



Groundwater potential assessment of the Sero plain using bivariate models of the frequency ratio, Shannon entropy and evidential belief function

SAEED KHOSHTINAT¹, BABAK AMINNEJAD^{2,*}, YOUSEF HASSANZADEH³ and HASAN AHMADI²

¹Ph.D. Candidate, Water Engineering, Department of Civil Engineering, Roudehen Branch, Islamic Azad University, Roudehen, Iran.

²Assistant Professor, Department of Civil Engineering, Roudehen Branch, Islamic Azad University, Roudehen, Iran.

³Professor, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran.

*Corresponding author. e-mail: aminnejad@riau.ac.ir

MS received 21 April 2018; revised 29 December 2018; accepted 4 January 2019; published online 4 June 2019

The goal of the present research is to evaluate three bivariate models of the frequency ratio, Shannon entropy (SE) and evidential belief function in the spatial prediction of groundwater at the Sero plain located in west Azerbaijan, Iran. In the first phase, well locations with groundwater yields $>11 \text{ m}^3/\text{hr}$ were identified (75 well locations). Ten groundwater conditioning factors affecting the occurrence of groundwater, namely, altitude, slope degree, curvature, slope aspect, rainfall, soil, land-use, geology and distance from the fault and the river, were selected for modelling. Finally, the groundwater potential map results were drawn from three implemented models and they were validated using testing data by area under the receiver operating characteristic curve (AUC). The AUCs of these models were 0.84, 81 and 85%, respectively. The results of the current study demonstrated that these models could be successfully employed for spatial prediction modelling. Moreover, the results of the SE model demonstrated that the most and the least important factors in groundwater occurrences in the area under study were altitude, curvature and rainfall, respectively. The results of this study are helpful for the Regional Water Authority of Urmia and the decision makers to comprehensively assess the groundwater exploration development and environmental management in future planning.

Keywords. Groundwater potential; frequency ratio; Shannon entropy; evidential belief function; Sero plain hydrology and water resource; remote sensing.

1. Introduction

Groundwater is defined as the main source of water demand in arid and semiarid areas. In comparison with surface water, the advantages of groundwater are having almost constant temperature, being less affected by drought, having better water quality and being tapped in when necessary, to mention

but a few (Manap *et al.* 2013). Groundwater is always exploited when there is inadequate surface water. Nowadays, the use of groundwater is increasing rapidly. Therefore, the identification of a proper method for the management and prediction of groundwater at the national, regional and local levels is necessary (Vaux 2011; Le Page *et al.* 2012). In Iran, as a semi-arid country of which

two-third is desert land without any green pasture, the water demand has been satisfied using groundwater (Bastani *et al.* 2010; Mehrdadi 2010; Nosrati and Van Den Eeckhaut 2012). The life of more than 70% of the rural population in Iran, for drinking and domestic requirements, depends on groundwater (Rahmati *et al.* 2015). Thus, groundwater potential mapping (GPM) in Iran is really necessary. Complex and costly instruments and methodology are required to gain information regarding groundwater resources. GPM is useful for the management of spatial groundwater resources, especially in regions with little data. The probability of groundwater occurrence in an area is defined as the groundwater potential (Jha *et al.* 2010). Traditional methods for groundwater potential identification and exploitation were accomplished through drilling. Moreover, geophysical methods involve more cost and time (Todd and Mays 1980; Roscoe 1990; Israil *et al.* 2006; Jha *et al.* 2010).

Nowadays, remote sensing and geographical information systems (GIS) are frequently used in studies related to water resources. Thus, many researchers have performed studies by considering some conditioning factors such as topography and geomorphology; drainage pattern; lineaments such as fault, lithology and geology and soil (Prasad *et al.* 2008; Chowdhury *et al.* 2009; Jasmin and Mallikarjuna 2011; Deepika *et al.* 2013; Chen *et al.* 2019). GIS is a useful tool with some advantages to manage huge amounts of data (Khosravi *et al.* 2016). Many studies have been carried out using the frequency ratio (FR) (Oh *et al.* 2011; Ozdemir 2011a; Manap *et al.* 2014; Davoodi *et al.* 2015; Naghibi *et al.* 2015), weights of evidence (Corsini *et al.* 2009; Lee *et al.* 2012; Pourtaghi and Pourghasemi 2014), logistic regression (Ozdemir 2011b), analytical hierarchy process (Awawdeh *et al.* 2014; Kaliraj *et al.* 2014) and evidential belief function (EBF) (Nampak *et al.* 2014). The results of all studies indicated that data-driven models have reasonable prediction ability and can be used in other areas to solve problems. In data-driven models, expert opinion does not have any impact on the results. Therefore, the results are more reliable.

The main aims of the current study are as follows: (i) spatial prediction of regions with high groundwater potential and (ii) identification and comparison of prediction abilities associated with the FR, Shannon entropy (SE) and EBF models in GPM. Evaluation of the GPM would be practical to the decision makers in groundwater management

and restoration, e.g., the Regional Water Authority of Urmia (RWAU), and in determining suitable locations for future drilling of productive wells. However, such studies have not been carried out in the entire province up to now. Thus, the present research is a leading work in this research area and plays a significant role in quick groundwater assessment.

2. Study area description

The Sero plain is located in the west Azerbaijan province in Iran. The plain is considered as one of the most important ones in the province upon which people are dependent for groundwater supply for both drinking and agricultural activities. It lies between $44^{\circ}37'30''$ – $44^{\circ}43'30''$ E and $37^{\circ}41'$ – $37^{\circ}47'$ N, and covers about 52 km² (see figure 1). Topographically, the altitude varies between 1490 and 1697 m above the sea level and the highest and lowest slopes are 26 and 0, respectively. Based on the reports of the RWAU, the mean annual precipitation in the region is about 440 mm. The most dominant soil is inceptisol soil and agricultural land use covers about 94% of the study area. Geologically, the Quaternary and Permian cover about 58% and 18% of the study area, respectively. Lithologically, most of the area is covered by K2l (light blue to pink, thick-bedded to massive, piagic limestone with calcite veins), OPm (ophiolite melange of ultrabasic rocks, serpentinite, meta gabbro, berreciated volcanic, grey to red shale and plagic limestone), Pd (sandstone and conglomerate), CM (tectonic mixture of serpentinitised ultramafic rocks, pillow lavas, gabbros, radiolarite and pelagic limestone), CMV (tectonic mixture mostly composed of pillow lava and sediments) and K (alternation of sandstone, shale and limestone at base followed by well grey limestone at the top (probably cretaceous). Also, a fence diagram of the study areas is shown in figure 2. Moreover, the distribution of major ions in the groundwater is as follows: $\text{Ca}^{2+} > \text{Mg}^{2+} > \text{Na}^{+} > \text{K}^{+}$ and $\text{HCO}_3^{-} > \text{SO}_4^{2-} > \text{Cl}^{-}$.

3. Methodology

3.1 Groundwater well inventory

Groundwater well inventory mapping as the first step in any spatial prediction modelling is necessary to identify the correlation of the distribution

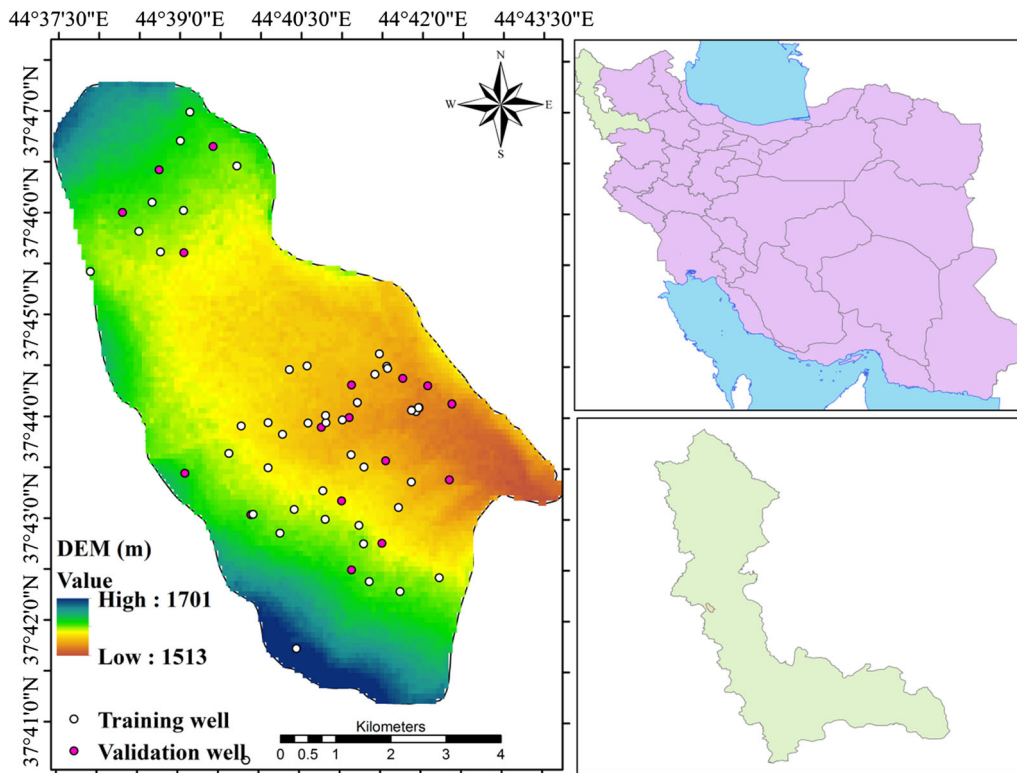


Figure 1. Well location map with the DEM of the Sero plain, Iran.

