Environmental Earth Sciences (2018) 77:369
https://doi.org/10.1007/s12665-018-7551-y

ORIGINAL ARTICLE

# Use of a maximum entropy model to identify the key factors that influence groundwater availability on the Gonabad Plain, Iran

Ali Golkarian[1] · Omid Rahmati[2]

Received: 23 January 2018 / Accepted: 6 May 2018 / Published online: 15 May 2018



## Abstract

The purpose of this study is to identify the key factors that influence the availability of groundwater resources in a 10-year period in the Gonabad region of Iran using a maximum entropy model. For this purpose, 165 qanats were selected in 2004 that had yields of more than 3 l/s. By reviewing similar studies, 13 factors were considered to have an influence on groundwater potential exploited by the qanats, including: slope aspect; drainage density; fault density; distance from faults or other fractures; land use; lithology; plan curvature; profile curvature; qanat density; distance from rivers; land slope; stream power index (SPI); and topographic wetness index (TWI). The results indicated that qanat density had the greatest influence on the groundwater potential in a given area. Additionally, the increased importance of this factor with the occurrence of drought indicates that qanats are appropriate structures for exploiting groundwater and that they were well sited when they were originally constructed. Other important factors that influence groundwater potential are SPI, TWI and lithology factors. This indicates the importance of the slope, water accumulation area and rock characteristics in the groundwater potential. Additionally, there were significant changes in the distribution of areas with a high to low groundwater potential over a 10-year period. Reducing the number of effective factors in 2014 modeling showed that with the occurrence of drought and the limitation of areas with groundwater potential, the factors affecting zoning of the groundwater potential are also limited.


**Keywords** Drought · Maximum entropy · Qanat · Groundwater potential · Temporal changes · Spatial changes

## Introduction

Groundwater is one of the main and most important sources of water used by communities for various purposes (Pradhan 2009; Neshat et al. 2014) because of low temperature variations, better sustainability than surface water against short-term droughts, easy access and low utilization costs (Jha et al. 2007). There are limited spaces in aquifers for water storage, so it is necessary to exploit these resources with knowledge of groundwater conditions (Bera and Bandyopadhyay 2012). The infiltration of rainwater and snowmelts into the soil and/or pore space and discontinuous rocks causes the natural recharge of aquifer (Nampak et al. 2014). Accordingly, the distribution and quantity of groundwater can be influenced by climatic conditions as well as surface and subsurface characteristics of the earth such as alluvial properties, quantity and quality of fractures in subsurface rocks, land use, geomorphic facies and topographic features (Saud 2010; Senthil Kumar and Shankar 2014).

In arid zones, most human civilizations are located on the basis of ready access to water resources, especially to groundwater (Bera and Bandyopadhyay 2012). In these regions, groundwater resources are an important tool for sustainable development (Subyani 2005; Bayumi 2008). Therefore, understanding the factors influencing the availability of groundwater is important in managing the quality and quantity of these resources.

A large number of statistical models and learner machine methods can predict how to distribute groundwater potential using independent natural variables. These include: the support vector machine (SVM); the boosted regression tree (BRT); multivariate adaptive regression splines (MARS); logistic multiple regression (LMR); generalized additive model (GAM); and random forest (RF) models (Naghibi et al. 2016, 2017a; Mair and El-Kadi 2013; Sorichetta et al.

✉ Ali Golkarian
Golkarian@um.ac.ir

[1] Faculty of Natural Resources and Environment, Ferdowsi University of Mashhad, Mashhad, Iran

[2] Young Researchers and Elites Club, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran

2013; Rodriguez-Galiano et al. 2014; Naghibi and Pourghasemi 2015; Gutiérrez et al. 2009; Zabihi et al. 2016). One of the learning machine models is the maximum entropy method that operates on the basis of the statistical-probabilistic distribution of key factors. This model has been used in different fields of the natural sciences for assessing issues such as the forecasting of species distribution, crop planting zonation, landslide susceptibility evaluation, processing of earth features and groundwater characteristics mapping (Ariyanto 2015; Bajat et al. 2011; Convertino et al. 2013; Davis and Blesius 2015; Dyke and Kleidon 2010; Elith et al. 2011; Huset 2013; Jaime et al. 2015; Kim et al. 2015; Kornejady et al. 2017a; Kumar and Stohlgren 2009; Liu et al. 2012; Mert et al. 2016; Moosavi and Niazi 2016; Park 2015; Peterson 2011; Phillips et al. 2004; Pueyo et al. 2007; Rahmati et al. 2016b; Shen et al. 2015; Thuiller et al. 2005; Townsend Peterson et al. 2007; Wahyudi et al. 2012; Wang and Bras 2011; Warren and Seifert 2011; Williams 2010; Wollan et al. 2008; Yackulic et al. 2013; Yu et al. 2014).

Maximum entropy (MaxEnt) model does not define strict assumptions prior to research which can be considered as a powerful advantage. Importantly, it can also handle data from various measurement scales. Assuming that the potential of groundwater in the study area is affected by the occurrence of droughts, the following goals and innovations were considered in this study: (1) the use of the natural drainage properties of a qanat as a suitable index of the groundwater potential; (2) the assessment of spatial and temporal variations of groundwater potential using the maximum entropy model in one of the great plains of eastern Iran in a 10-year period; and (3) the determination of the most effective factors in reducing or/and sustainability the groundwater potential in the study area.

## Study area

The authors selected the Gonabad Region in the eastern part of Iran as a case study to assess temporal and spatial changes of groundwater potential. The Gonabad Region is located in the Razavi Khorasan Province, Iran, between 34°02′ to 34°26′N latitude and 58°16′ to 59°01′E longitude (Fig. 1). It covers an area of approximately 2188.91 km$^2$. The topographical elevation of the study area varies between 894 and 2774 m above sea level. The mean annual point precipitation is recorded as 160 mm in the Gonabad weather station (IRIMO 2016). Based on information provided by the Geological Survey of Iran (GSI 1997), about 44% of the lithology covering the study area falls within the units described as Qt2 including low level pediment fan and valley terraces deposit. Most of the area (89%) is covered by rangeland land use type. Exploitation of groundwater resources in this region includes semi-deep and deep wells, springs, and qanats. The Gonabad Region faces major challenges in the management of scarce freshwater resources under pressures of increasing population, economic development, climate change, pollution and overabstraction. Assessment of groundwater is very vital within this region, since groundwater supplies drinking water and irrigation requirements. Also, the people living in this region depend on irrigated and dry farming agriculture.

## Methodology

The steps in this research are presented in Fig. 2. In the first step, data preparation was carried out. Four types of data were utilized in this study: digital elevation model (DEM)-derived factors, distance factors, density factors and land factors. Then, the best factors were selected. In this research, several conditions were considered for selecting the factors that could be used in modeling: firstly, the lack of correlation performed through the PCA test; secondly, the contribution rate in modeling; thirdly, the impact of each factor on the training and testing process; and fourthly, the similarity of accuracy in the training and testing process. To evaluate time variations, the model was implemented two times and with a 10-year interval. In the next step, the model performance was evaluated using the receiver operating with a characteristic) ROC) index. Finally, after determining the most effective factors, the results were interpreted.

### Qanat inventory map

Usually, qanats areas are constructed on plains or on large waterways in mountainous areas where the groundwater table is high. In this study, the locations of qanats in 2004 were obtained from a topographical map with 1:50,000-scale and this was ground-truthed by the e field visits using a GPS device (Fig. 1). The profile of qanat system and its components can be found in Naghibi et al. (2015).

### Geo-environmental factors

After reviewing the results of studies on the potential of groundwater, 13 effective and key factors were determined and their respective maps were prepared in ArcGIS 10.3.1 software. The methods used to develop the relevant factors from these data are discussed below.

#### Slope aspect

Changes in the slope aspect lead to differences in the amount of solar energy that is received by the land surface, which in turn affects soil moisture content, the type, extent and distribution of the canopy cover, and ultimately

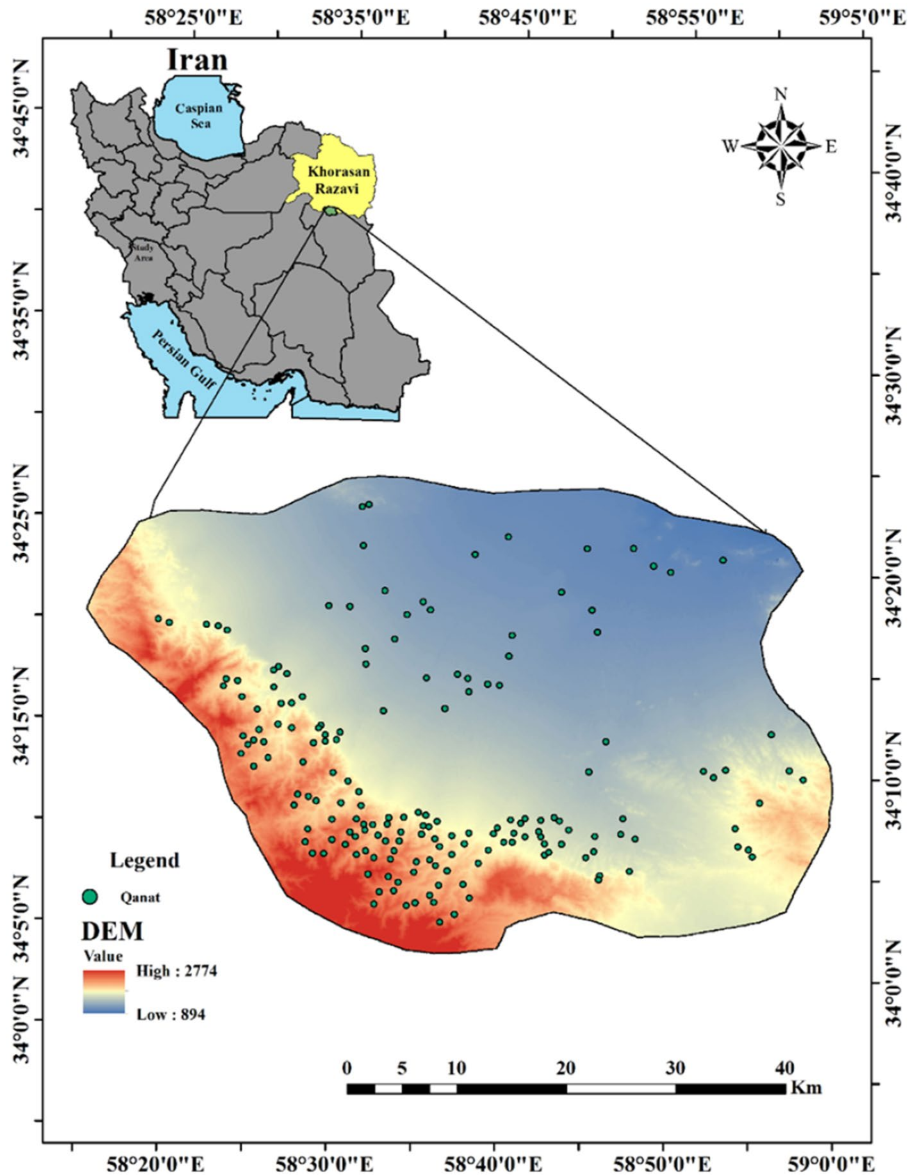



**Fig. 1** Location of Gonabad Region in Iran and Razavi Khorasan Province

evapotranspiration (Dai et al. 2001; Sidle and Ochiai 2006; Rahmati et al. 2016b; Naghibi et al. 2018). A slope aspect map was compiled from the DEM data (30 × 30 m) and has been classified into nine classes, four basic directions, four subsidiary directions and no direction or flat (Fig. 3j).

