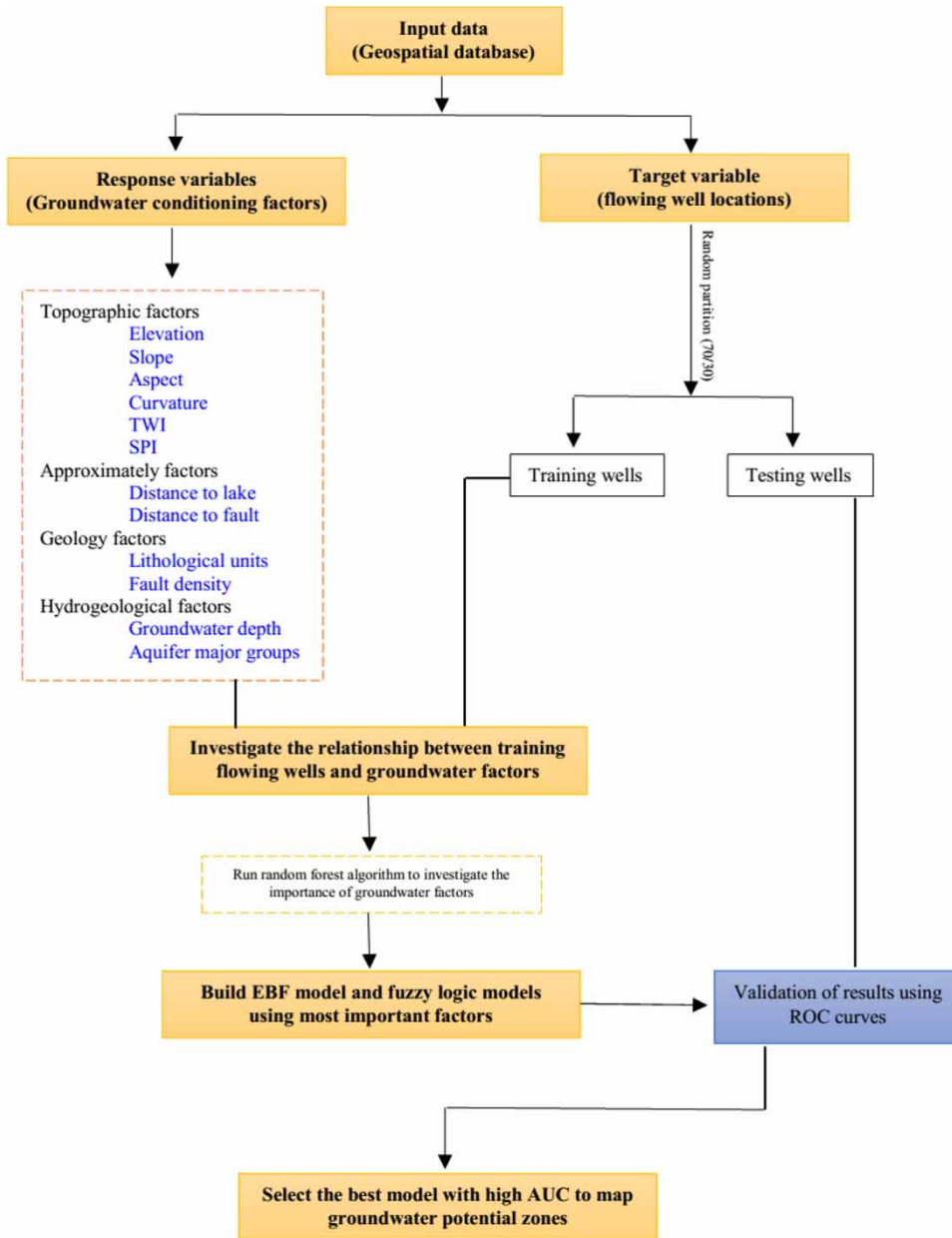
Figure 5 │ Flow chart for delineating flowing wells boundary using EBF and FL models.

condition were selected beyond this distance for every non-flowing well location. For the 130 selected grid cells, the values of groundwater factors were extracted using the extract multi-values to points module in ArcGIS 10.2™. The extracted values were stored as a comma delimited file (*.csv) and passed to R package to execute the RF algorithm.

For applying EBF and FL predictive models, the training flowing well inventory map was overlaid on the thematic layer of groundwater conditioning factors and the number of training wells for classes of each factor was determined. After that, the frequency ratio (FR) and fuzzy memberships were determined via Equations (14) and (15), respectively (Table 5). The mass functions of

Table 4 | Groundwater influential occurrence factors and their importance in the analysis of groundwater potential analysis (after Al-Abadi *et al.* (2016))

Factor	Its importance in groundwater potential
Elevation	Elevation is important in groundwater potential studies as climatic conditions vary at different elevations and this caused differences in soil and vegetation
Slope	Slope is a rise or fall of ground surface and it basically controls accumulation of water and thus controls groundwater recharge process
Curvature	Curvature is the second derivative of a surface and essentially affects the convergence and divergence of flow across the surface and hence affects the groundwater recharge process
Aspect	Aspect identifies the downslope direction of the maximum rate of change in value from each cell in a raster to its neighbors. Aspect strongly affects hydrologic processes via evapotranspiration, direction of frontal precipitation, and thus affects weathering process and vegetation and root development, especially in drier environments
TWI and SPI	These topographic indices have a basic role in the spatial variation of hydrological conditions such as soil moisture and groundwater flow
Distance to lake	Groundwater aquifers drain water to the river and lakes as a final destination, so this factor is considered in this study
Lithology	The hydraulic characteristics of the aquifer are strongly affected by variation in lithology. Therefore, it is considered an important factor in groundwater studies
Distance to faults and fault density	The rock fractures and other discontinuities promote groundwater movement and storage and thus allow groundwater storage to accumulate
Groundwater depth and aquifer major groups	These factors are directly related to aquifer storage and ground movement and play a significant role in groundwater potential

the EBF model were also computed for all classes of conditioning factors using Equations (5)–(10).

MODELING TECHNIQUES

Evidential belief functions

The EBF is a generalization of the Bayesian theory of subjective probability. Whereas a Bayesian accesses probabilities directly for the answer to a question of interest, a belief-function user assesses probabilities for related questions and then considers the implications of these probabilities for the question of interest (Dempster 1968). A brief mathematical description of the EBF is summarized here (An *et al.* 1992; Park 2011; Bui *et al.* 2012). Let X be the universe (the set which represents all the possible states of a system under study), then the power set 2^X is the set of all subsets of X including the empty set $\varnothing$. Now, if $X = \{a, b\}$, then $2^X = \{\varnothing, \{a\}, \{b\}, X\}$. Here, the elements of the power set are taken for representing propositions concerning the actual state of the system

being considered. A belief mass to each element of the power set is assigned using the theory of evidence as $m:2^X \rightarrow [0, 1]$. This equality is called a basic belief assignment and it has the following two properties: (i) the mass of the empty set is equal to zero ($m(\varnothing) = 0$); and (ii) the masses of the remaining member of the power set add up to a sum of $1 (\sum_{A \subseteq X} m(A) = 1)$. The upper and lower bounds of a probability interval can be defined according to the mass assignments. This interval is bounded by two non-additive continuous measures called belief and plausibility: $Bel(A) \leq P(A) \leq PL(A)$. The $Bel(A)$ for a set A is the total of all masses of subsets of the interested set.