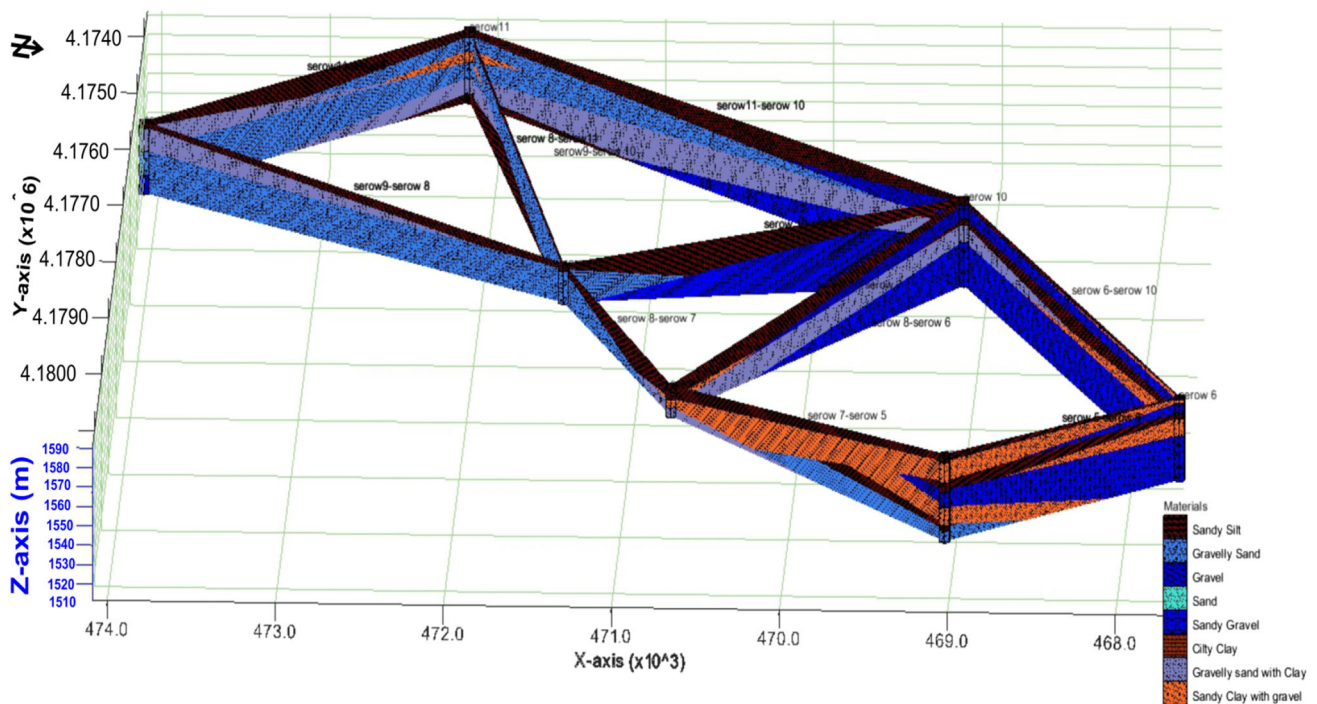


Figure 2. Fence diagram of the study area.

of groundwater well locations with the conditioning factors. Groundwater with higher potential was identified through the predictions made for the best groundwater potential (Nampak *et al.* 2014).

According to reports of UWRA and literature review, a threshold of 11 m³/hr was considered for groundwater yield (Nampak *et al.* 2014). The groundwater yield was identified according to the

actual pumping test. In the study area, only 75 well locations had groundwater yields $>11 \text{ m}^3/\text{h}$, which were randomly divided into two groups of sizes of 70% (53 well locations), as the training dataset used to develop the model, and 30% (22 well locations), as the testing dataset used for model validation (Nampak *et al.* 2014; Khosravi *et al.* 2016). An extensive field survey was carried out to identify the locations of wells with a GPS hand hole.

3.2 Groundwater conditioning factors

In the present research, 10 conditioning factors, namely, ground slope or slope degree, aspect, altitude, curvature, distance from the fault and the river, lithology, land use, rainfall and soil media, were taken into consideration. The conditioning factor selection was through reviewing the literature and using the data available (Mukherjee 1996; Oh *et al.* 2011; Ozdemir 2011a; Nampak *et al.* 2014).

At first, the digital elevation model (DEM) for the study area was downloaded from the ASTER Global DEM (spatial resolution of 30 m). Using DEM, four input variables, namely, slope degree, slope aspect, altitude and curvature maps, were fabricated. Slope degree had a high impact on groundwater occurrence. This factor was created and classified into five categories, namely, 0–3.1, 3.2–5.7, 5.8–8.7, 8.8–12.8 and 12.9–26.7, using quantile scheme classification (figure 3a). The slope aspects were classified after preparation into nine classes comprising the flat, north, northeast, east, southeast, south, southwest, west and the northwest (figure 3b). The third conditioning factor was altitude, which, after preparation, was divided into five classes, namely, 1491–1548, 1548–1573, 1573–1599, 1599–1633 and 1633–1697 m, using the quantile method (figure 3c). The curvature map shows the topography of the earth's surface and divides it into three classes, i.e., <-0.05 or concave, from -0.05 to 0.05 or flat, and >0.05 or convex (figure 3d) (Pham *et al.* 2017).

The fault in the study area was extracted from the fault of Iran, which was prepared using the Geological Survey of Iran (GSI) on a scale of 1:100,000. The factors of distance from the fault and the river were produced using the fault and the river of the study area in ArcGIS 10.2 software, and subsequently divided into five groups, namely, 0–100, 100–200, 200–500, 500–1000 and >1000 m (figure 3e and f) (Khosravi *et al.* 2018b).

Various lithologies have various infiltration rates; thus, lithology plays a crucial role in the identification of the groundwater potential occurrences (Pradhan 2009; Adiat *et al.* 2012; Nampak *et al.* 2014). This factor was prepared through the GSI. The formation ages of lithology were the Eocene, Middle-Eocene, Permian, Pre-Cambrian and the Quaternary (figure 3g). Land-use map was produced using supervised image classification techniques by Landsat 7 (ETM+), for which images were downloaded from the US Geological Survey (USGS). The images were divided into three classes, namely, irrigated land and agriculture, dry farming and moderate rangeland (figure 3h). The mean annual precipitation of four rain gauges in 20 yr was used to prepare the rainfall map which is a major conditioning factor in groundwater occurrences. The rainfall map was prepared using the Kriging method, due to its low RMSE, and divided into five classes, namely, 400–424, 425–446, 447–467, 468–488 and 489–510 mm (figure 3i). The characteristics of soil have a direct impact on water infiltration. Therefore, the soil map of the west Azerbaijan province with a scale of 1:50,000 was used in the analysis. The soil map was provided by UWRA, which was originally prepared by the Iranian Water Resources Department (IWRD). The dominant soil form in the area under study was inceptisols, followed by rock outcrop/inceptisols and rock outcrop/entisols (figure 3j).

3.3 Modelling of the groundwater potential assessment

Three models, namely, FR, SE and EBF, were applied for groundwater potential assessment and comparison at the Sero plain.