### Drainage density

In each region, the pattern and drainage density are closely related to the hydrological characteristics. Also factors such as geology, topography, soil type, infiltration, water absorption capacity in soil, canopy cover and climate affect the shape and density of drainage (Manap et al. 2013). High levels of drainage reduce the infiltration and increase the runoff. Therefore, areas with low drainage density are suitable for underground water development (Dinesh Kumar et al. 2007; Magesh et al. 2012).

The drainage network of the study area was extracted from topographic maps with a scale of 1:50,000 that were produced by Iran National Cartographic Center in 2013. Drainage density is the ratio of the total streams length to the area occupied by them (Singh and Prakash 2002). To calculate the drainage density, the total length of the stream per pixel (30 × 30 m) is divided into its area. Then the values of each pixel were assigned to its center. Finally, a drainage density map was developed using an interpolation technique. The drainage density calculated in the study area is in the range of 0–1.53 km/km$^2$ (Fig. 3e).

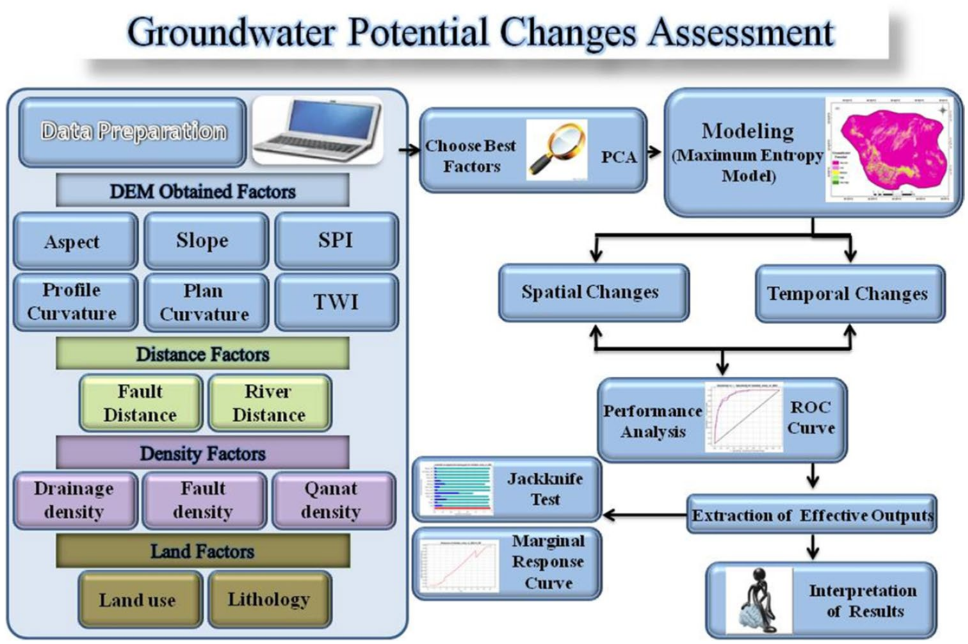



**Fig. 2** Flowchart study in this research

### Fault density

Lineaments such as faults and other deep fractures can cause surface and subsurface water exchange (Davoodi Moghaddam et al. 2013). The faults of the study area were extracted from geological maps with a scale of 1:100,000 that were produced by the Geological Survey Department of Iran (1997). The fault density map was prepared in the same way as the drainage density map and its range was between 0 and 18.64 km/km$^2$ (Fig. 3f).

### Fault distance

The map of distance to fault (m) was prepared based on the Euclidean distance algorithm using ArcGIS 10.3.1 software. The range of values for this factor is between 0 and 22,976 m (Fig. 3c).

### Land use

Land use can affect runoff, infiltration and the recharge of aquifers in different areas (Anbazhagan et al. 2005; Bhattacharya 2010; Naghibi et al. 2017b). Land use is the only factor that can change significantly during the 10-year period of study. Therefore, two land use maps for the years 2004 and 2014 were prepared. The land use map was prepared using Landsat 7 images and maximum likelihood algorithm via supervised classification (Rahmati et al. 2016a, b). In the studied area, five types of land use were identified including: agricultural land, fallow land, orchards, rangeland, and urban land. The largest and lowest land uses are rangeland and orchard, respectively (Fig. 3g).

### Lithology

Lithology influences various characteristics of an aquifer, especially its porosity and permeability (Chowdhury et al. 2010; Rahmati and Melesse 2016). These two characteristics affect the existence and movement of groundwater (Shahid et al. 2000; Ozdemir 2011). Usually, suitable aquifers have been reported in alluvial deposits of Quaternary formations (Rahmati et al. 2016b).

The lithology layer of the study area was extracted from geological maps with a scale of 1:100,000 that were produced by Geological Survey Department of Iran (1997). The lithology of investigation area consists of two sequences of sedimentary and volcanic rocks and unconsolidated deposits of Quaternary age (Fig. 3h).

### Plan curvature

The characteristics of landforms have been used as an effective factor in many geomorphological evaluation studies, such as the presence and distribution of groundwater, karst features, gullies, and landslides. The shape of the ground can alter surface drainage and sediment transfer to rivers (Jenness 2013). To prepare this layer, the digital elevation model (DEM) of the study area and ArcGIS 10.3.1 software were used (Fig. 3i).

### Profile curvature

The profile curvature can affect the flow velocity. Negative values (convex shape) increase the flow velocity and decrease infiltration, and positive values (concave shape)

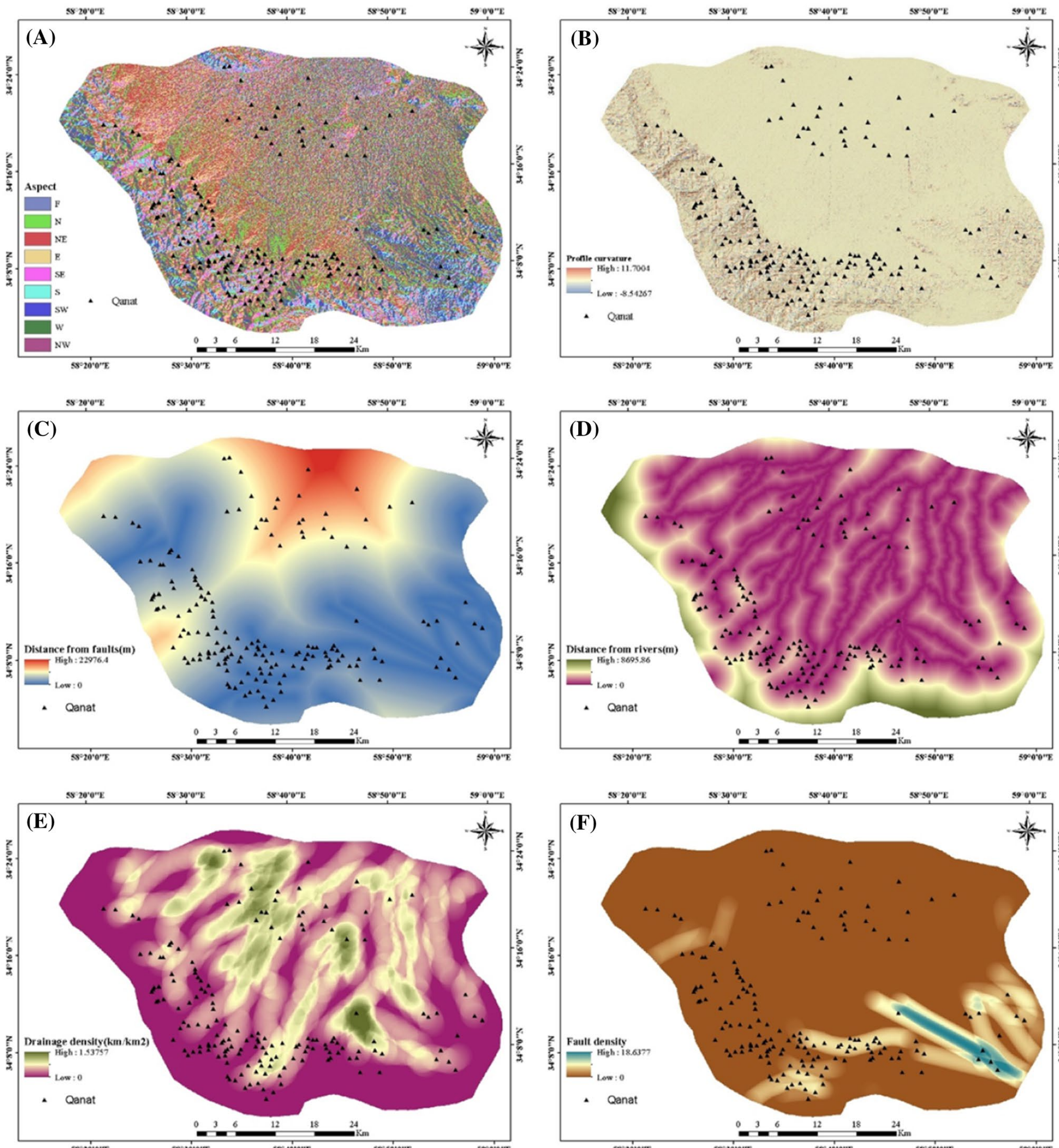



**Fig. 3** The map of factors used in the MaxEnt model. **a** Aspect, **b** profile curvature, **c** distance from faults, **d** distance from river, **e** drainage density, **f** fault density, **g** land use, **h** lithology, **i** plan curvature, **j** slope, **k** SPI, and **l** TWI

reduce the flow velocity and increase infiltration (Jenness 2013).

To prepare profile curvature like plan curvature, the digital elevation model of the study area and ArcGIS 10.3.1 software were used (Fig. 3b).