$$Bel(A) = \sum_{B|B \subseteq A} m(B) \tag{1}$$

In the same manner, the $PL(A)$ is the aggregate of the all masses of the set B that intersect the A:

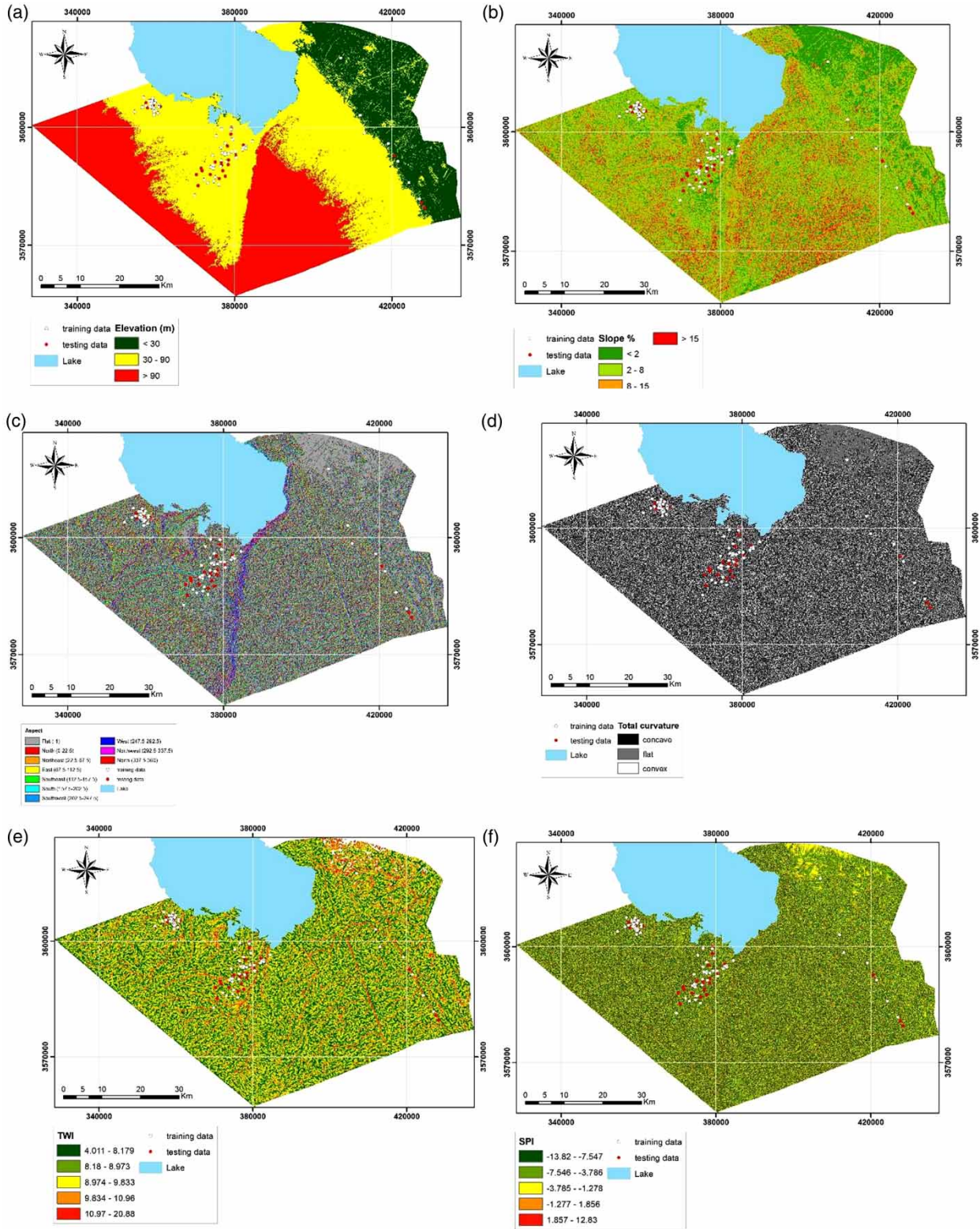$$PL(A) = \sum_{B|B \cap A \neq \varnothing} m(B) \tag{2}$$

Figure 6 | Groundwater condition factors: (a) elevation (m), (b) slope (%), (c) slope aspect, (d) curvature, (e) TWI, (f) SPI, (g) lithological units, (h) distance to faults (km), (i) fault density (km/km²), (j) distance to lake (km), and (k) groundwater depth (m). (*Continued.*)
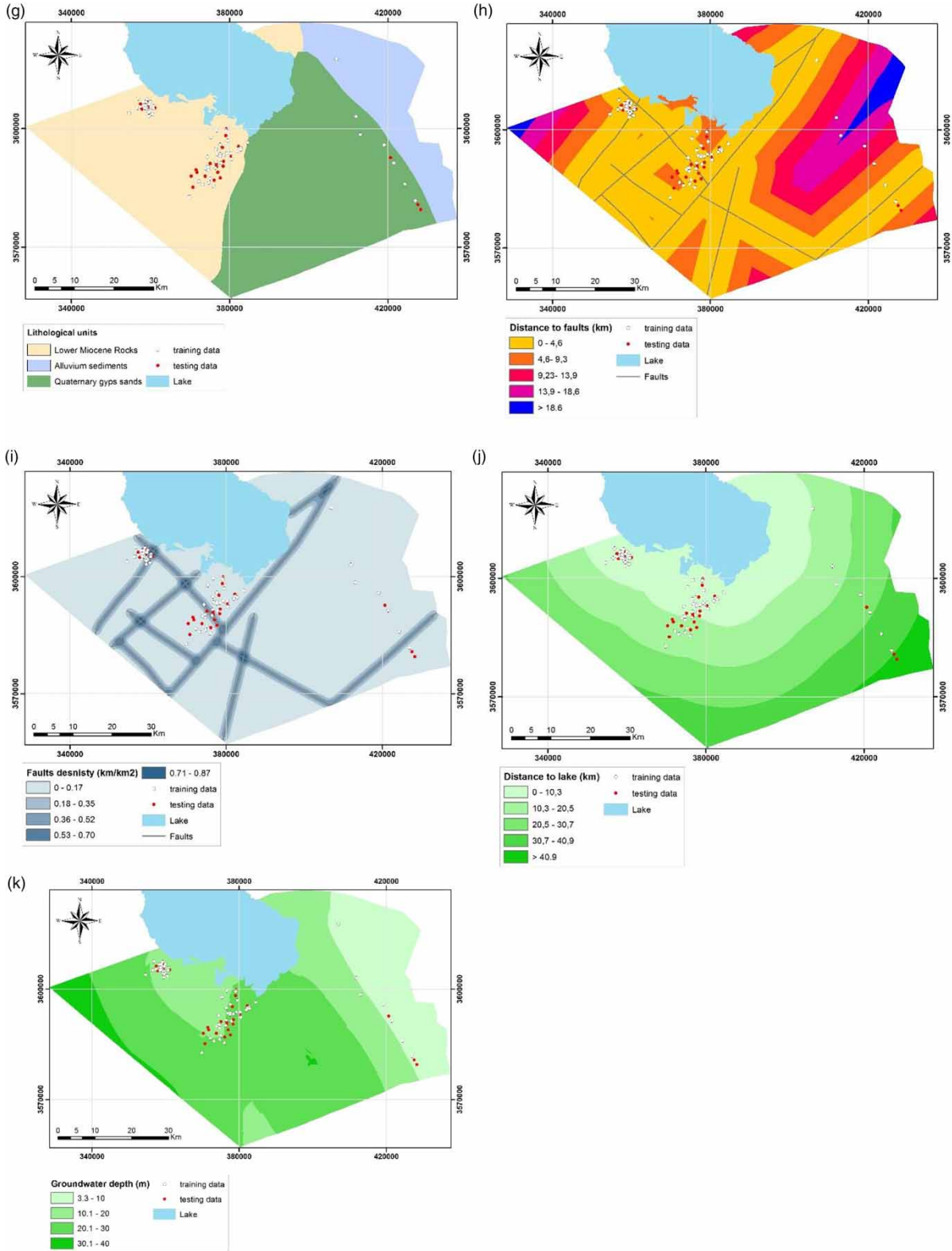
Figure 6 │ Continued.