3.3.1 FR model

FR is one of the well-known and simple geospatial bivariate models for determining the probabilistic correlation between the dependent (groundwater) and independent (conditioning) factors. The FR method was previously utilised in various applications, such as flood susceptibility mapping (Tehrany *et al.* 2013; Khosravi *et al.* 2016, 2018b), groundwater qanat potential mapping (Pourtaghi and Pourghasemi 2014) and landslide assessment (Youssef *et al.* 2015). The most important benefit of the current model is its simple execution and easy-to-understand results (Gogu *et al.* 2001; Yalcin 2011). In the present study, the FR

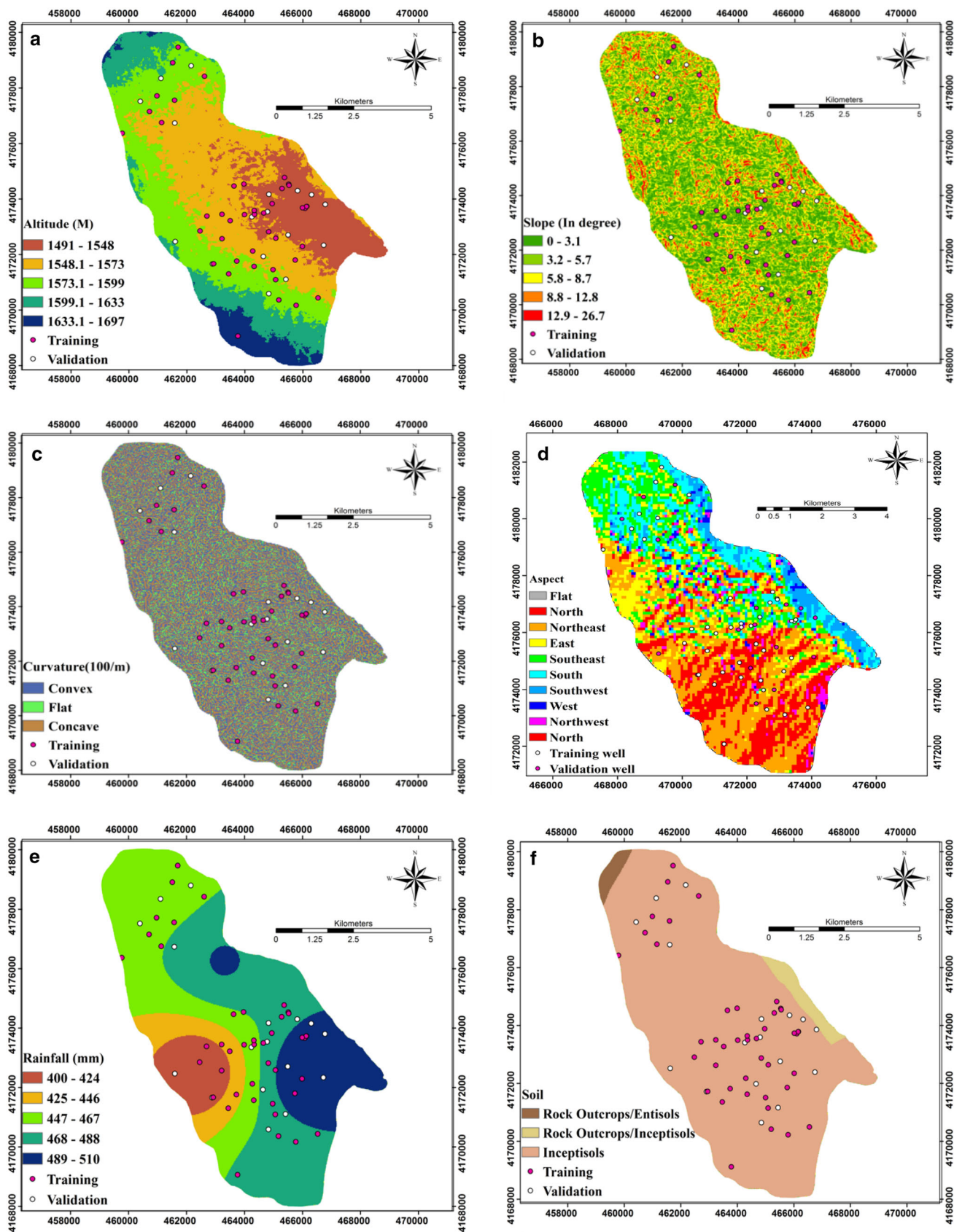


Figure 3. Conditioning factors of groundwater occurrences: (a) altitude, (b) slope angle, (c) curvature, (d) aspect, (e) rainfall, (f) soil, (g) land-use, (h) geology, (i) distance from the fault and (j) distance from the river.

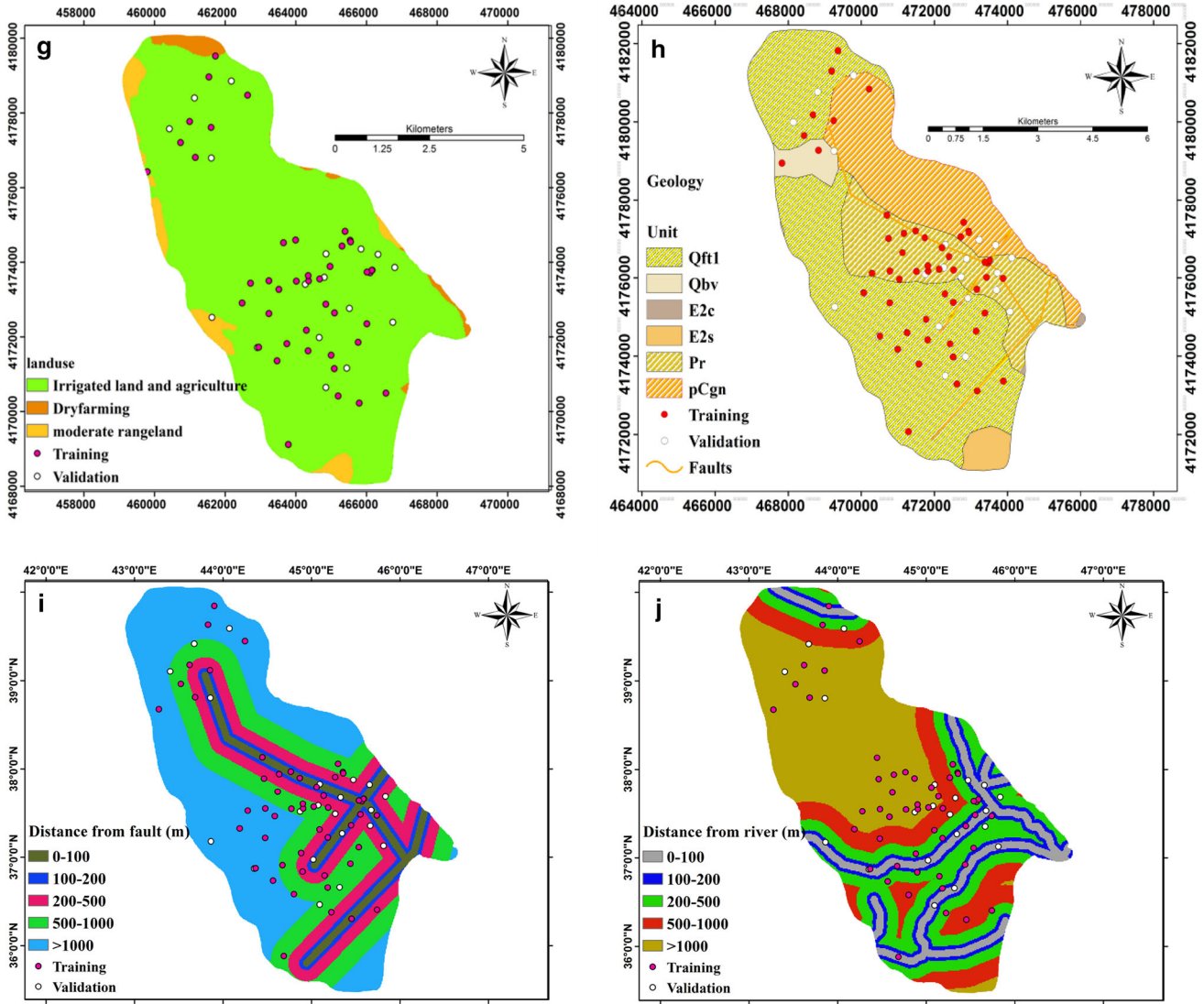


Figure 3. (Continued.)

model was applied to evaluate the impact of each conditioning factor used for groundwater occurrences (Khosravi *et al.* 2018b). Equations (4) and (5) were used to calculate the FR model and the groundwater well potential index (GWPI), respectively (Khosravi *et al.* 2018b):

$$FR = \frac{[N_{\text{pix}}(SX_i) / \sum_{i=1}^m SX_i]}{[N_{\text{pix}}(X_j) / \sum_{j=1}^n N_{\text{pix}}(X_j)]}, \quad (1)$$

$$FSI = \sum_{j=1}^n FR, \quad (2)$$

where $N_{\text{pix}}(SX_i)$ is the amount of pixels with groundwater well within the factor variable X of class i , $N_{\text{pix}}(X_j)$ denotes the amount of pixels within the factor variable X_j , m represents the

number of classes in the parameter variable X_i and n refers to the number of factors in the whole case study (Jaafari *et al.* 2014; Regmi *et al.* 2014).