## Qanat density

Qanats act as natural groundwater drainage systems (Perrier and Salkini 1991). The existing qanats in the study area are very old and have been gradually built according to the

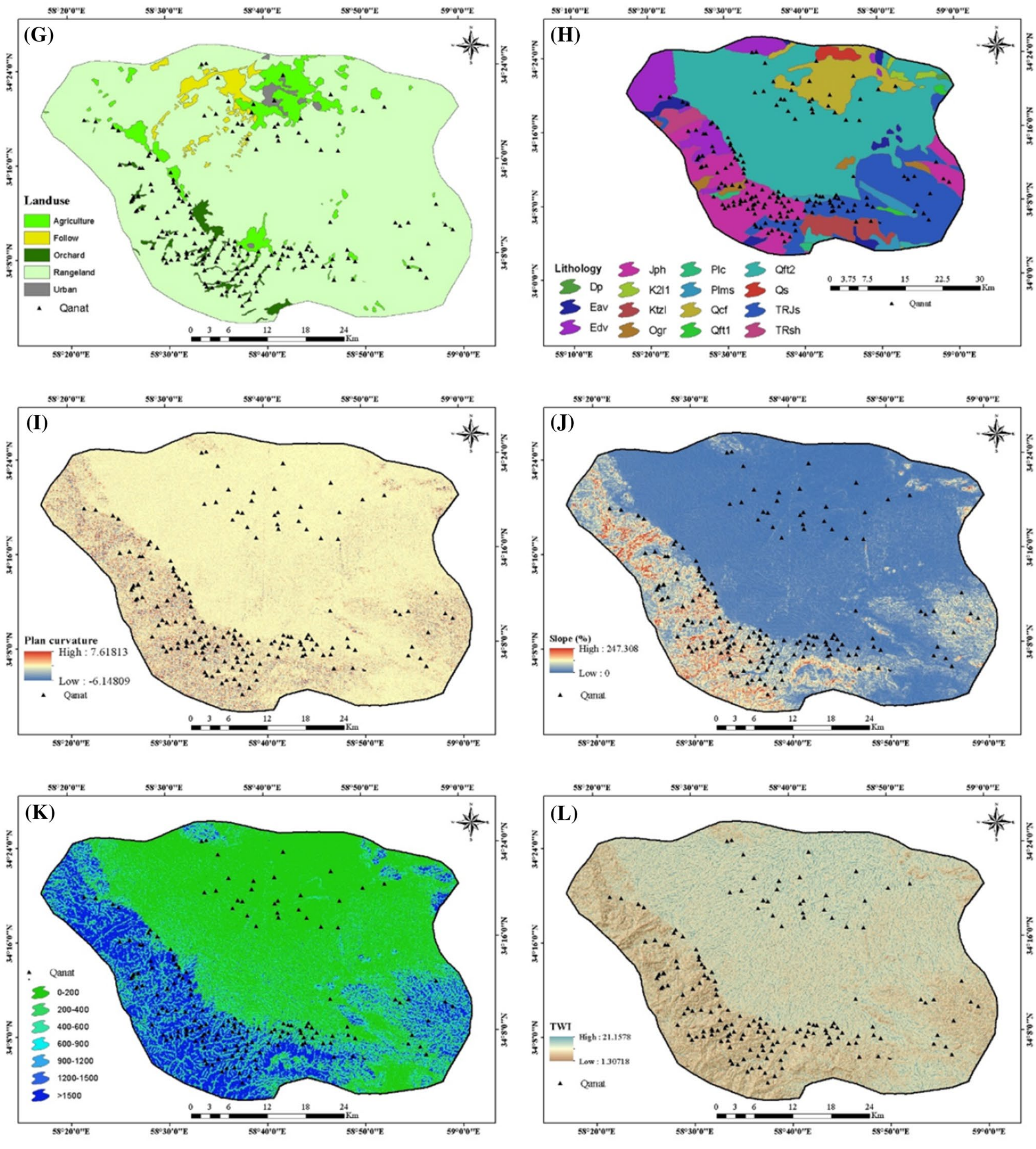



**Fig. 3** (continued)

potential of the aquifer and the need for water in the region. Therefore, an increasing density of qanats in each region indicates a suitable groundwater potential at the time of construction. The topographic maps of the study area with a scale of 1:50,000 that contain the district's qanat data were used to prepare the qanat density layer (Fig. 3m).

## Distance to river

Several studies have been conducted on the impact of river flows on groundwater status (Gooseff et al. 2005; Cardenas et al. 2004; Storey et al. 2003; Deepa et al. 2016). So increasing the distance from the rivers, especially rivers with permanent flows or rivers with a longer period of flow, can have a negative impact on groundwater potential. To provide this layer, topographic maps with a scale of 1:50,000 and Spatial Analyst extension in ArcGIS 10.3.1 were used (Fig. 3d).

#### Slope degree

Slope degree is an important factor in the study of groundwater conditions (Al Saud 2010; Ettazarini 2007; Kornejady et al. 2017b). This factor can affect the runoff, infiltration and recharge processes. As the slope increases, the runoff moves more quickly and this reduces the infiltration of water into the soil and reduces the recharge potential (Sarkar et al. 2001; Adiat et al. 2012). The DEM of the study area was used to prepare this layer (Fig. 3j).

#### Stream power index (SPI)

The stream power index is a measure of the erosive force of a stream (Moore et al. 1991). Increasing the SPI reduces the infiltration time, increases turbidity and reduces infiltration. Accordingly, this index was considered as an effective factor in determining the likelihood that recharge will take place to an aquifer from a flowing stream. The SPI was calculated using Eq. 1 (Moore et al. 1991) (Fig. 3k):

$$\mathrm{SPI} = B_{\mathrm{s}} \times \tan\alpha \tag{1}$$

where $B_{\mathrm{s}}$ is the specific catchment area ($\mathrm{m}^2$) and $\alpha$ is the local slope gradient (°).

#### Topographic wetness index

Topographic wetness index (TWI) quantifies the impact of topography on some hydrological characteristics. This index describes the process of collecting water at different places and the effect of gravity on the infiltration of water into an aquifer and to move downslope (Ghorbani Nejad et al. 2017). The index also determines the role of topography on the moisture content of soils in a given region and on the spatial distribution of soil moisture (Moore et al. 1991; Pourghasemi et al. 2013; Davoodi Moghaddam et al. 2013). The TWI index was calculated using Eq. 2 (Moore et al. 1991):

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

where $\alpha$ is the cumulative upslope area draining from a point (per unit contour length) and $\beta$ is the slope angle at the point (Fig. 3l).

### Combination of groundwater potential factors

The best combination of parameters used to assess the potential of groundwater was determined by checking the pairwise correlation between all factors using principal component analysis (PCA). The use of parameters that are well correlated with each other may cause errors in determining how prospective an area is for providing a groundwater resource. The optimal combination of factors is determined by eliminating inefficient factors. For this purpose, factors that are highly correlated with other factors are eliminated. The choice of these factors is based on the extent of correlation with other factors and by expert opinion. In this study, a correlation of 0.7 was considered as the maximum acceptable value for the correlation between factors (Kornejady et al. 2017a) (Table 1).

### Maximum entropy model

The theory of information and statistical principles are the basis of MaxEnt. The concept of entropy was first defined by Shannon (1948) as the expected amount (average) of the data contained in the conditions studied. This concept was presented based on Boltzmann's H-theorem as:

$$H(X) = E[I(X)] = E[-\ln(\hat{\pi}(x))], \tag{3}$$

where $E$ is the expected amount for a factor and $I$ is the data content. In this equation, the negative natural logarithm can be used to express the probability distribution of the factors. The logarithm with an increasing trend for independent sources is a suitable measure for entropy. According to the above equation, the average data were considered to be the expected amount. This is consistent with the probability theory that the mean of observational data is equal to the expected amount of a random variable (Ross 2014). Based on Jaynes' definition of maximum entropy, "the best probability distribution that can represent an unknown probability distribution function (pdf) subject to a set of testable information (expected values of different variables) is the one with the largest entropy" (Jaynes 1957a, b); if the results of a phenomenon are unknown, the information obtained will be greater and the entropy will be larger. Accordingly, when all restrictions are considered for the phenomenon under investigation, the best estimate of the pdf can be obtained, which can be accepted as the final result. In this case, the designated pdf has maximum entropy and the maximum information is obtained from the effective factors (Shannon 1948, 1951).

Like all other probability functions, two conditions for the entropy of $\hat{\pi}$ must be observed: (1) based on a statistical principle for all observable phenomena, assign a positive probability

**Table 1** Matrix of pairwise correlation between groundwater potential factors in the study area

| Factors | Aspect | Profile curvature | Fault distance | River distance | Drainage density | Fault density | Land use | Lithology | Plan curvature | Qanat density | Slope | SPI | TWI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aspect | 1.000 | 0.001 | −0.097 | 0.097 | −0.080 | 0.126 | 0.080 | −0.167 | 0.007 | −0.077 | 0.075 | 0.020 | −0.066 |
| Profile curvature | | 1.000 | −0.007 | 0.002 | −0.007 | 0.006 | −0.012 | −0.014 | −0.469 | −0.004 | 0.012 | 0.260 | 0.154 |
| Fault distance | | | 1.000 | −0.108 | 0.190 | −0.443 | −0.231 | 0.342 | −0.010 | 0.295 | −0.241 | −0.125 | 0.190 |
| River distance | | | | 1.000 | −0.699 | −0.063 | 0.113 | −0.449 | 0.009 | −0.256 | 0.376 | 0.134 | −0.224 |
| Drainage density | | | | | 1.000 | −0.013 | −0.123 | 0.464 | −0.013 | 0.297 | −0.413 | −0.152 | 0.243 |
| Fault density | | | | | | 1.000 | 0.055 | −0.363 | 0.008 | −0.115 | 0.048 | 0.044 | −0.062 |
| Land use | | | | | | | 1.000 | −0.141 | 0.010 | −0.047 | 0.141 | 0.036 | −0.121 |
| Lithology | | | | | | | | 1.000 | −0.022 | 0.282 | −0.412 | −0.174 | 0.261 |
| Plan curvature | | | | | | | | | 1.000 | −0.008 | 0.049 | −0.416 | −0.318 |
| Qanat density | | | | | | | | | | 1.000 | −0.186 | −0.084 | 0.127 |
| Slope | | | | | | | | | | | 1.000 | 0.288 | −0.483 |
| SPI | | | | | | | | | | | | 1.000 | 0.461 |
| TWI | | | | | | | | | | | | | 1.000 |

for each $x$; (2) the sum of all probabilities assigned to x must be one. By applying the two conditions above, Eq. 3 changes as:

$$H(X) = -\sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}(x). \tag{4}$$

In fact, in this study, there is a collection of random places $x$ from the larger set $X$, and appropriate qanat in each location is considered as the response variable ($Y$). In the spaces with these qanats, $y = 1$ and in the absence $y = 0$. MaxEnt is a generative model. This model performs learning through probability distribution $P(x, y)$ or conditional probability $P(x|y)$, while the discriminative models perform learning through the conditional probability $P(y|x)$. The implementation of discriminative models requires information about the presence and absence of phenomena, and the preparation of these data requires extensive studies) Vapnik and Vapnik 1998; Ng and Jordan 2001). Generative models can provide better predictions with little training information. This ability can be very useful in cases where accessible information is limited (Edvardsen et al. 2011; Edwards et al. 2005; Feeley and Silman 2011; Guisan et al. 2006; Niamir et al. 2011; Robertson et al. 2010).