Table 5 │ Values of fuzzy membership and EBFs for classes of groundwater factors

Factor	Classes ①	Class pixels ②	Pixels% ③ (a)	No. of wells ④	Wells% ⑤ (b)	FR ⑥	Fuzzy value ⑦	Mass functions ⑧ Bel	Dis	Unc	Pls
Elevation (m)	<30	857,997	0.191	6	0.092	0.484	0.302	0.202	0.400	0.398	0.600
	30–90	2,131,730	0.474	59	0.908	1.916	0.900	0.798	0.063	0.139	0.937
	>90	1,511,190	0.336	0	0.000	0.000	0.100	0.000	0.537	0.463	0.463
Distance to faults (km)	0–4.6	2,407,991	0.535	57	0.877	1.639	0.900	0.670	0.057	0.273	0.943
	4.6–9.3	1,014,359	0.225	5	0.077	0.341	0.267	0.140	0.255	0.605	0.745
	9.3–13.9	557,719	0.124	1	0.015	0.124	0.161	0.051	0.241	0.709	0.759
	13.9–18.6	405,912	0.090	2	0.031	0.341	0.267	0.139	0.228	0.632	0.772
	18.6–23.2	114,936	0.026	0	0.000	0.000	0.100	0.000	0.220	0.780	0.780
Distance to lake (km)	0–10.3	1,011,930	0.225	47	0.723	3.216	0.900	0.748	0.071	0.181	0.929
	10.3–20.5	1,215,858	0.270	14	0.215	0.797	0.298	0.185	0.214	0.601	0.786
	20.5–30.7	1,392,139	0.309	2	0.031	0.099	0.125	0.023	0.279	0.698	0.721
	30.7–40.9	739,108	0.164	2	0.031	0.187	0.147	0.044	0.231	0.726	0.769
	40.9–51.2	141,882	0.032	0	0.000	0.000	0.100	0.000	0.205	0.795	0.795
Lithological units	Miocene rocks	1,667,149	0.370	57	0.877	2.367	0.900	0.879	0.064	0.056	0.936
	Alluvium	761,079	0.169	1	0.015	0.091	0.100	0.034	0.391	0.576	0.609
	Alluvial fan	2,072,612	0.460	7	0.108	0.234	0.150	0.087	0.545	0.368	0.455
Groundwater depth (m)	3.3–10	855,958	0.190	6	0.092	0.485	0.291	0.149	0.275	0.576	0.725
	10–20	1,156,934	0.257	34	0.523	2.035	0.900	0.623	0.158	0.220	0.842
	20–30	2,313,984	0.514	25	0.385	0.748	0.394	0.229	0.311	0.460	0.689
	30–40	173,306	0.039	0	0.000	0.000	0.100	0.000	0.256	0.744	0.744
Aquifer group	Group 4	2,550,831	0.567	58	0.892	1.574	0.900	0.863	0.137	0.000	0.863
	Group 10	1,942,341	0.432	7	0.108	0.250	0.100	0.137	0.863	0.000	0.137

Notes: ① is the classes of each factor. ② is the number of pixels for each class (summation of these pixels gives the total number of pixels in the study area and equal to 4,500,917. (a) percentage of pixels, for example, for the first class of slope factor, the number of pixels = 857,997 and divided by the total number of pixels (4,500,917) and thus the percentage of this class is (857,997/4,500,917 = 0.191). ④ is the number of wells for each class (the total number of training wells is 65). ⑤ percentage of wells (for example, for the first class of slope factor the number of wells equals to 6 and divided by 65 gives 0.092, and so on. ⑥ is the ⑤/③ (b/a). ⑦ is obtained by applying Equation (14). ⑧ is obtained by applying Equations (5)–(10).

The $Bel(A)$ and $PL(A)$ are mathematically related as:

$$PL(A) = 1 - Bel(\bar{A}) \tag{3}$$

Once mass functions are calculated, Dempster's rule of combination is used to obtain the integrated mass functions (Dempster 1968). This rule of combination is a generalized scheme of Bayesian inference to aggregate evidence provided by disparate sources. Suppose that m_1 and m_2 are the basic probability assignments based on entirely distinct bodies of evidence D_1 and D_2. The belief functions Bel_1 and Bel_2 for the basic probability assignments m_1 and m_2, can be combined to generate a new belief function. Thus, for all $A_1, A_2, \ldots A_n \in 2^\Theta$, Dempster's rule produces a new probability assignment defined by $m(\phi) = 0$ and

$$m(H) = (1-k)^{-1} \sum_{\substack{i,j \\ A_i \cap B_j = H}} m_1(A_i)m_2(B_j) \tag{4}$$

where

$$k = \sum_{\substack{i,j \\ A_i \cap B_j = H}} m_1(A_i)m_2(B_j) < 1$$

For all non-empty $H \subset \Theta$. The quantity $(1-k)$ is a normalizing factor to compensate for the measure of belief committed to the empty set (Carranza *et al.* 2008).

In GIS practical applications, the maps of spatial evidences $A_i(i = 1, 2, \ldots, n)$ each with a specified number of variable classes $C_{ij}(j = 1, 2, \ldots, m)$ are given. If we assume that $N(T)$ is the total number of pixels of a given map (for example, elevation), $N(F)$ is the total number of flowing wells, $N(C_{ij})$ is the number of pixels in the variable class C_{ij}, and $N(C_{ij} \cap F)$ is the number of flowing wells in the class C_{ij}. The belief mass function for class C_{ij} is

estimated as (Carranza *et al.* 2008):

$$Bel_{C_{ij}} = \frac{W_{C_{ij}F}}{\sum_{j=1}^{m} W_{C_{ij}F}} \quad (5)$$

where

$$W_{C_{ij}F} = \frac{N(C_{ij} \cap F)/N(C_{ij})}{(N(F) - N(C_{ij} \cap F))/(N(T) - N(C_{ij}))} \quad (6)$$

The degree of disbelief mass function for the class C_{ij} is:

$$Dis_{C_{ij}} = \frac{W_{C_{ij}\bar{F}}}{\sum_{j=1}^{m} W_{C_{ij}\bar{F}}} \quad (7)$$

where

$$W_{C_{ij}\bar{F}} = \frac{(N(C_{ij}) - N(C_{ij} \cap F))/N(C_{ij})}{N(T) - N(F) - (N(C_{ij}) - N(C_{ij} \cap F))/(N(T) - N(C_{ij}))} \quad (8)$$

The uncertainty ($Unc_{C_{ij}}$) and plausible ($Pls_{C_{ij}}$) mass functions are calculated using Equations (9) and (10):

$$Unc_{C_{ij}} = 1 - Dis_{C_{ij}} - Bel_{C_{ij}} \quad (9)$$

$$Pls_{C_{ij}} = 1 - Dis_{C_{ij}} \quad (10)$$

The values of $Bel_{C_{ij}}$ and $Pls_{C_{ij}}$ are between 0 and 1.