3.3.2 Shannon entropy (SE) model

Overall, entropy shows the quantity of abnormality between events and results or decisions on various subjects under discussion (Wan 2009), and the entropy index is defined as the mean difference of the unit group to the whole system ratios (Theil 1972). SE, modified by the Boltzmann method, was used as the information theory (Pourghasemi *et al.* 2012). The SE model has been applied to the mapping of flood susceptibility (Khosravi *et al.* 2018a) and landslide susceptibility (Sharma *et al.* 2013) and has led to reasonable results. The following equations express the calculation of the weight of

the incorporated information. V_j is the value of the parameter from the total figure (Bednarik *et al.* 2010), which is determined based on the following equation:

$$E_{ij} = \frac{FR}{\sum_{j=1}^{M_j} FR}, \quad (3)$$

where FR represents the frequency ratio and E_{ij} denotes the probability density:

$$H_j = - \sum_{i=1}^{M_j} E_{ij} \log_2 E_{ij}, j = 1, \dots, n, \quad (4)$$

$$H_{j \max} = \log_2 M_j, M_j \text{--number of classes}, \quad (5)$$

$$I_j = (H_{j \max} - H_j / H_{j \max}), \\ I = (0, 1), j = 1, \dots, \quad (6)$$

$$V_j = I_j FR, \quad (7)$$

where H_j and $H_{j \max}$ denote the entropy values, I_j the information coefficient and M denotes the number of classes. Moreover, V_j represents the attained total weight value of the factor ranging between 0 and 1, where closer values to 1 indicate a higher disorder and imbalance (Khosravi *et al.* 2018a).

3.3.3 EBF model

The EBF model is developed based on the Dempster–Shafer theory of evidence. In the EBF model, evidential data estimation relates to a proposition (Shafer 1976; Dempster 2008). The EBF model consists of four parts, i.e., degrees of belief (Bel), disbelief (Dis), uncertainty (Unc) and plausibility (Pls), ranging between 0 and 1 (Carranza *et al.* 2005). The mapping of groundwater potential is performed based on the EBF model through the following equation (Shafer 1976; Dempster 2008):

$$\lambda(Tp) E_{ij} = N/D = [N(L \cap E_{ij}) / N(L)] \\ / [N(E_{ij}) - N(L \cap E_{ij})] \\ / (N(A) - N(L)), \quad (8)$$

$$\text{Bel} = \lambda(Tp) E_{ij} / \left(\sum \lambda(Tp) E_{ij} \right). \quad (9)$$

In this equation, $N(L \cap E_{ij})$ represents the number of groundwater well pixels in each class, $N(L)$ the total number of groundwater wells, $N(E_{ij})$ the number of pixels of each class, $N(A)$ the total number of pixels and N and D represent the

ratio of well location areas and the proportion of non-well areas, respectively. Similarly, Dis values are obtained by equations (10 and 11):

$$\lambda(T\bar{p}) E_{ij} = \frac{K}{H} \\ = [(N(L) - N(L \cap E_{ij})) / N(L)] \\ / [N(A) - N(L) - N(E_{ij}) \\ + N(L \cap E_{ij}) / (N(A) - N(L))], \quad (10)$$

$$\text{Dis} = \lambda(T\bar{p}) E_{ij} / \left(\sum \lambda(T\bar{p}) E_{ij} \right), \quad (11)$$

where K denotes the non-occurring well locations ratio and H the non-occurring well areas proportion in the attributes of the class.

Uncertainty and plausibility are calculated as follows:

$$\text{Unc} = (1 - \text{Bel} - \text{Dis}), \quad (12)$$

$$\text{Pls} = (1 - \text{Dis}). \quad (13)$$

4. Results

4.1 Results of the FR model

In the FR model, a higher frequency ratio shows the significant spatial relationship of conditioning factors with the occurrences of groundwater (Tehrany *et al.* 2013); the values >1 show a strong relationship and those smaller than 1 represent a weak relationship. This model was utilised to identify the relationship between each class of conditioning factors and groundwater occurrences and the results are presented in table 1.

The results indicate that in the study area, a lower altitude has a higher impact on groundwater occurrence. Therefore, the classes of 1491–1548 and 1548–1573 m with FRs of 2.2 and 1, respectively, have the highest impacts on the occurrences of groundwater. Moreover, the slopes of 0–3.1 and 8.6–12.7 with FRs of 1.49 and 1.45, respectively, have the highest impacts on groundwater occurrences. Thus, the results show that there is no specific pattern for the relationship between groundwater occurrences and the slope factor. With regard to curvature, in the area under study, the flat curvatures (2.59) have a strong impact on groundwater occurrence, which is in accordance with the nature of groundwater occurrences.

Table 1. Spatial relationship between well conditioning factors and well locations using the FR and SE models.

	No. of pixels in the domain	Percentage of domain	No. of wells	Percentage of floods	FR	E_{ij}	H_j	I_j	V_j
<i>Altitude (m)</i>									
1491–1548	13,164	18.30	22	41.51	2.27	0.52	1.71	0.27	0.23
1548–1573	24,492	34.05	18	33.96	1.00	0.23			
1573–1599	20,000	27.80	12	22.64	0.81	0.18			
1599–1633	10,080	14.01	0	0.00	0.00	0.00			
1633–1697	4196	5.83	1	1.89	0.32	0.07			
<i>Slope degree</i>									
0–3.1	19,992	27.79	22	41.51	1.49	0.35	1.88	0.19	0.16
3.1–5.7	24,805	34.48	16	30.19	0.88	0.20			
5.7–8.6	16,459	22.88	6	11.32	0.49	0.11			
8.6–12.7	8406	11.69	9	16.98	1.45	0.34			
12.7–26.6	2270	3.16	0	0.00	0.00	0.00			
<i>Curvature</i>									
Convex	31,125	43.27	17	32.08	0.74	0.18	1.31	0.17	0.23
Flat	9970	13.86	19	35.85	2.59	0.63			
Concave	30,837	42.87	17	32.08	0.75	0.18			
<i>Aspect</i>									
Flat	128	0.18	0	0.00	0.00	0.00	2.84	0.105357	0.09
North	10,609	14.75	15	28.30	1.92	0.24			
Northeast	10,931	15.20	2	3.77	0.25	0.03			
East	10,408	14.47	8	15.09	1.04	0.13			
Southeast	8901	12.37	6	11.32	0.91	0.11			
South	8577	11.92	6	11.32	0.95	0.12			
Southwest	7550	10.50	8	15.09	1.44	0.18			
West	7166	9.96	4	7.55	0.76	0.09			
Northwest	7662	10.65	4	7.55	0.71	0.09			
<i>Rainfall (mm)</i>									
400–423	5845	8.13	4	7.55	0.93	0.20	2.29	0.01	0.01
423–446	5977	8.32	4	7.55	0.91	0.19			
446–467	22,188	30.87	21	39.62	1.28	0.27			
467–487	25,923	36.06	18	33.96	0.94	0.20			
487–509	11,948	16.62	6	11.32	0.68	0.14			
<i>Soil</i>									
Rock outcrops/ entisols	1083	1.51	0	0.00	0.00	0.00	0.00	1.00	0.35
Rock outcrops/ inceptisols	2116	2.94	0	0.00	0.00	0.00			
Inceptisols	68,682	95.48	53	100.00	1.05	1.00			
<i>Land-use</i>									
Dry farming	1317	1.83	0	0.00	0.00	0.00	0.50	0.69	0.48
Moderate rangeland	2906	4.04	0	0.00	0.00	0.00			
Agriculture	67,658	94.13	53	100.00	1.06	0.50			
<i>Geology (age)</i>									
Quaternary	41,724	58.05	28	52.83	0.91	0.27	1.48	0.36	0.24
Eocene	1770	2.46	0	0.00	0.00	0.00			
Pre-Cambrian	14,871	20.69	8	15.09	0.73	0.22			
Middle Eocene	117	0.16	0	0.00	0.00	0.00			
Permian	13,399	18.64	17	32.08	1.72	0.51			
<i>Distance from the fault (m)</i>									
0–100	4611	6.41	11	20.75	3.24	0.44	2.06	0.11	0.17
100–200	4493	6.25	4	7.55	1.21	0.16			
200–500	12,910	17.96	13	24.53	1.37	0.19			

Table 1. (Continued.)