If the existence and absence data are available, MaxEnt can be implemented as a deterministic model (Baldwin 2009; Berger et al. 1996; Halvorsen 2012). In this study, only presence data was used. Accordingly, the generative model is introduced based on Byes' rule (Elith et al. 2010; Phillips and Dudík 2008; Phillips et al. 2009):

$$P(y = 1|x) = \frac{P(x|y = 1)P(y = 1)}{P(x)}, \tag{5}$$

where $P(y = 1|x)$ is the probability for the existence of qanat on site $x$, $P(x|y = 1)$ is the probability of existence in the site $x$ given that the qanat is present (equivalent to $\pi(x)$ in MaxEnt), $P(y = 1)$ is the complete occurrence, and $P(x)$ is the probability that the $x$ site will be selected ($P(x)$ is equivalent to $1/|X|$, that is, the probability of selecting the $x$ site from the $X$-set).

Another form of the above equation is:

$$P(y = 1|x) = \pi(x)P(y = 1)|X|. \tag{6}$$

According to the probability distribution in generative models, $P(x)$ can be rewritten as:

$$P(x) = \sum_{y} P(x|y) = P(x|y = 1)P(y = 1) + P(x|y = 0)P(y = 0). \tag{7}$$

If the probability of presence and absence is equal $(P(y = 0) = P(y = 1) = 0.5)$, the equation can be simplified as:

$$P(y = 1|x) = \frac{P(x|y = 1)}{P(x|y = 1) + P(x|y = 0)}. \tag{8}$$

In this study, the potential of groundwater in the study area was determined using the maximum entropy model in the MaxEnt software 3.3.3 k. MaxEnt is a machine learning model based on the presence of a phenomenon (Graham et al. 2008; Hernandez et al. 2008; Kornejady et al. 2017a; Phillips et al. 2006; Quinn et al. 2013; Wisz et al. 2008). One of the applications of these models is to study areas with inappropriate access.

Of course, it should be noted that if the distribution of input data is not appropriate, modeling only based on the phenomenon presence will be prone to error (Austin 2007; Hortal et al. 2008; Loiselle et al. 2008; McCarthy et al. 2011; Reddy and Dávalos 2003; Robertson et al. 2010; Veloz 2009; Wolmarans et al. 2010). In the present study to resolve this problem, all the qanats in the study area were investigated. MaxEnt is able to use the data in a continuous and categorized manner. This will make the results more accurate and reduce computational errors.

This model estimates the best probabilistic distribution of potential areas by connecting input factors and existing potential areas (Elith et al. 2010; Kleidon et al. 2010; Medley 2010; Moreno et al. 2011; Nieves et al. 2011). Initially, MaxEnt prepares a first map. This map has the same pixels in terms of potential groundwater probability. Then, by adding each of the influential factors, the characteristics and limitations of the susceptible areas are extracted and the accuracy of the first map is improved. This operation continues until the best probability distribution function (using the maximum entropy function) and groundwater potential map are obtained.

### Model validation

To validate the MaxEnt model, a success rate curve (SRC) and a prediction rate curve (PRC) should be plotted. These curves are plotted using both the training and testing datasets. In both curves, the vertical axis is related to the correct detection rate of places with groundwater potential and the horizontal axis is related to the correct detection rate of places without groundwater potential. The area under the SRC (AUSRC) represents the accuracy and the area under the PRC (AUPRC) represents the prediction power or generalization of the model. When this value is closer to one, the accuracy of the model increases (Metz 1978; Pearce and Ferrier 2000; Pontius and Schneider 2001).

Determining the most important factors affecting the potential of groundwater can be useful in managing water resources. In the MaxEnt model, the jackknife test is used to determine the most important factors. In addition, the different classes and ranges of each factor have a different effect on the groundwater potential. Determining the most important range of each factor was determined using the response curve.

### Data format and the determination of parameters in MaxEnt

The map of all factors studied was converted to ASCII format and the coordinates of the used qanats were prepared in the form of a CSV file; then, both sets of data were entered into the model. The background points and convergence thresholds were selected as 10,000 and $10^{-5}$, respectively. To better learn the proper pattern and ensure the effectiveness of the plateau, the number of replicates was increased to 5000. Since the type of features and the number of samples are different, selecting the “auto” option allows the model to use the appropriate adjustment values (Phillips et al. 2006).

## Results and discussion

### Initial selection of appropriate factors

According to the results of the PCA test (Table 1), the pairwise correlation between the drainage density and the river distance has a maximum value ($-0.699$) and a pairwise correlation between the profile curvature and the slope aspect has a minimum value (0.001). Also, the results showed that the pairwise correlation of all factors was less than 0.7, so none of the considered factors were eliminated and all were entered into the model.

### Running the model at two different times

As stated above, in the present study, for the purpose of examining time variation, the potential of groundwater in the 2 years of 2004 and 2014 was investigated. With the initial implementation of the model, it became clear that aspect, plan curvature and profile curvature in 2004 and aspect, drainage density, fault distance, fault distance, plan curvature and land use in 2014 had no effect on the accuracy of the model and, even with their elimination, the accuracy of the model increased in the implementation of 2014. To compare the results of groundwater potential in the two studied times, groundwater potential maps of the study area were classified according to natural break distance into five classes as very low, low, medium, high and very high (Pourghasemi and Rossi 2017; Pourghasemi and Kerle 2016; Rahmati et al. 2016a) (Fig. 4). The density of suitable qanats, the percentage of different groundwater potential classes, the kappa index and the Chi square test results for two-time implementation are provided in Table 2. In all cases, the largest area is dedicated to classes with the least groundwater potential. By improving the groundwater potential (increase class), the area decreases. Accordingly, the percentages of different groundwater potential classes in both years 2004 and 2014 have statistically significant

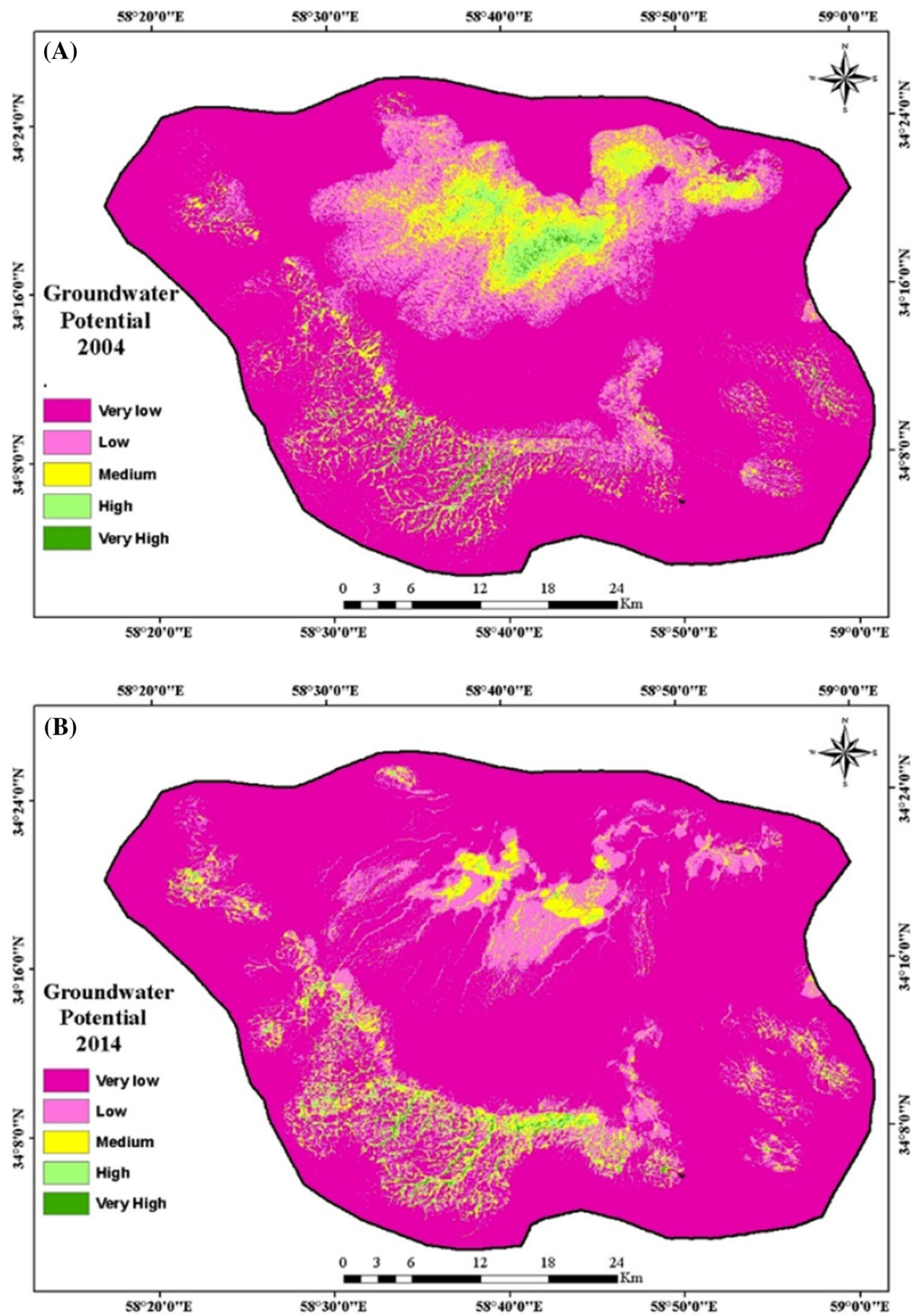



**Fig. 4** Groundwater potential maps: **a** 2004, **b** 2014

**Table 2** Areal percentage of models' classes, qanat densities, discordance and concordance (2004–2014)

| Class | 2004 | | | 2014 | | | 2004–2014 | |
|---|---|---|---|---|---|---|---|---|
| | Class area (%) | Chi square | Qanat density | Class area (%) | Chi square | Qanat density | Kappa index | Level of agreement |
| Very low | 65.73 | 139.59 | 0.0000094 | 76.81 | 207.95 | 0.0000059 | 0.77 | Strong |
| Low | 18.04 | $P$ value $<0.05$ | 0.0000870 | 14.87 | $P$ value $<0.05$ | 0.0000300 | 0.19 | Very weak |
| Medium | 11.17 | | 0.0002000 | 5.72 | | 0.0000720 | 0.13 | Very weak |
| High | 4.55 | | 0.0004200 | 2.15 | | 0.0001500 | 0.12 | Very weak |
| Very high | 0.50 | | 0.0011000 | 0.45 | | 0.0006000 | 0.14 | Very weak |

disagreement at 95% confidence level ($P$ value $b$ 0.05). This result is common in natural processes and shows the role of different natural conditions in the existence or absence of the phenomenon studied. The Kappa test results show that the groundwater potential in 2004 and 2014 have a strong agreement in very low classes and very weak agreement in low, medium, high and very high classes. The results of Table 2 indicate a decrease in the area percentage of all classes except the very low class in 2014. This result indicates that there has been a reduction in groundwater potential in the interval between the two studies. Investigating the density of appropriate qanats showed that in classes with a higher groundwater potential, the density of qanats increased. This result indicates that the model is functioning correctly.