Fuzzy logic

The FL is a cognitive artificial intelligence technique initiated in 1965 by Professor Lotfi Zadeh at the University of California, USA. In a broad sense, FL is viewed as a system of concepts, principles, and methods for dealing with modes of reasoning that are approximate rather than exact. It provides a means of representing uncertainties and vagueness that characterize human perception, judgmental reasoning, and decisions. Contrary to the crisp set, the boundary of a fuzzy set is not precise. Crisp sets only allow full membership, whereas fuzzy sets allow partial membership which can take values ranging from 0 to 1:

$$\mu_A(x) = X \rightarrow [0, 1] \quad (11)$$

where X is the universal set defined in a specific problem and $\mu_A(x)$ is the grade of membership for elements x in fuzzy set A (Yager & Zadeh 1992). The most important concept in FL is the definition of the membership functions for input and output data sets.

Fuzzy inference system (FIS) is the process of formulating from a given input space to an output space using FL. Basically, the FIS consists of four interconnected components (Figure 7): fuzzification module, knowledge base, inference engine, and defuzzification module. In the fuzzification step, the input space (crisp numbers) is transformed into fuzzy sets through applying the fuzzification function. In the knowledge base part, the IF-THEN rules provided by experts are formulated and stored. The IF part of a rule is its *antecedent*, and the THEN part of a rule is its
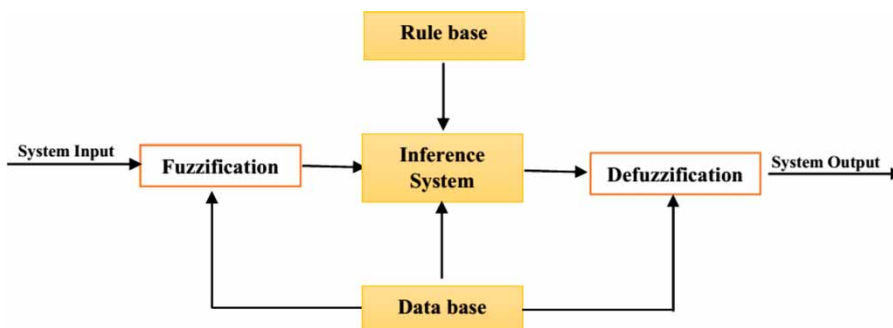

Figure 7 │ Fuzzy inference system diagram.

consequent. The human reasoning process is simulated by making fuzzy inference on the inputs and formulated IF-THEN rules using an inference engine. Finally, the fuzzy set obtained by the inference engine is transformed into a crisp value through the defuzzification module.

Applying fuzzy principles in GIS-related groundwater potential studies requires three main steps: (i) identification of groundwater occurrence factors and their classes; (ii) assigning fuzzy membership function values; and (iii) integration of the layer classes (modified after Bui *et al.* (2012)). Assigning fuzzy membership values is crucial to a proper fuzzy model. In this context, various approaches have been developed which can be either classified into knowledge-based, data-driven, or a combination of both. In the first method, values of fuzzy membership are mainly determined using expert opinion, while in the second method, fuzzy memberships are assigned through investigating the relationship between groundwater borehole locations and groundwater conditioning factors used in the analysis. The most widely used approaches for assigning fuzzy membership values in earth science are the cosine amplitude (CA) method and the FR method. The CA method formulates fuzzy membership values according to the following equation:

$$CA = \frac{N(C_{ij} \cap F)}{\sqrt{N(T) \cdot N(F)}} \tag{12}$$

In contrast, the FR is calculated as:

$$FR_{ij} = \frac{N(C_{ij} \cap F)/N(F)}{N(C_{ij})/N(T)} \tag{13}$$

Equation (12) gives values in the range of (0, 1), and thus the output values do not have to be normalized. To transform FR to membership values, the output of Equation (13) is normalized into (0, 1) interval using the Max-Min normalization process as:

$$\mu(C_{ij}) = \frac{FR_{ij} - Min(FR_{ij})}{Max(FR_{ij}) - Min(FR_{ij})} \left[max(\mu(C_{ij})) - min(\mu(C_{ij})) \right] \\ + min(\mu(C_{ij})) \tag{14}$$

where $\mu(C_{ij})$ is the fuzzy membership value; $Max(FR_{ij})$ and $Min(FR_{ij})$ are the upper and lower normalization bounds. In this paper, the values of $Max(FR_{ij})$ and $Min(FR_{ij})$ are taken as 0.9 and 0.1, respectively (Pradhan 2011). The FR method was selected here to calculate fuzzy membership values because the CA method gives very small values and thus is not regarded as suitable for further analysis.

Once the fuzzy membership values are calculated for the selected conditioning factors of groundwater occurrence, the next step is to integrate these layers to produce a groundwater potential zoning map using the fuzzy combining operators. There are different types of fuzzy combining operators, for example, fuzzy AND, fuzzy OR, fuzzy SUM, fuzzy PRODUCT, and fuzzy GAMMA can be used for integrating layers in the GIS environment. The fuzzy AND combines fuzzy membership values using the minimum operators as:

$$\mu_{AND} = MIN\left[\mu(C_{ij})\right] \tag{15}$$

while fuzzy OR combines fuzzy membership value using the fuzzy maximum operator as:

$$\mu_{OR} = MAX\left[\mu(C_{ij})\right] \tag{16}$$

Fuzzy SUM is complementary to the fuzzy product and is defined as:

$$\mu_{SUM} = 1 - \prod_{i=1}^{n} \left[1 - \mu(C_{ij})\right] \tag{17}$$

The output is always greater than the largest contributing fuzzy membership value. Fuzzy PRODUCT combines fuzzy membership by multiplying their values according to this formula:

$$\mu_{PRODUCT} = \prod_{i=1}^{n} \mu(C_{ij}) \tag{18}$$

And finally, the fuzzy GAMMA is defined as:

$$\mu_{GAMMA} = Fuzzy\ SUM^{\lambda} \cdot Fuzzy\ PRODUCT^{1-\lambda} \tag{19}$$

The values of λ are chosen to be in the range of [0, 1].

RF machine learning technique

RF is an ensemble machine learning technique to solve problems of both classification and regression types. It is a model to estimate either the Bayes classifier or the regression function. The RF algorithm handles random binary trees which use a subset of the observations through bootstrapping techniques; from the original data set a random choice of the training data is sampled and used to build the model (Breiman 2001). Data that are not included in the model training are named as 'out-of-bag' (OOB) (Catani *et al.* 2013). The most attractive aspect of the RF model is the ability to estimate the importance of a variable by looking for how much the prediction error increases when the OOB error for that variable is permuted while all other variables are fixed unchanged (Liaw & Wiener 2002). This capability is applied here to investigate the relative importance of the groundwater conditioning factors used in the analysis of groundwater potentiality, a critically important but often neglected aspect of groundwater potentiality studies.