	No. of pixels in the domain	Percentage of domain	No. of wells	Percentage of floods	FR	E_{ij}	H_j	I_j	V_j
500–1000	18,262	25.41	14	26.42	1.04	0.14			
>1000	31,605	43.97	11	20.75	0.47	0.06			
<i>Distance from the river (m)</i>									
0–100	7369	10.25	13	24.53	2.39	0.41	2.13	0.08	0.09
100–200	6850	9.53	5	9.43	0.99	0.17			
200–500	16,216	22.56	14	26.42	1.17	0.20			
500–1000	13,254	18.44	6	11.32	0.61	0.10			
>1000	28,192	39.22	15	28.30	0.72	0.12			

Therefore, in a flat area, the runoff infiltrates more than other areas do. In the case of slope aspects, the north has a high influence on groundwater occurrences (1.9) as it receives sunlight less than other slope aspects do; it is followed by the southwest (1.44) and the east (1.04). The result of the relationship between rainfall and groundwater by the FR model shows that a rainfall amount of 446–467 mm has a high impact on groundwater occurrences (1.28). In the case of soil type, the results of the FR model show that inceptisols have high impacts on groundwater occurrences, as the total wells are located in the inceptisols. It is worth mentioning that the area is mostly covered by the aforementioned soil. For land-use, results show that agricultural land-use has a high impact on groundwater occurrences (1.06), as the whole wells are located in the agricultural land and this land-use covers about 94% of the study area (the major crops are both irrigated- and dry farming wheat alongside with a School and Community Assistance for Recycling and Composting Education garden). For the conditioning factor of geology (age), the Permian age has a high impact on groundwater occurrences (1.72), as 52% of the wells are located in the Quaternary age. However, a large part of the area does not have enough FR. With regard to the distance from the fault, the FR results show that only the classes of >1000 m do not affect groundwater occurrences and the distances between 0 and 1000 m have a high impact on groundwater occurrences; the first class, i.e., 0–100 m (3.24), has the most significant effect on the occurrences of groundwater and the more the distance from fault, the lower will be the impact on groundwater occurrences. The results of distance from the fault show that the first class, i.e., 0–100 m, has the highest effect on the occurrences of groundwater (2.39), followed by the class of 200–500 m (1.17).

4.2 Results of the SE model

The SE model has been applied to determine the correlation between each conditioning factor and groundwater well location (table 1). The weights of the conditioning factors of plan aspect, slope angle, altitude and curvature are equal to 0.09, 0.16, 0.23 and 0.23, respectively. This shows that altitude and curvature, among all topographical factors, have the highest impact, and the next are slope angle and slope aspect. The weight of rainfall, as a hydrological factor, is 0.01. The estimated weights of SE for other conditioning factors are as follows: soil (0.35), land-use (0.48), geology (0.24), distance from the fault (0.17) and distance from the river (0.09). Generally, the most and the least important factors for groundwater occurrences in the study area are altitude, curvature and rainfall, respectively.

4.3 Results of the EBF model

The correlation results for the well locations and each conditioning factor were achieved through the EBF model (Bel, Dis, Unc and Pls) and are presented in table 2. Nampak *et al.* (2014) stated that an important weakness of the EBF model was that if there was no value for Bel in a given class, there would be no well occurrence in the same class.

The spatial correlation between the location of the well and altitude with the EBF model demonstrates that the highest Bel values (0.515 and 0.226) belong to two classes of 1491–1548 m and 1458–1573 m, implying that the groundwater occurrence probability for these classes of altitude is higher than that for the other classes. The 1599–1633 m class has the lowest Bel (0). That is, there is no well location in this class of altitudes. In the case of slope degree, the slopes of 0–3.1 have the highest impact (0.346), followed by 8.6–12.7 (0.337).

Table 2. *Spatial relationship between well conditioning factors and well locations using the EBF model.*

	Percentage of domain	Percentage of wells	Bel	Dis	Unc	Pls
<i>Altitude (m)</i>						
1491–1548	41.51	18.30	0.515	0.143	0.341	0.857
1548–1573	33.96	34.05	0.226	0.201	0.573	0.799
1573–1599	22.64	27.80	0.185	0.215	0.601	0.785
1599–1633	0.00	14.01	0.000	0.233	0.767	0.767
1633–1697	1.89	5.83	0.073	0.209	0.718	0.791
<i>Slope degree</i>						
0–3.1	41.51	27.79	0.346	0.162	0.492	0.838
3.1–5.7	30.19	34.48	0.203	0.213	0.584	0.787
5.7–8.6	11.32	22.88	0.115	0.230	0.655	0.770
8.6–12.7	16.98	11.69	0.337	0.188	0.475	0.812
12.7–26.6	0.00	3.16	0.000	0.207	0.793	0.793
<i>Curvature</i>						
Convex	32.08	43.27	0.182	0.382	0.436	0.618
Flat	35.85	13.86	0.635	0.238	0.127	0.762
Concave	32.08	42.87	0.183	0.380	0.437	0.620
<i>Aspect</i>						
Flat	0.00	0.18	0.000	0.111	0.889	0.889
North	28.30	14.75	0.241	0.093	0.666	0.907
Northeast	3.77	15.20	0.031	0.126	0.843	0.874
East	15.09	14.47	0.131	0.110	0.759	0.890
Southeast	11.32	12.37	0.115	0.112	0.773	0.888
South	11.32	11.92	0.119	0.112	0.769	0.888
Southwest	15.09	10.50	0.180	0.105	0.714	0.895
West	7.55	9.96	0.095	0.114	0.791	0.886
Northwest	7.55	10.65	0.089	0.115	0.796	0.885
<i>Rainfall (mm)</i>						
400–423	7.55	8.13	0.196	0.202	0.603	0.798
423–446	7.55	8.32	0.191	0.202	0.607	0.798
446–467	39.62	30.87	0.271	0.175	0.554	0.825
467–487	33.96	36.06	0.199	0.207	0.595	0.793
487–509	11.32	16.62	0.144	0.213	0.643	0.787
<i>Soil</i>						
Rock outcrops/entisols	0.00	1.51	0.000	0.496	0.504	0.504
Rock outcrops/inceptisols	0.00	2.94	0.000	0.504	0.496	0.496
Inceptisols	100.00	95.48	1.000	0.000	0.000	1.000
<i>Land-use</i>						
Dry farming	0.00	1.83	0.000	0.494	0.506	0.506
Moderate rangeland	0.00	4.04	0.000	0.506	0.494	0.494
Agriculture	100.00	94.13	1.000	0.000	0.000	1.000
<i>Geology</i>						
Quaternary	52.83	58.05	0.271	0.222	0.507	0.778
Eocene	0.00	2.46	0.000	0.203	0.797	0.797
Pre-Cambrian	15.09	20.69	0.217	0.212	0.571	0.788
Middle Eocene	0.00	0.16	0.000	0.198	0.802	0.802
Permian	32.08	18.64	0.512	0.165	0.323	0.835
<i>Distance from fault (m)</i>						
0–100	20.75	6.41	0.442	0.164	0.393	0.836
100–200	7.55	6.25	0.165	0.191	0.644	0.809
200–500	24.53	17.96	0.186	0.178	0.635	0.822
500–1000	26.42	25.41	0.142	0.191	0.667	0.809
>1000	20.75	43.97	0.064	0.274	0.661	0.726

Table 2. (Continued.)