The results obtained indicate that the MaxEnt model has a high level of accuracy (Fig. 5) for determining the groundwater potential of the study area in both 2004 and 2014. Despite the decrease in the number of suitable qanats in 2014, implementing the model at the two mentioned times, with AUC of about 0.9 for both training and testing, represents the excellent ability of the model to assess and predict the groundwater potential (Pearce and Ferrier 2000).

## Most effective factors

The relative contribution of all factors studied in the training process of the MaxEnt model in both 2004 and 2014 is presented in Table 3. The results show that in both 2004 and

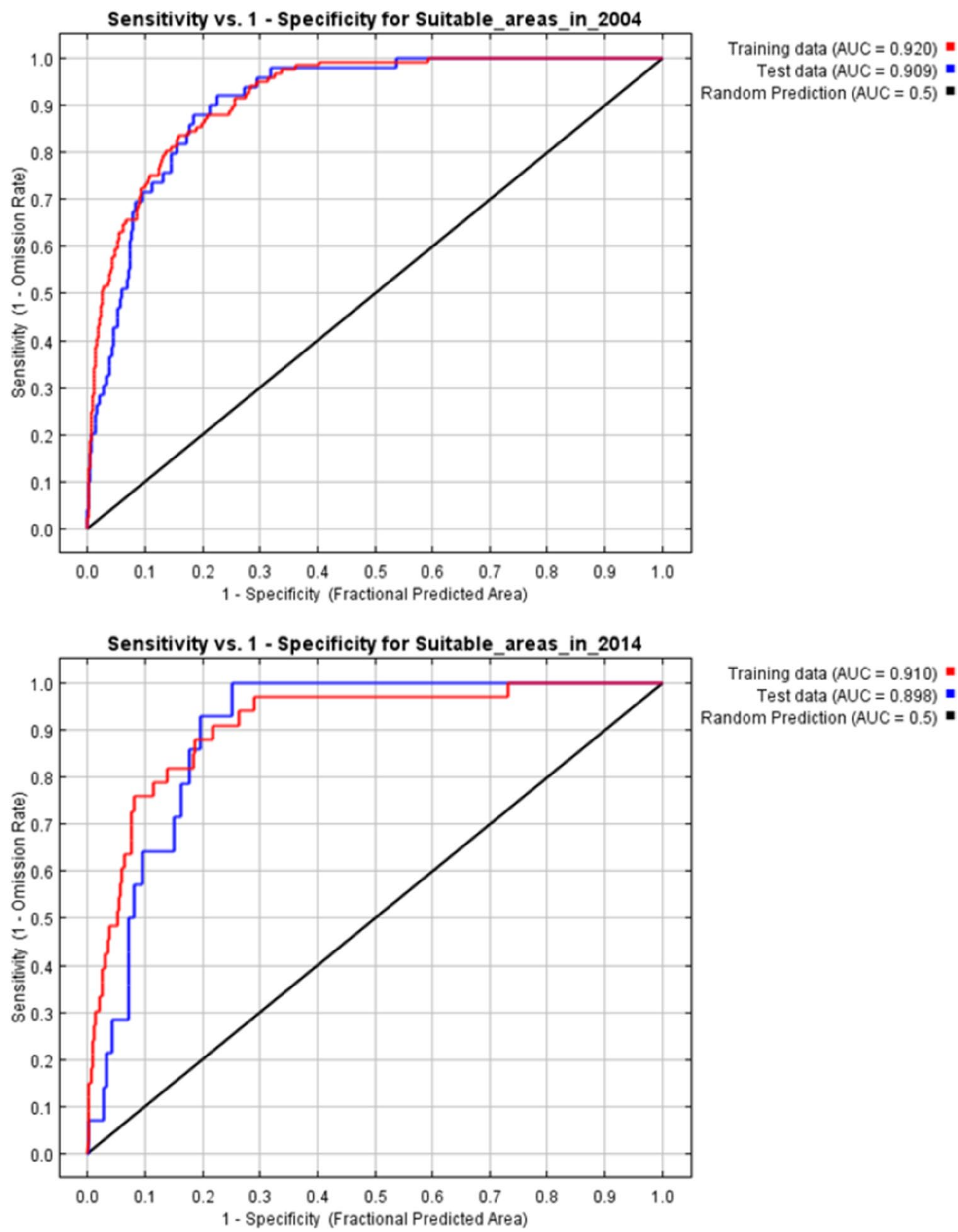



**Fig. 5** Accuracy of the model in training and validation steps in **a** 2004 and **b** 2014

**Table 3** Contribution rate of the various factors on groundwater potential modeling in the MaxEnt model

| Factor | Qanat density | SPI | Lithology | Land use | Slope | Drainage density | TWI | River distance | Fault Density | Fault distance | Profile curvature |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *2004* | | | | | | | | | | |
| Contribution (%) | 47 | 17.7 | 7.7 | 3.5 | 7.1 | 3.5 | 7.1 | 3.8 | 1.7 | 0.9 | – |
| | *2014* | | | | | | | | | | |
| Contribution (%) | 63 | 12.8 | 5.7 | – | 5.6 | – | 9.2 | 1.8 | – | – | 1.8 |

2014, the relative contribution of qanat densities is higher than other factors and is about 50% and more than the contribution of various factors in modeling dependent on this factor. The results also indicate a relatively large change in the contribution of the investigated factors over time. In 2014, due to the reduction of groundwater potential in different parts of the study area, the contribution of qanat density, with a 16% increase, has the highest increase, and the contributions of SPI index, lithology and river distance with 4.9, 2.0 and 2.0% decrease, respectively, have the highest reduction. As there are a number of algorithms in the MaxEnt model, it is difficult to determine the specific contribution of a given factor. These algorithms can have different results in determining the contribution of factors (Phillips 2012). Accordingly, the assigned contribution, rather than reflecting the importance of factor, relates to the role of each factor in the modeling process.

The jackknife test using test data for 2004 and 2014 was implemented and its results are presented in Fig. 6. The graph obtained from the jackknife test includes three types of information: (1) model AUPRC when each factor enters the model alone (blue part); (2) the AUPRC of the model when the model runs with all the factors other than the one mentioned at the beginning of the bar (in other words, the accuracy of the model with the removal of each factor and the presence of the remaining factors—the green part); and

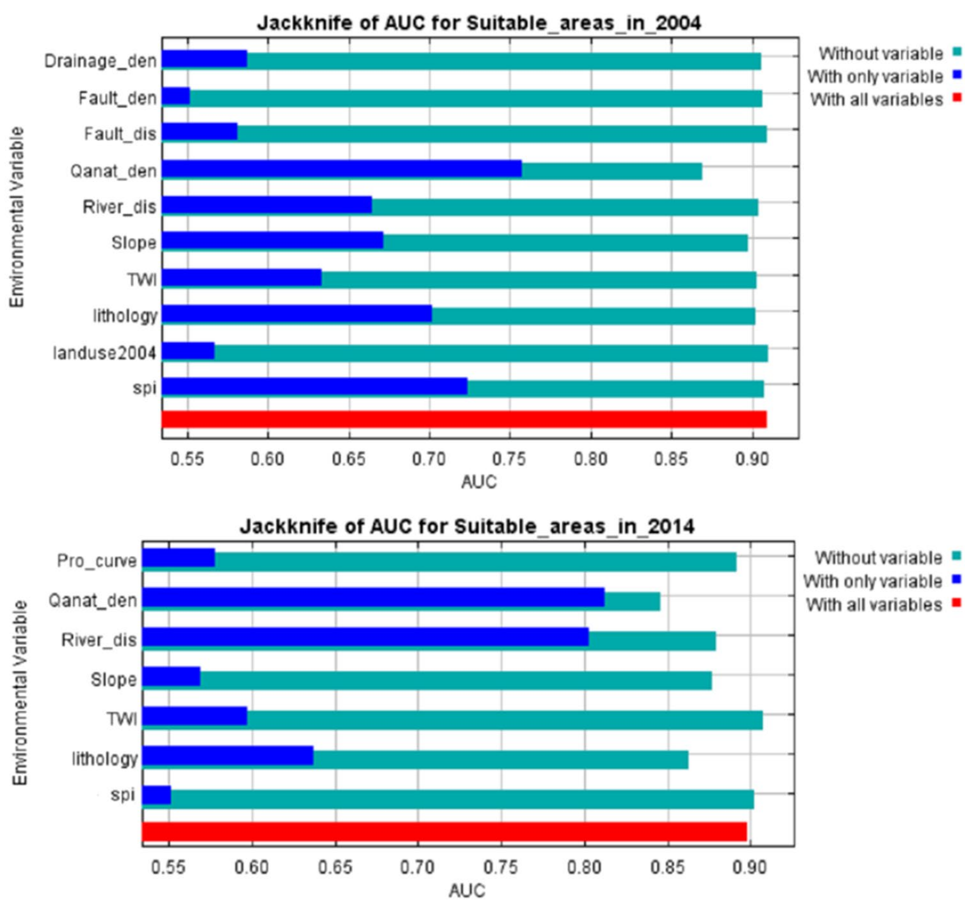



**Fig. 6** Jackknife of AUC for MaxEnt model using test data. **a** 2004, **b** 2014

(3) model AUPRC when all factors are involved in modeling (red part).

The implementation of the model with each of the factors independently showed that the qanat density and the SPI index in 2004 and the qanat density and the river distance in 2014 with the highest AUPRC had the greatest impact on the potential for groundwater availability. The results of the implementation of the progressive removal of each factor (green bar) indicate that the highest decrease in AUPRC was obtained by eliminating qanat density from the model. Therefore, the sensitivity of the model to the changes in this factor is high. The uniqueness of the impact this factor has had on other factors cannot compensate for the lack of it and the accuracy of the model is reduced. The sensitivity of the model to the SPI and river distance factors is not high and the lack of these two factors is compensable by other factors. These results are consistent with the relative contribution of the factors.

The results also showed that the drainage density, distance and density of faults and the land uses with the lowest AUPRC had little effect on the modeling of 2004, and was eliminated due to the ineffectiveness of the 2014 modeling process. A review of the data show that the limited availability of information of the number and length of faults, the predominance of plain lands and the low diversity of the land use is likely to have influenced the low impact of these factors on the model output.

The factors can be eliminated from the modeling process when they have no effect on the accuracy of the model (Kornejady et al. 2017a). With this argument, the factors of slope aspect and the plan curvature in 2004 and the slope aspect, drainage density, fault density, fault distance, plan curvature and land use in 2014 were eliminated in the modeling process. The results show that the reduction in the number of factors affecting the modeling process in 2014 indicates that with the occurrence of drought, many factors lose their role in the groundwater potential and determining the optimal areas for the potential of groundwater is limited to several key factors. However, in 2004, after qanat density and the SPI index factors which have a high impact on model output, lithology, TWI, slope and river distance factors had a moderate impact.