Relative operating characteristic for model validation

It is well known that any predictive model needs validation (verification) before it can be effectively utilized for prediction. The relative operating characteristic (ROC) curve was used here to validate the developed models. The ROC is a powerful plot for determining the predictive capability of systems. The area under the ROC curve (AUC) characterizes the quality of a forecast system by describing the system's ability to anticipate correctly the occurrence and non-occurrence of predefined events (Mason & Graham 2002). In the ROC curve, the true positive rate (sensitivity) is plotted against the false positive rate (1 – specificity). The AUC ranges from 0.5 (random prediction represented by a diagonal straight line) to 1 (perfect prediction). More specifically, the relation between AUC and model prediction accuracy is as follows: poor (0.5–0.6); average (0.6–0.7); good (0.7–0.8); very good (0.8–0.9); and excellent (0.9–1). The AUC is often computed for both success and prediction rates. The success rate explains how well the resulting prediction map classified the area of existing flowing wells, while the prediction rate is utilized as a measure for

predictive capability of the model and only used the testing data set (unseen data) to investigate the prediction performance. In this study, the AUC for prediction rate was only demonstrated because it is more representative of model performance.

MODEL SETUP, RESULTS, AND DISCUSSION

Selection of the most important factors using RF algorithm

To execute the RF algorithm requires that two main parameters should be determined: *ntree* and *mtry*. *ntree* is the total number of growing trees, while *mtry* is the number of factors that are randomly chosen at each split node (Al-Abadi & Shahid 2016; Zabihi *et al.* 2016). To minimize the generalization error and correlation among growing decision trees, Breiman (2001) suggested that *mtry* should be less than $\log2(M+1)$, in which M is the number of predictors (groundwater conditioning factors) used in the analysis. As the total number of predictors used in the current research was 12, the *mtry* should be less than 3 ($int(\log2(12+1))$). In this study, *ntree* was taken equal to 1,000 as this value gives more stable prediction and a relatively low prediction error (Rodriguez-Galiano & Chica-Olmo 2012). Two parameters were used here to assess the importance of factors in the analysis of groundwater potentiality: mean decrease accuracy (*MeanDecreaseAccuracy*) and mean decrease in Gini coefficient (*MeanDecreaseGini*) (Naghibi *et al.* 2017). The first index measures the decrease in model fit as the specified factor is removed from the analysis, while Gini coefficient measures how each variable contributes to the homogeneity of the nodes and leaves in the RF results. The plot of these measures for the first run of the RF algorithm is shown in Figure 8. This plot identified elevation, groundwater depth, geology, distance to lake, aquifer groups, and distance to faults as the most importance factors. The Gini index also identified the previous factors as the most important factors, but with a different rank. The less effective factors, i.e., slope angle, curvature, aspect, TWI, SPI, and fault density were dropped from the analysis and the RF model was rerun only with the most important factors.
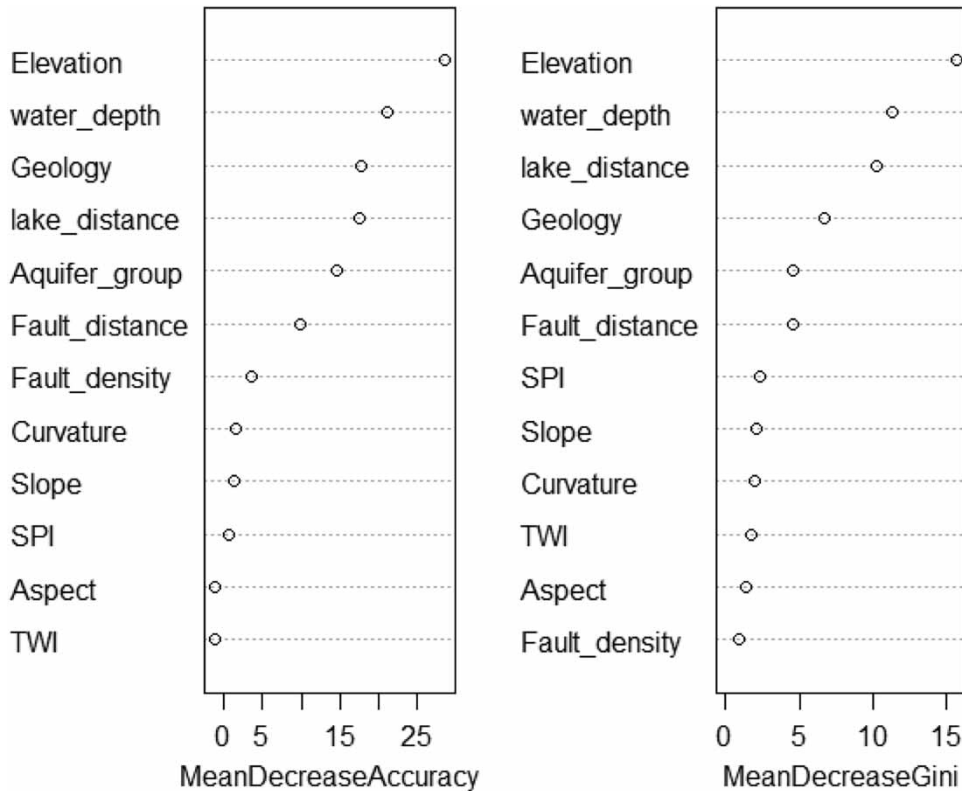
Figure 8 | Mean decrease accuracy and mean decrease GINI.

Table 6 | Accuracy of RF models

Parameters	Developed RF models	
	All variables	The most important variables
OOB error	0.0625	0.0547
Correctly classified instances	120 (92.96%)	123 (95.31%)
Incorrectly classified instances	10 (7.03%)	7 (4.68%)
Kappa statistic	0.86	0.91
Mean absolute error	0.13	0.09
Root mean square error	0.23	0.19

The overall accuracy for the two runs is summarized in Table 6. It is clear from Table 4 that excluding the less important factors increases the FR model accuracy from 93% to 95%. All other error statistics support this conclusion. Therefore, the less important factors were removed from the analysis and the rest of the factors were used to build the EBF and FL models.