	Percentage of domain	Percentage of wells	Bel	Dis	Unc	Pls
<i>Distance from river (m)</i>						
0–100	24.53	10.25	0.407	0.166	0.427	0.834
100–200	9.43	9.53	0.168	0.198	0.634	0.802
200–500	26.42	22.56	0.199	0.188	0.613	0.812
500–1000	11.32	18.44	0.104	0.215	0.681	0.785
>1000	28.30	39.22	0.122	0.233	0.644	0.767

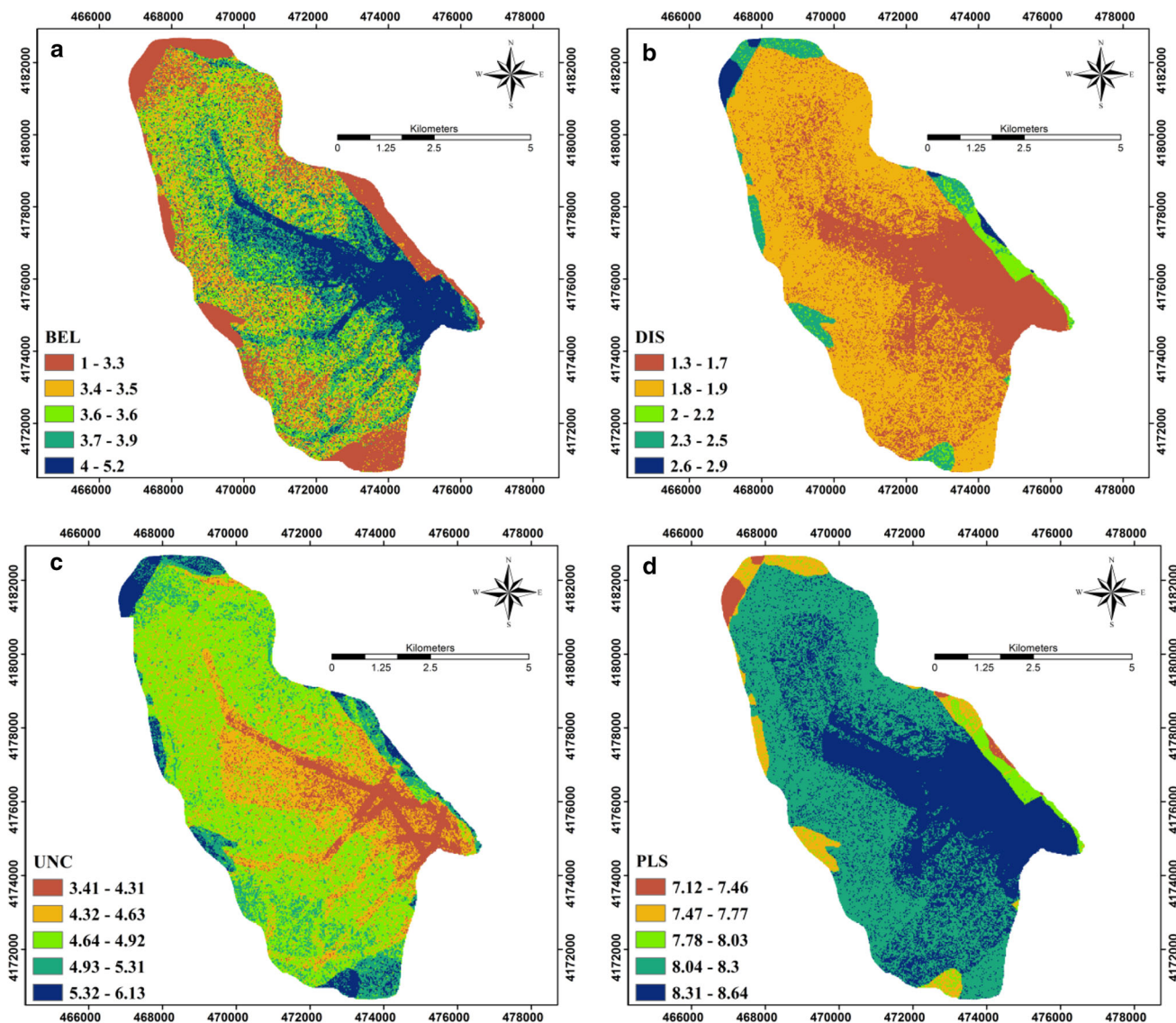


Figure 4. Integrated EBF model parameter: (a) Bel, (b) Dis, (c) Unc and (d) Pls.

In contrast, the last slope class, i.e., 12.7–26.6, has the lowest impact on groundwater occurrences (Bel = 0). Regarding curvature, the flat regions have higher Bel (0.635) than other curvatures. In terms of slope aspect, the north aspect has the highest value of Bel (0.241) and the lowest Bel

belongs to flat (0) areas. Results of the relationship between rainfall and groundwater well location show that the 446–467 mm class has the highest Bel (0.271). The highest (Bel) and lowest (Dis) probabilities of the occurrence of groundwater well are for the regions covered by inceptisols

with values of 1 and 0, respectively. The highest Bel (1) belongs to the areas with agricultural land-use, and the dry-farming and moderate rangelands have the lowest Bel (0). In terms of geological (age) conditions, the highest (0.512) and the lowest (0) values of Bel belong to the Permian, and the Eocene and middle Eocene, respectively. Considering the distance from the fault, the highest Bel value as well as the lowest Dis value is in the distance between 0 and 100 m. It reveals that the probability of groundwater well occurrences decreases with the increase in the distance from the fault. In terms of the distance from the river, the highest Bel belongs to the 0–100 m class, i.e., the lower the distance from the river, the higher will be the Bel value and the more the impact on groundwater occurrences.

The integrated results of the EBF model are depicted in figure 4(a–d). The Bel (figure 4a) and Dis (figure 4b) maps have been compared with each other. The results of the comparison show high Bel values in regions with low Dis and vice versa. Hence, these two maps are in contrast to each other. Overall, high groundwater well potential occurrences belong to the areas with high Bel degrees and low Dis degrees. Moreover, low Unc values belong to the areas with high Bel degrees. As a result, the areas with high potential of groundwater occurrence have a low uncertainty. The Pls map results show that the values of Pls are high in regions with high Bel values and low Unc values. The results of the current research are in accordance with those of Nampak *et al.* (2014) and Pourghasemi and Beheshtirad (2015). According to Althwaynee *et al.* (2014), the main benefit of the EBF mode is the modelling of the uncertainty degree (Dis) along with the predictive mapping of zones.

The calculated GPI indices were classified into five different classes of low, moderate, high and very high potential based on a natural break classification scheme (Zare *et al.* 2013). The results are shown in figure 5(a–c).

5. Validation of the achieved maps for groundwater potential

Validation of the achieved maps by different models is a very important phase in any modelling, without which the results of modelling lose their

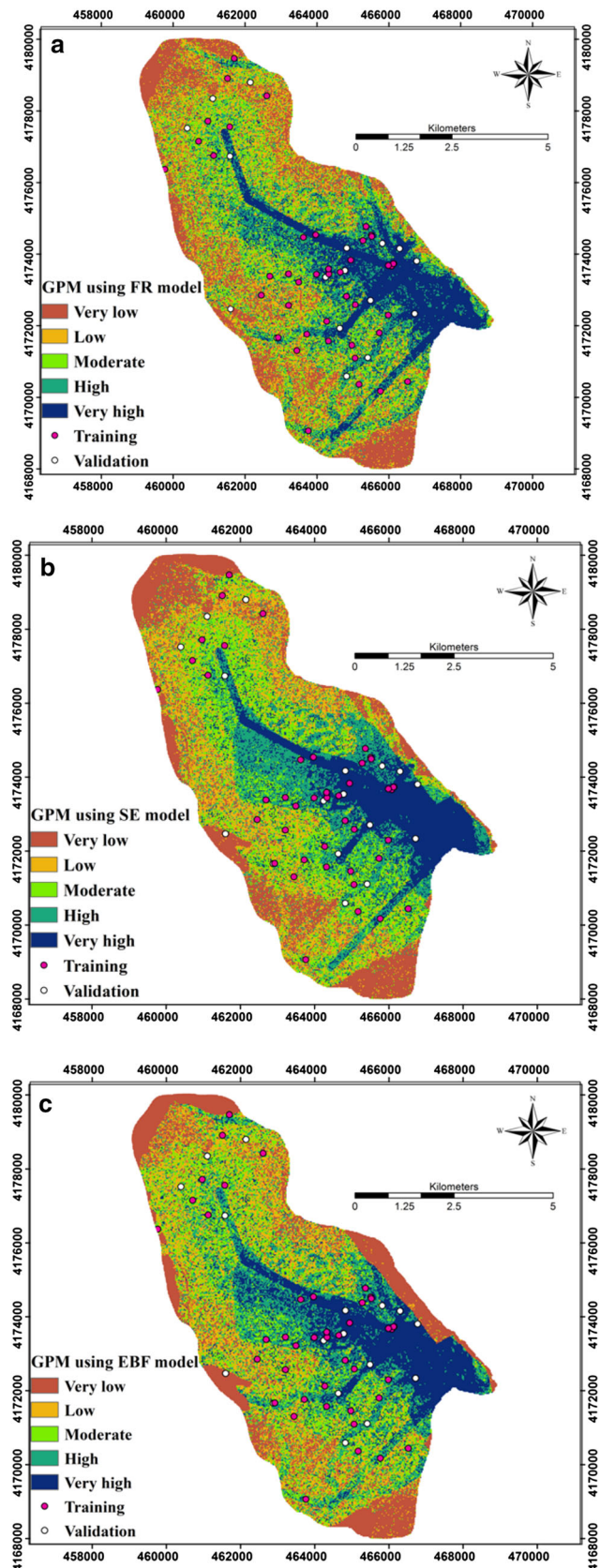


Figure 5. GPM developed by the (a) FR, (b) SE and (c) EBF models.