## The response of different factors to the groundwater potential

By increasing the qanat density in both 2004 and 2014, its effect on the groundwater potential increases. Increasing the density of the qanats increases the volume of water extracted from aquifers. This parameter should reduce the groundwater potential, but increasing the groundwater potential by increasing the qanat density indicates the proper identification of potential areas and a proper site selection when constructing these structures (Figs. 7, 8a).

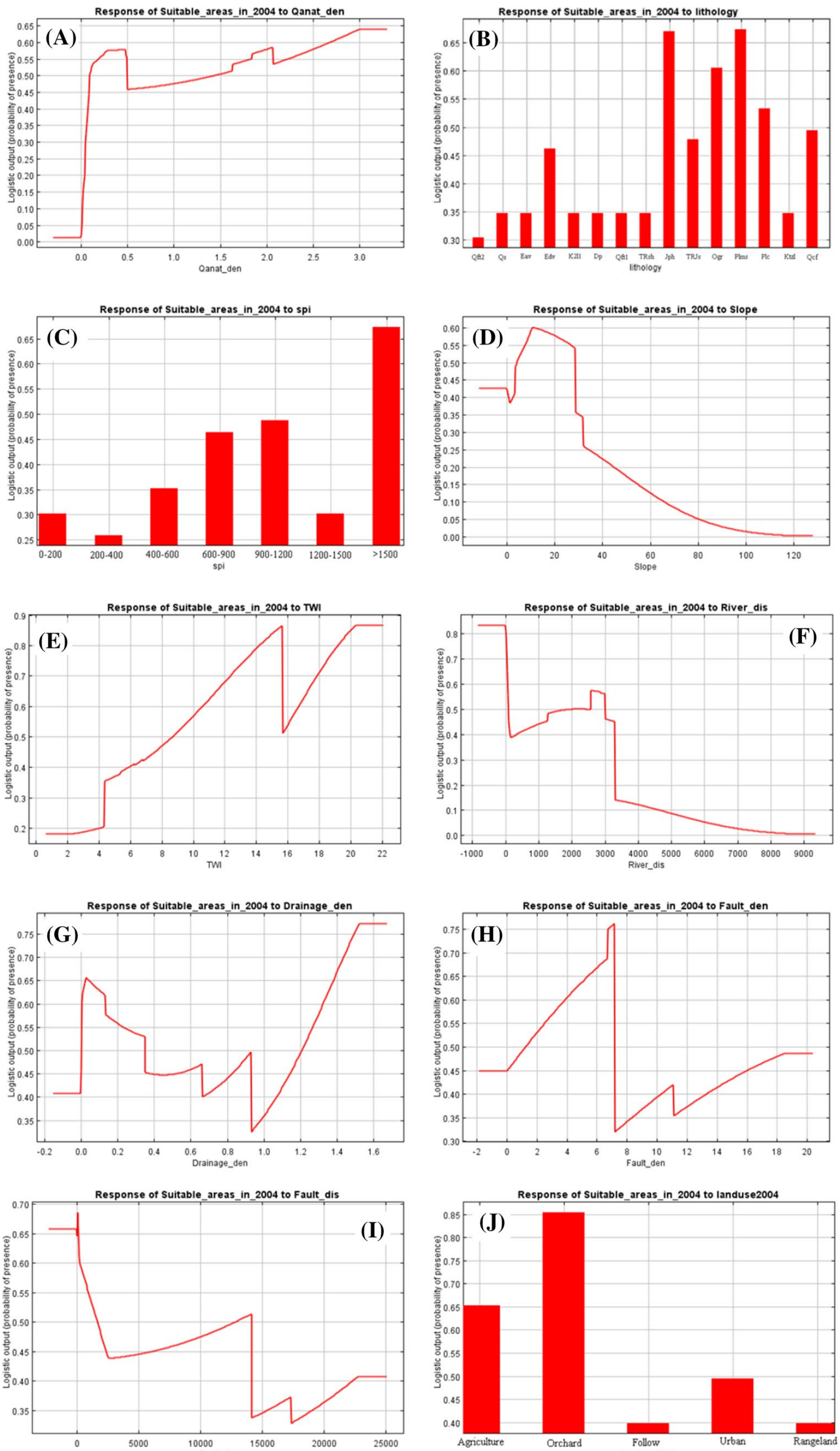
In general, in both 2004 and 2014, the groundwater potential would be reduced by increasing the slope. In 2004, the highest groundwater potential was found where slopes were about 10–30%, but in 2014 the highest groundwater potential was found where slopes were negligible. This is due to qanats in mountainous areas which were observed to have high discharge rates in 2004, but either dried up or the discharge rate dropped significantly in 2014. In 2004, the occurrence of high-yielding qanats in mountainous areas increased the groundwater potential for areas with a high slope. But with the focus of desirable qanats in plain areas in 2014, the potential of mountainous areas decreased. Thus, with the occurrence of drought, the groundwater potential in mountainous areas reduced at a greater rate than in lowland areas (Figs. 7, 8d).

The implementation of the model in both 2004 and 2014 indicates an increase in groundwater potential by increasing the TWI index. Since the TWI is increased in low slopes and extensive accumulation area, the highest TWI is related to the plain area. The existence of an increased trend in the groundwater potential with increasing TWI indicates the importance of plain areas in terms of this factor. It is likely that the decline in the 2004 curve is due to the role of the existing qanats in mountainous regions (Figs. 7, 8e).

By investigating the SPI values in the study area, it is determined that the high SPI values are related to mountainous regions and low SPI values are related to plain areas. In 2014, areas with a minimum and maximum SPI have the highest potential. This result can be interpreted based on the SPI index which is directly related to the contributing area and the slope degree. In 2004, groundwater potential increased in areas with high SPI values (class 7), but in areas with decreased SPI values it was low (class 1). Other areas with moderate SPI values (class 4 and 5) showed an increase in groundwater potential. The difference in the distribution and amount of precipitation in 2004 compared to 2014 can raise the amount of recharge to aquifers and, by compensating for the slope effect, increase the groundwater potential in higher SPI classes (Figs. 7, 8c).

Profile curvature was not used due to the lack of significant impact on modeling in 2014. Garden and agricultural land uses had the greatest impact on the model output in 2004. Since the extracted water from existing qanats is used for the development of garden and agricultural lands, increasing the groundwater potential in these two land use categories is completely reasonable (Fig. 7j).

In both 2004 and 2014, the groundwater potential was reduced with increase in distances from rivers. This is consistent with rivers in the study area being important recharge features and with their influence on groundwater recharge declining with distance. With increasing



(A)
Response of Suitable_areas_in_2004 to Qanat_den
Logistic output (probability of presence)
Qanat_den
(B)
Response of Suitable_areas_in_2004 to lithology
lithology
(C)
Response of Suitable_areas_in_2004 to spi
0-200
200-400
400-600
600-900
900-1200
1200-1500
>1500
spi
(D)
Response of Suitable_areas_in_2004 to Slope
Slope
(E)
Response of Suitable_areas_in_2004 to TWI
TWI
(F)
Response of Suitable_areas_in_2004 to River_dis
River_dis
(G)
Response of Suitable_areas_in_2004 to Drainage_den
Drainage_den
(H)
Response of Suitable_areas_in_2004 to Fault_den
Fault_den
(I)
Response of Suitable_areas_in_2004 to Fault_dis
Fault_dis
(J)
Response of Suitable_areas_in_2004 to landuse2004
Agriculture
Orchard
Follow
Urban
Rangeland
landuse2004

◂**Fig. 7** Response of different factors to the groundwater potential in 2004. **a** Qanat density, **b** lithology, **c** SPI, **d** slope, **e** TWI, **f** river distance, **g** drainage density, **h** fault density, **i** fault distance and **j** land use

'distance from river', the groundwater potentiality values reduced drastically (Figs. 7, 8f).

Survey of the lithology map of the area showed that the greatest impact on the groundwater potential is related to sedimentary rocks, unconsolidated sediments of Quaternary age and igneous rocks. Sedimentary rocks in 2004 have the highest potential and, with the transfer of potential areas to the plains, the impact of the Quaternary formations in 2014 was increased. Additionally, unconsolidated sediments of Quaternary age are considered to have a high groundwater potential, particularly in lowland areas, and fractured granite has the most potential of igneous rocks (Figs. 7, 8b).

This factor was not used due to the lack of significant impact on modeling in 2004. The impact diagram of this factor is a bell-shaped curve, and the greatest impact occurs within the range of 0–1.5. Reduced potential in these parts is due to area constraints (Fig. 8g).

The distance from the fault factor was only used in the 2004 modeling. In general, with the increase in the distance from a fault, the groundwater potential is reduced, but the reduction of impact does not have a regular trend. Due to the role of water transfer from upstream areas by faults, the reason for the potential reduction on getting away from the fault can be explained (Fig. 7i).

Fault density was only used in the 2004 modeling. The impact of this factor does not show a steady trend. Up to a density of about 7, increasing fault density causes impact increase on groundwater potential. In the following, after a sharp drop, the increasing trend continues. The reason for this result can be related to the dual role of faults in the transfer of groundwater (Fig. 7h).

Drainage density was only used in the 2004 modeling. Drainage density in the study area did not have a uniform trend. Up to drainage density factor of about one, the impact on groundwater potential decreases, and in areas with a drainage factor greater than one, the impact increases. The drainage system of the area has a dual function. On the one hand, by increasing the drainage density, the rate of water discharge during rain increases, which leads to a decrease in infiltration. On the other hand, the bed and banks of the river are a good place for infiltration and the recharge of aquifer, and with increasing drainage density, the extent of these areas is also increased. Of course, according to Fig. 7g, the declining trend is very irregular and the increasing trend is regular.

## Conclusions

Determining the spatial distribution of areas that have a high potential for groundwater availability and how these areas change over time provide information to help manage groundwater resources. Drought is one of the phenomena that can change the groundwater potential. In this study, using the MaxEnt model, which is a machine learning model, the groundwater potential was modeled for 2 years, 2004 and 2014, from the arid to semi-arid Gonabad region of eastern Iran. The major difference between these 2 years is the occurrence of a drought p in the years leading up to 2014. For the modeling of groundwater potential, 13 factors including aspect, slope, profile curvature, fault distance, drainage density, fault density, land use, lithology, plan curvature, qanat density, slope, SPI and TWI were used.

The modeling accuracy for training and testing data for both years 2004 and 2014 achieved an AUC of about 0.9, which is indicative of the very good accuracy of the model in zoning the groundwater potential. Groundwater potential was classified into five categories: very low, low, medium, high and very high. The results indicate that in 2 years of model implementation, the concordance between four categories (low, medium, high and very high) was weak.

This result also shows significant changes in the groundwater potential between 2004 and 2014. However, in both years the qanat density factor has the greatest impact on the groundwater potential in the study area. As the existing qanats are very old, the importance of this factor in the model suggests that these structures were well sited when they were constructed.

An investigation of the most significant factors that contribute to groundwater potential indicated that a number of factors could be removed from the model using the 2014 dataset without reducing the integrity of the model.