Groundwater potential mapping using EBF

A high value of Bel mass functions means a high probability of existing artesian condition, while a low value means a low probability. It can be seen from Table 5 that the elevation range (30–90 m) has the highest value of Bel (0.798) and the lowest values of Dis (0.064), indicating the highest probability of artesian potential. The elevation values <30 m have a relatively low value of Bel (0.202) and a high value of Dis (0.40), meaning that this class has a low probability of artesian potential. The last class of elevation (>90) has a Bel value equal to 0, indicating a low probability of groundwater flowing wells conditions in this class (Park 2011). For distance to fault, the only class having the highest probability of groundwater potentiality is the first class (0–4.6 km), as this class has a high Bel (0.670) and a low Dis (0.057). The rest of the classes generally have relatively low values of Bel, thus implying low probability of groundwater occurrence. This result confirms the importance of a structural setting to control groundwater movement and occurrence in the study
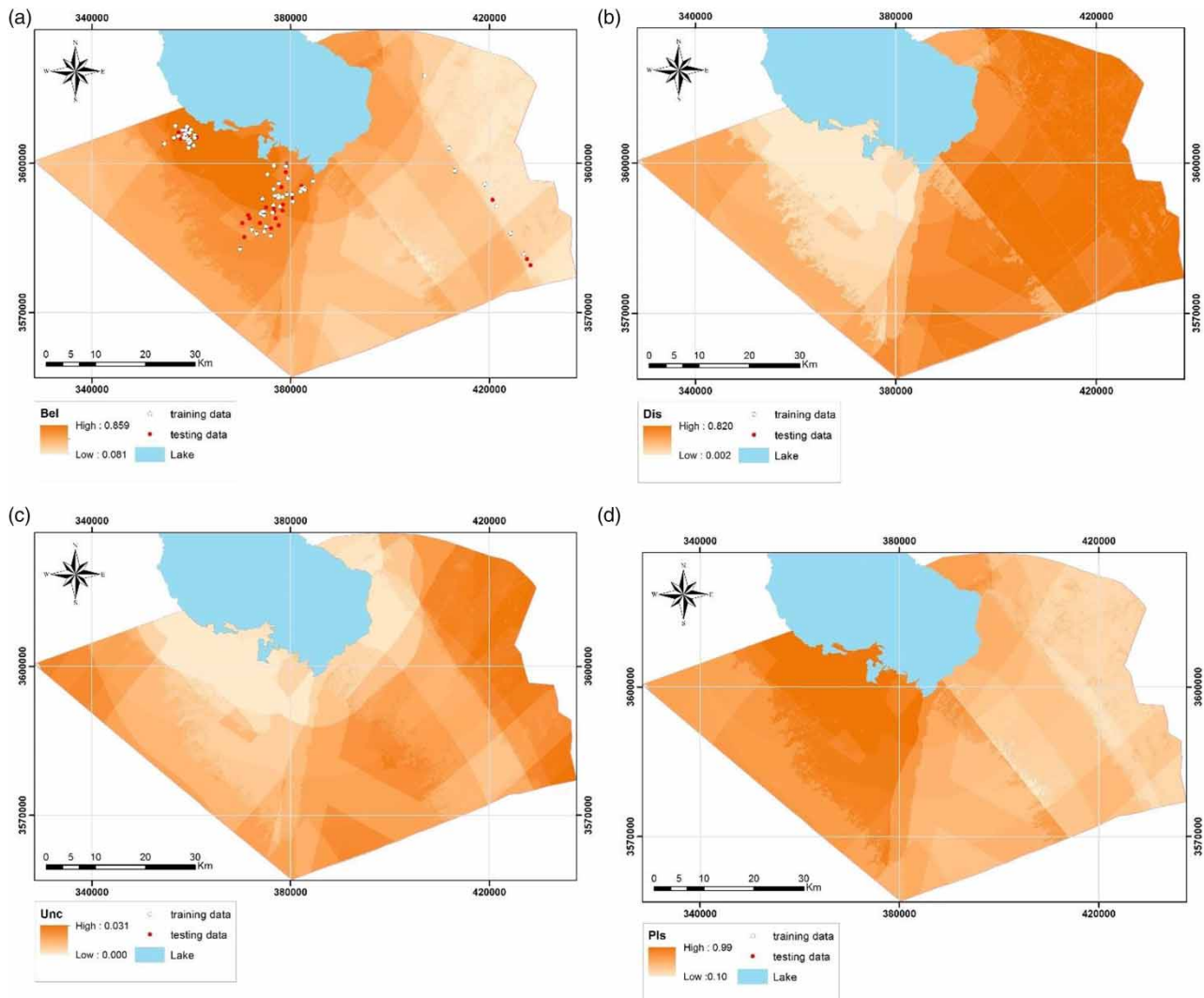
Figure 9 │ EBF mass functions: (a) belief, (b) disbelief, (c) uncertainty, and (d) plausibility.

area. In the case of distance to Rezzaza Lake, the only high Bel value (0.748) is seen for the range (0–10.3 km), indicting a high probability of groundwater artesian potentiality. The remaining classes have low Bel values, indicating low probability of groundwater occurrence. In the case of lithology, the Micocene rocks have high Bel values and thus high probability of artesian condition. In contrast, Quaternary sediments have low Bel values that indicate low probability of getting artesian conditions. For groundwater depth layer, the high Bel and low Dis values for the range (10–20 m) indicate a direct relationship with groundwater potential. The remaining classes have low values of Bel, implying a low probability of groundwater occurrence. For the last factor,

aquifer group, the high probability of groundwater potentiality is associated with the major group 4 with Bel value equal to 0.863. This group represents the karstified and highly fractured limestone with high confining pressure conditions and, therefore, there is a direct relationship with artesian conditions between this group and Bel mass function. The major group 10 has a low Bel value (0.137) and, therefore, a low probability of artesian potential condition.

Once the mass function of EBF was calculated, Dempster's rule of combination was utilized to get the EBF mass functions. Combining the mass functions is an iterative process. Therefore, the two mass functions of elevation and distance to faults conditioning factors were first combined.

The resulting mass functions of the first combining process was combined with distance to lake mass function using Dempster's rule to produce the next integrated mass function, and so on. In total, five combining processes were executed to obtain the final integrated mass functions (Figure 9(a)–9(d)). Comparing between *Bel* (Figure 9(a)) and *Dis* (Figure 9(b)) showed that there is a negative relationship between these mass functions. In areas with high *Bel*, the *Dis* values are low and vice versa. On the other hand, the *Unc* map (Figure 9(c)) reveals a lack of information supporting uncertainty in producing a potential map (Nampak *et al.* 2014). From Figure 9(c) and compared to the *Bel* map (Figure 9(a)), the *Unc* values are high in areas with low values of *Bel*. The *Pls* map (Figure 9(d)) is similar to the *Bel* map except that the contrast between lower and higher degrees was less apparent than the Bel map.

The integrated *Bel* values were used to provide a groundwater flowing well potential index (GFWPI) and were classified into five classes ranging from very low to very high using a natural break classification scheme (Figure 10(a)). There are many different classification systems used in groundwater potential mapping such as the natural breaks, quantiles, equal interval, standard deviation, and manual methods. The most common method for a reclassification system in groundwater potential studies is natural breaks, which is used in this study.

Groundwater potential mapping using FL

In the case of the FL model, the relative importance of condition factors in groundwater potentiality can be deduced from the fuzzy membership values; high values mean high groundwater potentiality and vice versa. From Table 5, the high fuzzy membership values (equal to 0.9) are associated with elevation in range (30–90), distance to faults in range (0–4.6 km), distance to lake range (0–10.3 km), Miocene rocks, groundwater depth range (10–20 m), and aquifer group 4. The remaining fuzzy membership values are relatively less than 0.5 and therefore are regarded as less important in determining the artesian condition in the study area. To generate the GFWPI using the FL technique, the fuzzy membership values of six factors were integrated using the FL overlay method in ArcGIS 10.2™ environment.

Five fuzzy combining operators were used in this study, namely, fuzzy AND, fuzzy OR, fuzzy SUM, fuzzy PRODUCT, and fuzzy GAMMA. For fuzzy GAMMA operator, five lambda λ values were used. These are 0.1, 0.3, 0.5, 0.7, and 0.9. For each fuzzy operator, GFWPI were generated and classified into five zones in similar fashion as the EBF model. A total of nine GFWPI maps were generated. For visualization purposes, only the maps of the fuzzy AND, fuzzy PRODUCT, and fuzzy GAMMA ($\lambda = 0.3$, 0.7, and 0.9) are presented in Figure 10(b)–10(f).

Validation of the results

The prediction rates of the developed models are shown in Figure 11. According to the AUC values, all of the prediction models built in this study have very good prediction accuracy. The results of validation tests revealed that the AND fuzzy models perform the best with AUC equal to 0.860 and the fuzzy SUM showed the lowest prediction accuracy (AUC = 0.831). The second highest capability model for spatial predicting groundwater artesian zone boundary was the EBF with AUC = 0.844.