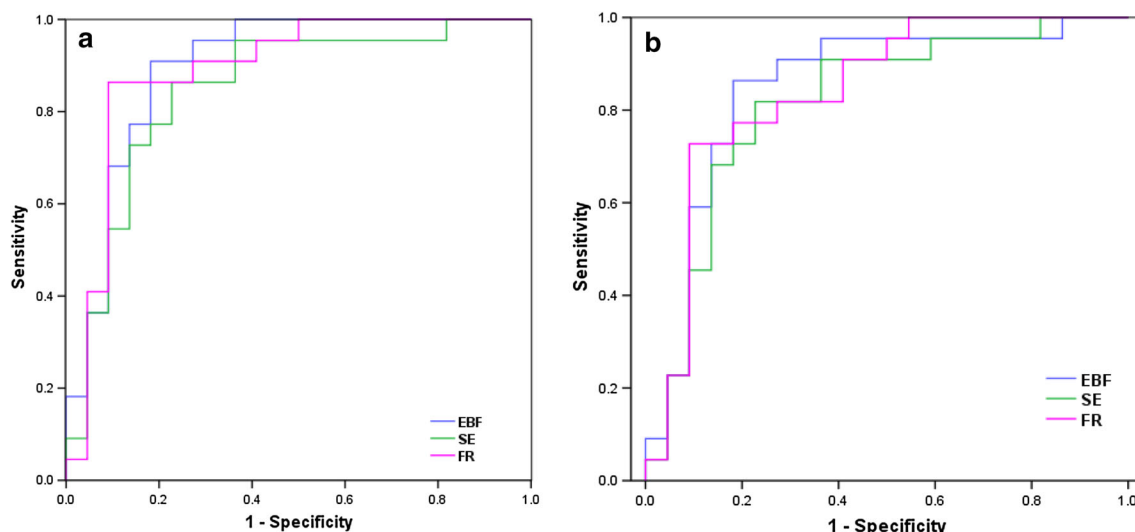


Figure 6. Validation of the GPM in the Sero plain: (a) the success rate curve and (b) the prediction rate curve.

scientific significance (Chung and Fabbri 2003). The training dataset was overlaid with final maps and, finally, the receiver operating characteristic curves (ROC curves) were plotted as the success-rate curves. The area under the ROC curve (AUC) was quantitatively calculated as the result of the prediction capability of the model in the SPSS 18 software (Akgun 2012; Mohammady *et al.* 2012; Pourghasemi *et al.* 2012). The ROC curve is a graphical plot demonstrating the diagnostic capability of a binary classifier system. In the ROC plot, X-axis and Y-axis are false positive rate and true positive rate, respectively. The range of AUC is between 0.5 and 1, and higher AUC values exhibit more prediction ability of the model. This range was classified as: 0.5–0.6, poor; 0.61–0.7, average; 0.71–0.8, good; 0.81–0.9, very good and finally, 0.9–1, excellent (Yesilnacar 2005). As the success rate curve was calculated with the training dataset, which was used for modelling, it could not be utilised as the result of model validation and it only showed how much the model built was useful in GPM. The prediction rate curve, which was attained using the final groundwater potential maps and testing dataset, revealed how efficient the built model was for GPM. Thus, its AUC was used as the model’s prediction capability. The AUC success rate values of FR, SE and EBF models are equal to 0.888, 0.847 and 0.897, respectively, which correspond to 88, 84 and 89% modelling accuracy rates (figure 6a and table 3). The AUC values of the prediction rate for the FR, SE and EBF models are 0.84, 0.81 and 0.85, which correspond to 0.84, 81 and 85% of the prediction accuracy rates (figure 6b and table 3).

Table 3. AUC results of the used models.

Models	AUC success rate	AUC prediction rate
FR	0.88	0.84
SE	0.84	0.81
EBF	0.89	0.85

Based on the results, the FR, SE and EBF bivariate models can be applied as a practical and simple tool to evaluate and manage groundwater. The advantages of the bivariate models are (i) simple implementation, (ii) reasonable accuracy in spatial prediction, and (iii) the ability to identify the factors or combinations of factors in the assessment (Van Westen *et al.* 2003).

6. Conclusion

As groundwater resources have an increasingly important role in water supplies all over the world, the evaluation of groundwater potential zonation has turned into a hot topic for those holding responsible positions in authority in governments, groundwater resources managers and private and land-use planners. In the present paper, three data-driven models, namely, FR, SE and EBF, were successfully applied to delineate the groundwater potential of the Sero plain in west Azerbaijan province, Iran, as one of the principal regions in the province for agricultural activities. In the first step, 75 well locations with groundwater yields >11 m³/h were identified and divided into two categories with the sizes of 70% (53 locations) to develop the model and 30% (22 locations) to

validate the model. Afterwards, 10 groundwater conditioning factors were selected for modelling. Finally, using the above-mentioned three models, three groundwater potential maps were produced. The results of model validation showed that the AU-success rate and AU-prediction rate curves of FR, SE and EBF were equal to 88, 84 and 89% and 0.84, 81 and 85%, respectively, and, according to ROC classification, their accuracy was very good. The results revealed that the EBF model was the best model for groundwater assessment of the Sero plain in both training and validation phases; the next models were FR and SE. One of the major benefits of EBF was that it allowed for (i) the predictive mapping of favourable zones and (ii) modelling of the uncertainty degree in prediction. Based on the results of the SE model, the most and the least important factors of groundwater occurrences in the study area were altitude and curvature and rainfall, respectively.

Overall, the achievements of the current study revealed that the models investigated could be applied successfully in spatial prediction modelling. The results of this study are helpful for related agencies in Iran, e.g., the RWAU and the Bureau of Soils and Water Management, and decision makers in comprehensive assessment of the development of groundwater exploration and environmental management in future planning. The bivariate models provided rapid, accurate and cost-effective results. In future work, the water quality of the case study, especially in areas with high groundwater occurrence potential, may be investigated and classified according to its usability in drinking and agriculture for better management of the groundwater resources.

Acknowledgement

We would like to thank all who helped us during the accomplishment of the study.

References

- Adiat K A, Nawawi M N and Abdullah K 2012 Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool – A case of predicting potential zones of sustainable groundwater resources; *J. Hydrol.* **440** 75–89.
- Akgun A 2012 A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: A case study at İzmir, Turkey; *Landslides* **9**(1) 93–106.
- Althuwaynee O F, Pradhan B, Park H J and Lee J H 2014 A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping; *Catena* **114** 21–36.
- Awawdeh M, Obeidat M, Al-Mohammad M, Al-Qudah K and Jaradat R 2014 Integrated GIS and remote sensing for mapping groundwater potentiality in the Tulul al Ashaqif, Northeast Jordan; *Arab. J. Geosci.* **7**(6) 2377–2392.
- Bastani M, Kholghi M and Rakhshandehroo G R 2010 Inverse modeling of variable-density groundwater flow in a semi-arid area in Iran using a genetic algorithm; *Hydrogeol. J.* **18**(5) 1191–1203.
- Bednarik M, Magulová B, Matys M and Marschalko M 2010 Landslide susceptibility assessment of the Kral'ovany–Liptovský Mikuláš railway case study; *Phys. Chem. Earth A/B/C* **35**(3–5) 162–171.
- Carranza E J, Woldai T and Chikambwe E M 2005 Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi district, Zambia; *Nat. Resour. Res.* **14**(1) 47–63.
- Chen W, Panahi M, Khosravi K, Pourghasemi H R, Rezaie F and Parvinnezhad D 2019 Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and biogeography-based optimization; *J. Hydrol.* **572** 435–448.
- Chowdhury A, Jha M K, Chowdary V M and Mal B C 2009 Integrated remote sensing and GIS-based approach for assessing groundwater potential in West Medinipur district, West Bengal, India; *Int. J. Remote Sens.* **30**(1) 231–250.
- Chung C J and Fabbri A G 2003 Validation of spatial prediction models for landslide hazard mapping; *Nat. Hazards* **30** 451–472.
- Corsini A, Cervi F and Ronchetti F 2009 Weight of evidence and artificial neural networks for potential groundwater spring mapping: An application to the Mt. Modino area (Northern Apennines, Italy); *Geomorphology* **111**(1–2) 79–87.
- Davoodi M D, Rezaei M, Pourghasemi H R, Pourtaghi Z S and Pradhan B 2015 Groundwater spring potential mapping using a bivariate statistical model and GIS in the Taleghan watershed, Iran; *Arab. J. Geosci.* **8**(2) 913–929.
- Deepika B, Avinash K and Jayappa K S 2013 Integration of hydrological factors and demarcation of groundwater prospect zones: Insights from remote sensing and GIS techniques; *Environ. Earth Sci.* **70**(3) 1319–1338.
- Dempster A P 2008 Upper and lower probabilities induced by a multivalued mapping; In: *Classic works of the Dempster-Shafer theory of belief functions*, Springer, Berlin, Heidelberg, pp. 57–72.
- Gogu R, Carabin G, Hallet V, Peters V and Dassargues A 2001 GIS-based hydrogeological databases and groundwater modelling; *Hydrogeol. J.* **9**(6) 555–569.
- Israil M, Al-Hadithi M and Singhal D C 2006 Application of a resistivity survey and geographical information system (GIS) analysis for hydrogeological zoning of a piedmont area, Himalayan foothill region, India; *Hydrogeol. J.* **14**(5) 753–759.