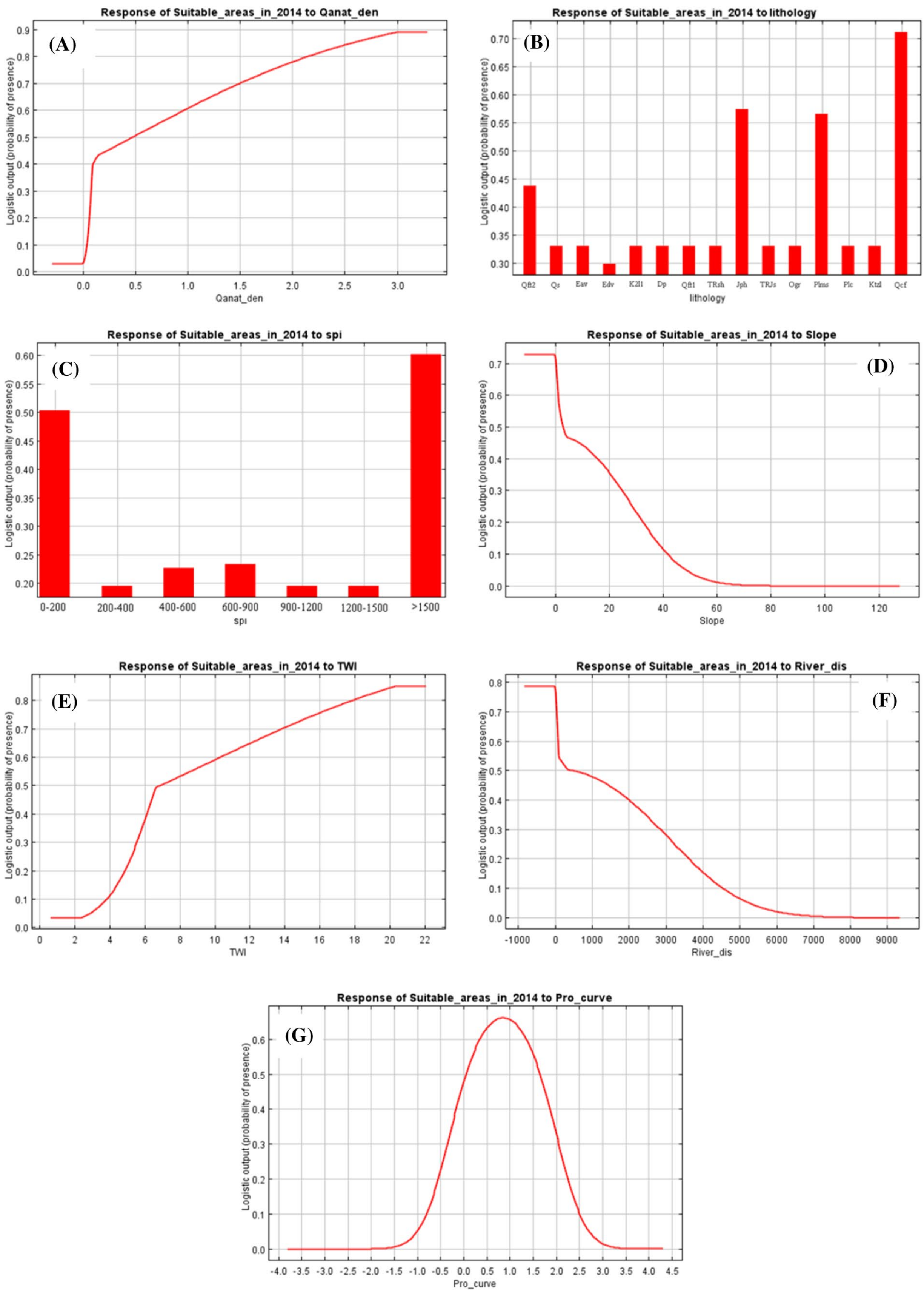



**Fig. 8** Response of different factors to the groundwater potential in 2014. **a** Qanat density, **b** lithology, **c** SPI, **d** slope, **e** TWI, **f** river distance, **g** profile curvature

## References


Adiat KAN, Nawawi MNM, 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. https://doi.org/10.1016/j.jhydrol.2012.03.028

Al Saud M (2010) Mapping potential areas for groundwater storage in Wadi Aurnah Basin, western Arabian Peninsula, using remote sensing and geographic information system techniques. Hydrogeol J 18:1481–1495. https://doi.org/10.1007/s10040-010-0598-9

Anbazhagan S, Ramasamy SM, Gupta DS (2005) Remote sensing and GIS for artificial recharge study, runoff estimation and planning in Ayyar basin, Tamil Nadu, India. Environ Geol 48:158–170

Ariyanto AC (2015) Mapping of possible corridors for Javan Leopard (*Panthera pardus* ssp. *melas*) between Gunung Merapi and Gunung Merbabu National Parks, Indonesia. Doctoral dissertation, University of Twente

Austin M (2007) Species distribution models and ecological theory, a critical assessment and some possible new approaches. Ecol Model 200(1):1–19

Bajat B, Hengl T, Kilibarda M, Krunić N (2011) Mapping population change index in Southern Serbia (1961–2027) as a function of environmental factors. Comput Environ Urban Syst 35(1):35–44

Baldwin RA (2009) Use of maximum entropy modeling in wildlife research. Entropy 11(4):854–866

Bayumi T (2008) Quantitative groundwater resources evaluation in the lower part of Yalamlam basin, Makkah Al Mukarramah, Western Saudi Arabia. JKAU Earth Sci 19:35–56

Bera K, Bandyopadhyay J (2012) Ground water potential mapping in Dulung watershed using remote sensing and GIS techniques, West Bengal, India. Int J Sci Res Publ 2(12):1–7

Berger AL, Pietra VJD, Pietra SAD (1996) A maximum entropy approach to natural language processing. Comput Linguist 22(1):39–71

Bhattacharya AK (2010) Artificial ground water recharge with a special reference to India. Int J Res Rev Appl Sci 4:214–221

Cardenas MB, Wilson JL, Zlotkik VA (2004) Impact of heterogeneity, bed forms, and stream curvature on subchannel hyporheic exchange. Water Resour Res 40(8):W083071–W0830713

Chowdhury A, Jha MK, Chowdary VM (2010) Delineation of groundwater recharge zones and identification of artificial recharge sites in West Medinipur District, West Bengal using RS, GIS and MCDM techniques. Environ Earth Sci 59(6):1209–1222

Convertino M, Troccoli A, Catani F (2013) Detecting fingerprints of landslide drivers, a MaxEnt model. J Geophys Res Earth Surf 118(3):1367–1386

Dai FC, Lee CF, Li J, Xu ZW (2001) Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environ Geol 40(3):381–391

Davis J, Blesius L (2015) A hybrid physical and maximum-entropy landslide susceptibility model. Entropy 17(6):4271–4292

Davoodi Moghaddam D, Rezaei M, Pourghasemi HR, Pourtaghie ZS, Pradhan B (2013) Groundwater spring potential mapping using bivariate statistical model and GIS in the Taleghan watershed, Iran. Arab J Geosci 8(2):913–929

Deepa S, Venkateswaran S, Ayyandurai R et al (2016) Groundwater recharge potential zones mapping in upper Manimuktha Sub basin Vellar river Tamil Nadu India using GIS and remote sensing techniques. Model Earth Syst Environ 2:137. https://doi.org/10.1007/s40808-016-0192-9

Dinesh Kumar PK, Gopinath G, Seralathan P (2007) Application of remote sensing and GIS for the demarcation of groundwater potential zones of a river basin in Kerala, southwest coast of India. Int J Remote Sens 28(24):5583–5601

Dyke J, Kleidon A (2010) The maximum entropy production principle, its theoretical foundations and applications to the earth system. Entropy 12(3):613–630

Edvardsen A, Bakkestuen V, Halvorsen R (2011) A fine-grained spatial prediction model for the red-listed vascular plant *Scorzonera humilis*. Nord J Bot 29(4):495–504

Edwards TC, Cutler DR, Zimmermann NE, Geiser L, Alegria J (2005) Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86(5):1081–1090

Elith J, Phillips S, Hastie T, Dudík M, Chee Y, Yates C (2010) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17:43–47

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17(1):43–57

Ettazarini S (2007) Groundwater potential index: a strategically conceived tool for water research in fractured aquifers. Environ Geol 52:477–487

Feeley KJ, Silman MR (2011) Keep collecting, accurate species distribution modeling requires more collections than previously thought. Divers Distrib 17(6):1132–1140

Geological Survey Department of Iran (GSDI) (1997). http://www.gsi.ir/Main/Lang_en/index.html. Accessed 20 May 2017

Ghorbani Nejad S, Falah F, Daneshfar M, Haghizadeh A, Rahmati O (2017) Delineation of groundwater potential zones using remote sensing and GIS-based data-driven models. Geocarto Int 32(2):167–187

Gooseff MN, Anderson JK, Wondzell SM et al (2005) A modelling study of hyporheic exchange pattern and the sequence, size, and spacing of stream bedforms in mountain stream networks, Oregon, USA. Hydrol Proc 19(15):2915–2929. https://doi.org/10.1002/hyp.5790

Graham CH, Elith J, Hijmans RJ, Guisan A, Townsend Peterson A, Loiselle BA (2008) The influence of spatial errors in species occurrence data used in distribution models. J Appl Ecol 45(1):239–247

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006) Using niche-based models to improve the sampling of rare species. Conserv Biol 20(2):501–511

Gutiérrez AG, Schnabel S, Contador JFL (2009) Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. Ecol Model 220:3630–3637

Halvorsen R (2012) A gradient analytic perspective on distribution modelling. Sommerfeltia 35:1–165

Hernandez PA, Franke I, Herzog SK, Pacheco V, Paniagua L, Quintana HL et al (2008) Predicting species distributions in poorly-studied landscapes. Biodivers Conserv 17(6):1353–1366

Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos 117(6):847–858

Huset R (2013) A GIS-based analysis of the environmental predictors of dispersal of the emerald ash borer in New York. MA thesis, Syracuse University

Jaime R, Alcántara JM, Bastida JM, Rey PJ (2015) Complex patterns of environmental niche evolution in Iberian columbines (genus *Aquilegia*, Ranunculaceae). J Plant Ecol 8(5):457–467

Jaynes ET (1957a) Information theory and statistical mechanics. Phys Rev 106(4):620

Jaynes ET (1957b) Information theory and statistical mechanics. II. Phys Rev 108(2):171

Jenness J (2013) DEM surface tools for ArcGIS. Jenness Enterprises. http://www.jennessent.com/arcgis/surface:area.htm. Accessed Mar

Jha MK, Chowdhury A, Chowdary VM, Peiffer S (2007) Groundwater management and development by integrated remote sensing and geographic information systems: prospects and constraints. Water Resour Manag 21:427–467