Based on these results, the fuzzy AND was selected to demarcate groundwater artesian zone boundary in the Karbala Governorate. The fuzzy values of this model ranged from 0.1 to 0.9 and were classified into five classes: very low, low, moderate, high, and very high. The first two classes (very low and low) cover an area of about 50% (2,041 km^2) of the total area. In contrast, the moderate zone encompasses only 25% (994 km^2). The last two classes (high and very high) encompass an area of about 25% (1,016 km^2). The high and very high zones are mainly distributed in the northern part of the area of interest and close to Rezzaza Lake, while the remaining zones cover the rest of the study area.

CONCLUSIONS

In the study, the data-driven EBFs and data-knowledge FL approaches were used for mapping groundwater flowing well potential in the Karbala Governorate in central Iraq. Twelve groundwater conditioning factors with flowing well inventory maps were used to build the predictive models
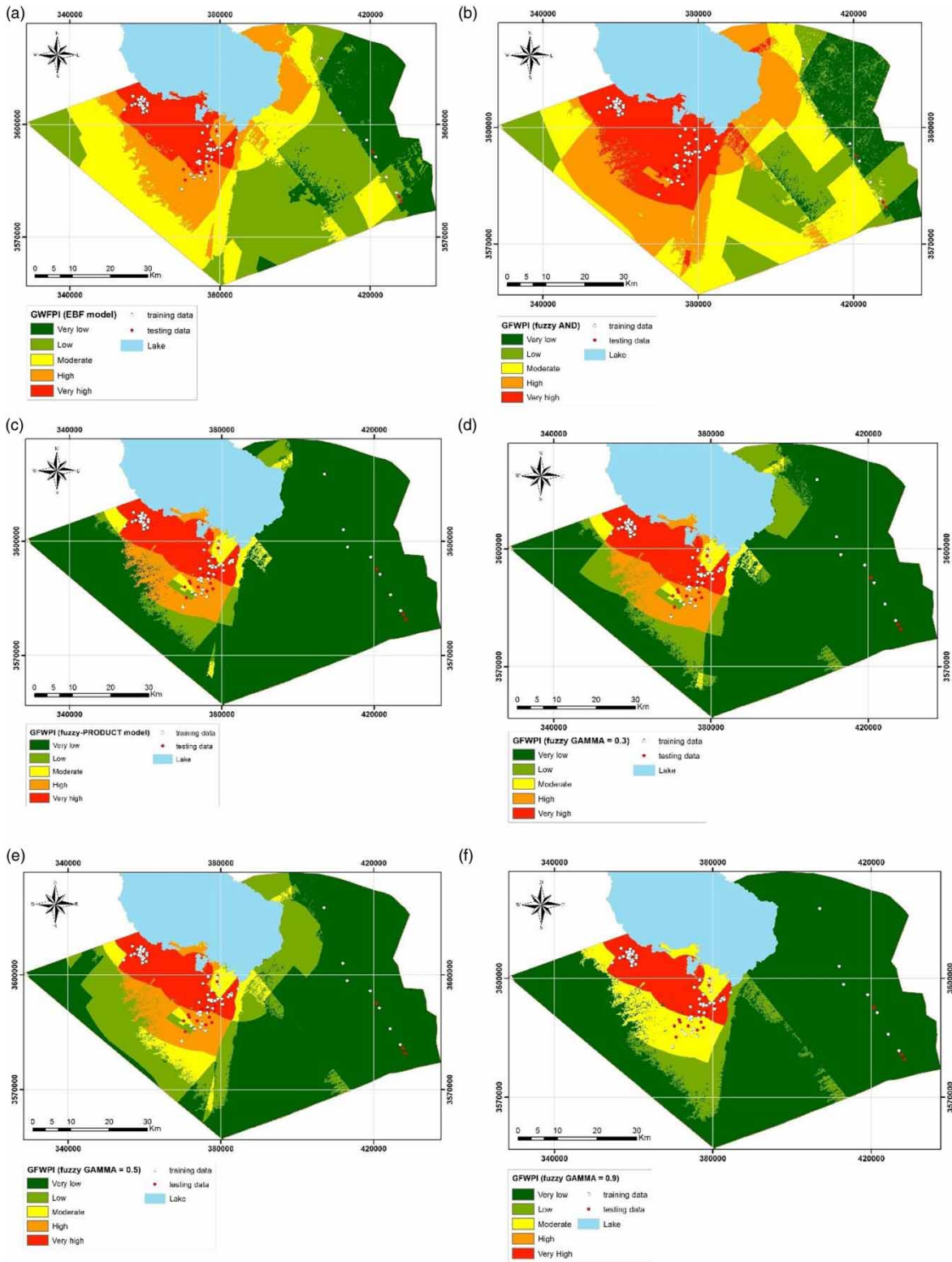
Figure 10 │ Groundwater flowing well potential index (GFWPI) using different predictive models: (a) EBF, (b) fuzzy AND, (c) fuzzy PRODUCT, and (d) to (f) fuzzy GAMMA ($\lambda = 0.3$, 0.5, and 0.9, respectively).
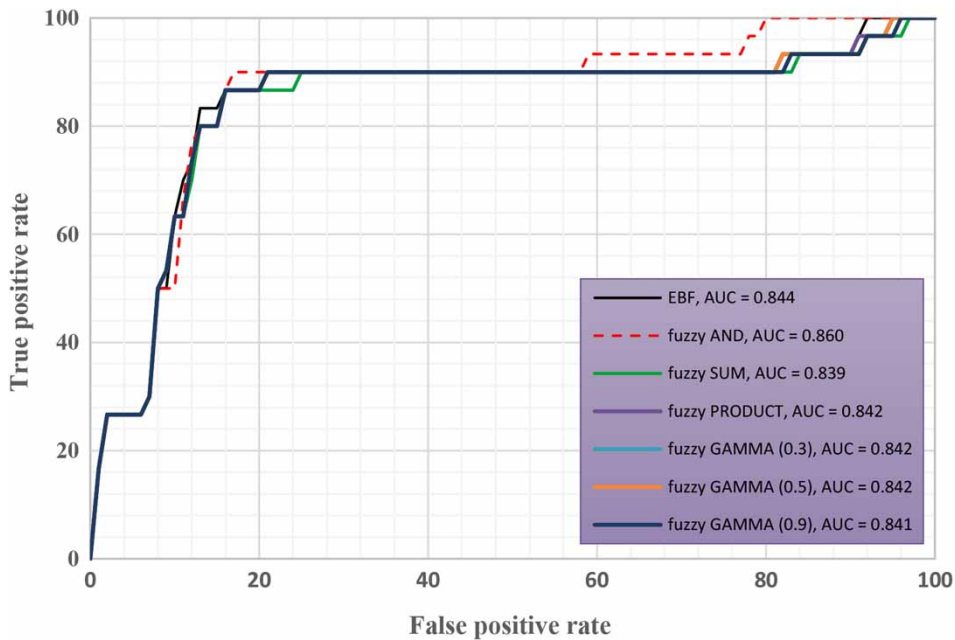
Figure 11 │ Validation results of predictive models.

in the GIS platform. The RF machine learning technique was used to investigate the contribution of groundwater conditioning factors in mapping groundwater flowing well potentiality. Results of the RF algorithm confirmed that six out of the twelve factors were most important and excluded the less important factors, which contributed in increasing the prediction accuracy from 92% to 95%. The FL and EBF models were built using the most important factors. The mass functions of EBFs were estimated through overlying a training flowing well inventory map with six groundwater conditioning factors. The FR was utilized to generate fuzzy membership values of each class in groundwater factors. The validation results using a ROCs curve showed that the best model was the AND fuzzy model followed by EBFs. The lowest accuracy model was fuzzy SUM (AUC = 0.431). Results also confirm that the integration of RF with EBFs and FL models contributes to developing predictive models with high prediction accuracy. Results also indicated that the northern part of the study area is a more promising area to drill wells with a higher probability of artesian conditions. The findings of this study could be used by decision-makers and hydrogeologists to develop groundwater and manage aquifer systems in a more efficient manner.