- Jaafari A, Najafi A, Pourghasemi H R, Rezaeian J and Sattarian A 2014 GIS-based frequency ratio and index of entropy models for landslide susceptibility assessment in the Caspian forest, northern Iran; *Int. J. Environ. Sci. Technol.* **11**(4) 909–926.
- Jasmin I and Mallikarjuna P 2011 Satellite-based remote sensing and geographic information systems and their application in the assessment of groundwater potential, with particular reference to India; *Hydrogeol. J.* **19**(4) 729–740.
- Jha M K, Chowdary V M and Chowdhury A 2010 Groundwater assessment in Salboni Block, West Bengal (India) using remote sensing, geographical information system and multi-criteria decision analysis techniques; *Hydrogeol. J.* **18**(7) 1713–1728.
- Kaliraj S, Chandrasekar N and Magesh N S 2014 Identification of potential groundwater recharge zones in Vaigai upper basin, Tamil Nadu, using GIS-based analytical hierarchical process (AHP) technique; *Arab. J. Geosci.* **7**(4) 1385–1401.
- Khosravi K, Nohani E, Maroufinia E and Pourghasemi H R 2016 A GIS-based flood susceptibility assessment and its mapping in Iran: A comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique; *Nat. Hazards* **83**(2) 947–987.
- Khosravi K, Panahi, M and Tien Bui D 2018a Spatial prediction of groundwater spring potential mapping based on adaptive neuro-fuzzy inference system and metaheuristic optimization; *Hydro. Earth Syst. Sci.* **22** 4771–4792.
- Khosravi K, Pham B T, Chapi K, Shirzadi A, Shahabi H, Revhaug I, Prakash I and Tien Bui T 2018b A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran; *Sci. Total Environ.* **627** 744–755.
- Le Page M, Berjamy B, Fakir Y, Bourgin F, Jarlan L, Abourida A, Benrhanem M, Jacob G, Huber M, Sghrer F and Simonneaux V 2012 An integrated DSS for groundwater management based on remote sensing: The case of a semi-arid aquifer in Morocco; *Water Resour. Manag.* **26**(11) 3209–3230.
- Lee S, Song K Y, Kim Y and Park I 2012 Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model; *Hydrogeol. J.* **20**(8) 1511–1527.
- Manap M A, Sulaiman W N, Ramli M F, Pradhan B and Surip N 2013 A knowledge-driven GIS modeling technique for groundwater potential mapping at the Upper Langat Basin, Malaysia; *Arab. J. Geosci.* **6**(5) 1621–1637.
- Manap M A, Nampak H, Pradhan B, Lee S, Sulaiman W N and Ramli M F 2014 Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS; *Arab. J. Geosci.* **7**(2) 711–724.
- Mehrdadi N 2010 Groundwater quality degradation of an aquifer in Iran central desert; *Desalination* **260**(13) 264–275.
- Mohammady M, Pourghasemi H R and Pradhan B 2012 Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster–Shafer, and weights-of-evidence models; *J. Asian Earth Sci.* **61** 221–236.
- Mukherjee S 1996 Targeting saline aquifer by remote sensing and geophysical methods in a part of Hamirpur–Kanpur, India; *Hydrogeol. J.* **19** 1853–1864.
- Naghbi S A, Pourghasemi H R, Pourtaghi Z S and Rezaei A 2015 Groundwater qanat potential mapping using frequency ratio and Shannon’s entropy models in the Moghan watershed, Iran; *Earth Sci. Inf.* **8**(1) 171–186.
- Nampak H, Pradhan B and Manap M A 2014 Application of GIS based data driven evidential belief function model to predict groundwater potential zonation; *J. Hydrol.* **513** 283–300.
- Nosrati K and Van Den Eeckhaut M 2012 Assessment of groundwater quality using multivariate statistical techniques in Hashtgerd Plain, Iran; *Environ. Monit. Assess.* **65**(1) 331–344.
- Oh H J, Kim Y S, Choi J K, Park E and Lee S 2011 GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea; *J. Hydrol.* **399**(3–4) 158–172.
- Ozdemir A 2011a GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison; *J. Hydrol.* **411**(3–4) 290–308.
- Ozdemir A 2011b Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey); *J. Hydrol.* **405**(1–2) 123–136.
- Pham B T, Khosravi K and Prakash I 2017 Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at Pauri Garhwal Area, Uttarakhand, India; *Environ. Process* **4**(3) 711–730.
- Pourghasemi H R and Beheshtirad M 2015 Assessment of a data-driven evidential belief function model and GIS for groundwater potential mapping in the Koohrang Watershed, Iran; *Geocarto Int.* **30**(6) 662–685.
- Pourghasemi H R, Mohammady M and Pradhan B 2012 Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran; *Catena* **97** 71–84.
- Pourtaghi Z S and Pourghasemi H R 2014 GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran; *Hydrogeol. J.* **22**(3) 643–662.
- Pradhan B 2009 Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques; *Open Geosci.* **1**(1) 120–129.
- Prasad R K, Mondal N C, Banerjee P, Nandakumar M V and Singh V S 2008 Deciphering potential groundwater zone in hard rock through the application of GIS; *Environ. Geol.* **55**(3) 467–475.
- Rahmati O, Samani A N, Mahdavi M, Pourghasemi H R and Zeinivand H 2015 Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS; *Arab. J. Geosci.* **8**(9) 7059–7071.
- Regmi A D, Devkota K C, Yoshida K, Pradhan B, Pourghasemi H R, Kumamoto T and Akgun A 2014 Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya; *Arab. J. Geosci.* **7**(2) 725–742.

- Roscoe M C 1990 *Handbook of groundwater development*; Wiley, New York.
- Shafer G A 1976 *Mathematical theory of evidence*; Princeton University Press, Princeton.
- Sharma L P, Patel N, Ghose M K and Debnath P 2013 Synergistic application of fuzzy logic and geo-informatics for landslide vulnerability zonation – A case study in sikkim Himalayas, India; *Appl. Geomatics* **5(4)** 271–284.
- Tehrany M S, Pradhan B and Jebur M N 2013 Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS; *J. Hydrol.* **504** 69–79.
- Theil H 1972 *Statistical decomposition analysis; with applications in the social and administrative sciences*; North-Holland Publ. Co, Amsterdam.
- Todd D K and Mays L W 1980 *Groundwater hydrology* (2nd edn); Wiley, Canada.
- Van Westen C J, Rengers N and Soeters R 2003 Use of geomorphological information in indirect landslide susceptibility assessment; *Nat. Hazards* **30(3)** 399–419.
- Vaux H 2011 Groundwater under stress: The importance of management; *Environ. Earth Sci.* **62(1)** 19–23.
- Wan S 2009 A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map; *Eng. Geol.* **108(3–4)** 237–251.
- Yalcin A A 2011 Geotechnical study on the landslides in the Trabzon Province, NE, Turkey; *Appl. Clay Sci.* **52(1–2)** 11–19.
- Yesilnacar E K 2005 *The application of computational intelligence to landslide susceptibility mapping in Turkey*; PhD thesis, Department of Geomatics, University of Melbourne.
- Youssef A M, Al-Kathery M and Pradhan B 2015 Landslide susceptibility mapping at Al-Hasher area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models; *Geosci. J.* **19(1)** 113–134.
- Zare M, Pourghasemi H R, Vafakhah M and Pradhan B 2013 Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: A comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms; *Arab. J. Geosci.* **6(8)** 2873–2888.

Corresponding editor: ARKOPROVO BISWAS