REFERENCES

Al-Abadi, A. M. & Shahid, S. 2016 Spatial mapping of artesian zone at Iraqi southern desert using a GIS-based random forest machine learning model. *Modeling Earth System and Environment* **2** (96), 1–17. doi:10.1007/s40808-016-0150-6.

Al-Abadi, A. M., Al-Temmeme, A. A. & Al-Ghanimy, M. A. 2016 A GIS-based combining of frequency ratio and index of entropy approaches for mapping groundwater availability zones at Badra–Al Al-Gharbi–Teeb areas, Iraq. *Sustainable Water Resource Management* **2** (3), 265–283. doi:10.1007/s40899-016-0056-5.

Al-Amiri, H. 1979 *Structural Interpretation of the Landsat Satellite Imagery for the Southern Desert of Iraq*. GEOSURV, Internal Report No. 988.

Al-Jiburi, H. K. & Al-Basrawi, N. H. 2015 Hydrogeological map of Iraq, Scale 1:100000, 2nd edition, 2013. *Iraqi Bulletin of Geology and Mining* **11** (1), 17–26.

An, P., Moon, W. M. & Bonham-Carter, G. F. 1992 On knowledge-based approach on integrating remote sensing, geophysical and geological information. In: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS '92)*, Houston, Texas, pp. 34–38.

Aouragh, M. H., Essahlaoui, A., El Ouali, A., El Hmaidi, A. & Kamel, S. 2016 Groundwater potential of Middle Atlas plateaus, Morocco, using fuzzy logic approach, GIS and remote sensing. *Geomatics, Natural Hazards and Risks*. doi:10.1080/19475705.2016.1181676.

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.

Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I. & Dick, O. B. 2012 Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* **96**, 28–40. doi:10.1016/j.catena.2012.04.001.

Carranza, E. J. M. & Laborte, A. G. 2014 Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computer and Geosciences* **74**, 60–70. doi:10.1016/j.cageo. 2014.10.004.

Carranza, E. J. M., Hale, M. & Faassen, C. 2008 Selection of coherent deposit-type locations and their application in data-driven mineral prospectivity mapping. *Ore Geology Reviews* **33**, 536–558. doi:10.1016/j.oregeorev.2007.07.001.

Catani, F., Lagomarsino, D., Segoni, S. & Tofani, V. 2013 Landslide susceptibility estimation by random forest technique: sensitivity and scaling issues. *Natural Hazards Earth System and Science* **13**, 2815–2831. doi:10.5194/nhess-13-2815-2013.

Chacón, J., Irigaray, C., Fernández, T. & El Hamdouni, R. 2006 Engineering geology maps: landslides and GIS. *Bulletin of Engineering Geology and the Environment* **65** (4), 341–411. doi:10.1007/s10064-006-0064-z.

Dempster, A. P. 1968 A generalization of Bayesian inference. *Journal of the Royal Statistical Society* **30** (2), 205–247.

Jassim, S. Z. & Goff, J. C. 2006 *Geology of Iraq.* Dolin, Prague and Moravian Museum, Brno, Czech Republic, 431 pp.

Liaw, A. & Wiener, M. 2002 Classification and regression by randomForest. *R News: The Newsletter of the R Project* **2** (3), 18–22. http://cran.r-project.org/doc/Rnews/.

Mason, S. J. & Graham, N. E. 2002 Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society* **128**, 2145–2166.

Mogaji, K. A., Lim, H. S. & Abdullah, K. 2014 Regional prediction of groundwater potential mapping in a multifaceted geology terrain using GIS-based Dempster-Shafer model. *Arabian Journal of Geoscience* **8** (5), 3235–3258. doi:10.1007/s12517-014-1391-1.

Naghibi, S. A., Pourghasemi, H. R. & Abbaspour, K. 2017 A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theoretical and Applied Climatology*. doi:10.1007/s00704-016-2022-4.

Nampak, H., Pradhan, B. & Manap, M. A. 2014 Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology* **513**, 283–300. doi:10.1016/j. jhydrol.2014.02.053.

Ozdemir, A. 2011 Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *Journal of Hydrology* **405**, 123–136. doi:10.1016/j.jhydrol. 2011.05.015.

Park, N. W. 2011 Application of Dempster–Shafer theory of evidence to GIS-based landslide susceptibility analysis. *Environmental Earth Science* **62** (2), 367–376. doi:10.1007/s12665-010-0531-5.

Park, I., Kim, Y. & Lee, S. 2014 Groundwater productivity potential mapping using evidential belief function. *Groundwater* **52** (S1), 201–207. doi:10.1111/gwat.12197.

Pourghasemi, H. R. & Beheshtirad, M. 2015 Assessment of a data-driven evidential belief function model and GIS for groundwater potential mapping in the Koohrang Watershed, Iran. *Geocarto International* **30** (6), 662–685. doi:10.1080/10106049.2014.966161.

Pradhan, B. 2011 Manifestation of an advanced fuzzy logic model coupled with geoinformation techniques to landslide susceptibility mapping and their comparison with logistic regression modelling. *Environmental and Ecological Statistics* **18**, 471–493. doi:10.1007/s10651-010-0147-7.

Rahmati, O., Pourghasemi, H. R. & Melesse, A. 2016 Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran. *Catena* **137**, 360–372.

Rather, J. A. & Andrabi, Z. A. 2012 Fuzzy logic based GIS modeling for identification of groundwater potential zones in the Jhagrabaria watershed of Allahabad district, Uttar Prahesh, India. *International Journal of Advanced Remote Sensing and GIS* **1** (2), 218–233.

Reineking, T. 2014 *Belief Functions: Theory and Algorithms.* Doctoral Thesis, Bremen University, Bremen, Germany, 165p.

Rodriguez-Galiano, V. & Chica-Olmo, M. 2012 Land cover change analysis of a Mediterranean area in Spain using different sources of data: multi-seasonal Landsat images, land surface temperature, digital terrain models and texture. *Applied Geography* **35** (1–2), 208–218. doi:10.1016/j.apgeog.2012.06.014.

Sissakian, V. K. 2000 *Geological Map of Iraq,* 3rd edn. Scale 1:1,000,000, GEOSURV, Bagdad, Iraq.

Tahmassebipoor, N., Rahmati, O., Noormohamadi, F. & Lee, S. 2005 Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arabian Journal of Geoscience* **9**. doi: 10.1007/s12517-015-2166-z.

Thiam, A. K. 2005 An evidential reasoning approach to land degradation evaluation: Dempster-Shafer theory of evidence. *Transactions in GIS* **9**, 507–520.

Yager, R. R. & Zadeh, L. A. 1992 *An Introduction to Fuzzy Logic Applications in Intelligent Systems.* Kluwer, Boston, MA, USA.

Zabihi, M., Pourghasemi, H. R., Pourtaghi, Z. S. & Behzadfar, M. 2016 GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environmental Earth Sciences* **75**, 665. doi:10.1007/s12665-016-5424-9.