Kim HG, Lee DK, Park C, Kil S, Son Y, Park JH (2015) Evaluating landslide hazards using RCP 4.5 and 8.5 scenarios. Environ Earth Sci 73(3):1385–1400

Kleidon A, Malhi Y, Cox PM (2010) Maximum entropy production in environmental and ecological systems. Philos Trans R Soc B 365(1545):1297–1302

Kornejady A, Ownegh M, Bahremand A (2017a) Landslide susceptibility assessment using maximum entropy model with two different data sampling methods. Catena 152:144–162

Kornejady A, Ownegh M, Rahmati O, Bahremand A (2017b) Landslide susceptibility assessment using three bivariate models considering the new topo-hydrological factor: HAND. Geocarto Int. https://doi.org/10.1080/10106049.2017.1334832

Kumar S, Stohlgren TJ (2009) Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. J Ecol Nat Environ 1(4):094–098

Liu Y, Guo Q, Tian Y (2012) A software framework for classification models of geographical data. Comput Geosci 42:47–56

Loiselle BA, Jørgensen PM, Consiglio T, Jiménez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections, does climate bias in collection sampling influence model outcomes? J Biogeogr 35 (1): 105–116

Magesh NS, Chandrasekar N, Soundranayagam JP (2012) Delineation of groundwater potential zones in Theni district, Tamil Nadu, Environ Earth Sci using remote sensing, GIS and MIF techniques. Geosci Front 3(2):189–196

Mair A, El-Kadi AI (2013) Logistic regression modeling to assess groundwater vulnerability to contamination in Hawaii, USA. J Contam Hydrol 153:1–23. https://doi.org/10.1016/j.jconhyd.2013.07.004

Manap MA, Sulaiman WNA, Ramli MF, Pradhan B, Surip N (2013) A knowledge-driven GIS modeling technique for groundwater potential mapping at the Upper Langat Basin, Malaysia Arab. J Geosci 6:1621–1637. https://doi.org/10.1007/s12517-011-0469-2

McCarthy E, Moretti D, Thomas L, DiMarzio N, Morrissey R, Jarvis S et al (2011) Changes in spatial and temporal distribution and vocal behavior of Blainville's beaked whales (*Mesoplodon densirostris*) during multiship exercises with mid-frequency sonar. Mar Mamm Sci 27(3):E206–E226

Medley KA (2010) Niche shifts during the global invasion of the Asian tiger mosquito, *Aedes albopictus* Skuse (Culicidae), revealed by reciprocal distribution models. Glob Ecol Biogeogr 19(1):122–133

Mert A, Özkan K, Şentürk Ö, Negiz MG (2016) Changing the potential distribution of Turkey oak (*Quercus cerris* L.) under climate change in Turkey. Pol J Environ Stud 25(4):1633–1638

Metz CE (1978) Basic principles of ROC analysis. Semin Nucl Med 8(4):283–298 **(WB Saunders)**

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modeling: a review of hydrological, geomorphological and biological applications. Hydro Process 5:3–30

Moosavi V, Niazi Y (2016) Development of hybrid wavelet packet-statistical models (WP-SM) for landslide susceptibility mapping. Landslides 13(1):97–114

Moreno R, Zamora R, Molina JR, Vasquez A, Herrera M (2011) Predictive modeling of microhabitats for endemic birds in South Chilean temperate forests using Maximum entropy (Maxent). Ecol Inform 6(6):364–370

Naghibi A, Pourghasemi HR (2015) A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods for groundwater potential mapping in Iran. Water Resour Manag 29(14):5217–5236. https://doi.org/10.1007/s11269-015-1114-8

Naghibi SA, Pourghasemi HR, Pourtaghi ZS, 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

Naghibi SA, Pourghasemi HR, Dixon B (2016) Groundwater spring potential using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ Monit Assess. https://doi.org/10.1007/s10661-015-5049-6

Naghibi SA, Ahmadi K, Daneshi A (2017a) Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resour Manag 31:2761

Naghibi SA, Moghaddam DD, Kalantar B, Pradhan B, Kisi O (2017b) A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. J Hydrol 548:471–483

Naghibi SA, Pourghasemi HR, Abbaspour K (2018) A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. Theor Appl Climatol 131(3–4):967–984

Nampak H, Pradhan B, Manap MA (2014) Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. J Hydrol. https://doi.org/10.1016/j.jhydrol.2014.02.053

Neshat A, Pradhan B, Pirasteh S, Shafri HZM (2014) Estimating groundwater vulnerability to pollution using a modified DRASTIC model in the Kerman agricultural area, Iran. Environ Earth Sci 71(7):3119–3131

Ng AY, Jordan MI (2001) On discriminative versus generative classifiers, a comparison of logistic regression and naive Bayes. Adv Neural Inf Process Syst 14:605–610

Niamir A, Skidmore AK, Toxopeus AG, Munoz AR, Real R (2011) Finessing atlas data for species distribution models. Divers Distrib 17(6):1173–1185

Nieves V, Wang J, Bras RL (2011) Statistics of multifractal processes using the maximum entropy method. Geophys Res Lett 38:17

Ozdemir A (2011) 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 411:290–308

Park NW (2015) Using maximum entropy modeling for landslide susceptibility mapping with multiple geoenvironmental data sets. Environ Earth Sci 73(3):937–949

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Model 133(3):225–245

Perrier E, Salkini AB (1991) Supplemental Irrigation in the Near East and North Africa. Kluwer Academic Publisher, Norwell

Peterson AT (2011) Ecological niches and geographic distributions (MPB-49), vol 49. Princeton University Press, Princeton

Phillips S (2012) A brief tutorial on Maxent. Lessons Conserv 3:107–135

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent, new extensions and a comprehensive evaluation. Ecography 31(2):161–175

Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. In: Proceedings of the twenty-first international conference on machine learning. ACM, New York, p. 83

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190(3):231–259

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models, implications for background and pseudo-absence data. Ecol Appl 19(1):181–197

Pontius RG, Schneider LC (2001) Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agric Ecosyst Environ 85(1):239–248

Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province. Iran Environ Earth Sci 75(3):1–17

Pourghasemi HR, Rossi M (2017) Landslide susceptibility modeling in a landslide prone area in Mazandarn Province, north of Iran: a comparison between GLM, GAM, MARS, and M-AHP methods. Theor Appl Climatol 130(1–2):609–633

Pourghasemi HR, Pradhan B, Gokceoglu C, Mohammadi M, Moradi HR (2013) Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab J Geosci 6(7):2351–2365

Pradhan B (2009) Groundwater potential zonation for basaltic watersheds using satellite remote sensing data and GIS techniques. Cent Eur J Geosci 1(1):120–129

Pueyo S, He F, Zillio T (2007) The maximum entropy formalism and the idiosyncratic theory of biodiversity. Ecol Lett 10(11):1017–1028

Quinn SA, Gibbs JP, Hall MH, Petokas PJ (2013) Multi scale factors influencing distribution of the eastern hellbender salamander (*Cryptobranchus alleganiensis alleganiensis*) in the northern segment of its range. J Herpetol 47(1):78–84

Rahmati O, Melesse AM (2016) Application of Dempster–Shafer theory, spatial analysis and remote sensing for groundwater potentiality and nitrate pollution analysis in the semi-arid region of Khuzestan, Iran. Sci Total Environ 568:1110–1123

Rahmati O, Haghizadeh A, Stefanidis S (2016a) Assessing the accuracy of GIS-based analytical hierarchy process for watershed prioritization; Gorganrood River Basin, Iran. Water Resour Manag 30(3):1131–1150

Rahmati O, Pourghasemi HR, Melesse AM (2016b) 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

Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. J Biogeogr 30(11):1719–1727

Robertson MP, Cumming GS, Erasmus BFN (2010) Getting the most out of atlas data. Divers Distrib 16(3):363–375

Rodriguez-Galiano V, Mendes MP, Garcia-Soldado MJ, Chica-Olmo M, Ribeiro L (2014) Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain). Sci Total Environ 476–477:189–206

Ross SM (2014) Introduction to probability models. Academic Press, Orlando

Sarkar B, Deota B, Raju P, Jugran D (2001) A geographic information system approach to evaluation of groundwater potentiality of Shamri micro watershed in the Shimla Taluk, Himachal Pradesh. J Indian Soc Remote Sens 29(3):151–164

Saud MA (2010) Mapping potential areas for groundwater storage in Wadi Aurnah Basin, western Arabian Peninsula, using remote sensing and geographic information system techniques. Hydrogeol J 18:1481–1495

Senthil Kumar GR, Shankar K (2014) Assessment of groundwater potential zones using GIS. Front Geosci 2(1):1–10

Shahid S, Nath SK, Roy J (2000) Groundwater potential modeling in a soft rock area using a GIS. Int J Remote Sens 21(9):1919–1924

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656 **[Mathematical Reviews (MathSciNet), MR10, 133e]**

Shannon CE (1951) Prediction and entropy of printed English. Bell Syst Tech J 30(1):50–64

Shen G, Pimm SL, Feng C, Ren G, Liu Y, Xu W et al (2015) Climate change challenges the current conservation strategy for the giant panda. Biol Conserv 190:43–50

Sidle R, Ochiai H (2006) Processes, prediction, and land use. Water resources monograph. American Geophysical Union, Washington

Singh AK, Prakash SR (2002) An integrated approach of remote sensing, geophysics and GIS to evaluation of groundwater potentiality of Ojhala sub-watershed, Mirjapur district, UP, India. In: Asian conference on GIS, GPS, aerial photography and remote sensing, Bangkok, Thailand

Sorichetta A, Ballabio C, Masetti M, Robinson GR Jr, Sterlacchini S (2013) A comparison of data-driven groundwater vulnerability assessment methods. Ground Water 51(6):866–879. https://doi.org/10.1111/gwat.12012

Storey RG, Howard KWF, Williams DD (2003) Factors controlling riffle-scale hyporheic exchange flows and their seasonal changes in a gaining stream: a three-dimensional groundwater flow model. Water Resour Res 39(2):1034

Subyani A (2005) Hydrochemical identification and salinity problem of ground-water in Wadi Yalamlam basin, Western Saudi Arabia. J Arid Environ 60:53–66

Thuiller W, Richardson DM, Pyšek P, Midgley GF, Hughes GO, Rouget M (2005) Niche-based modeling as a tool for predicting the risk of alien plant invasions at a global scale. Glob Chang Biol 11(12):2234–2250

Townsend Peterson A, Papeş M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling, a comparison of GARP and Maxent. Ecography 30(4):550–560

Vapnik VN, Vapnik V (1998) Statistical learning theory. Wiley, New York

Veloz SD (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. J Biogeogr 36(12):2290–2299

Wahyudi AD, Bartzke M, Küster E, Bogaert P (2012) Maximum entropy estimation of a benzene contaminated plume using ecotoxicological assays. Environ Pollut 172C:170–179

Wang J, Bras RL (2011) A model of evapotranspiration based on the theory of maximum entropy production. Water Resour Res 47(3):1–10. https://doi.org/10.1029/2010WR009392

Warren DL, Seifert SN (2011) Ecological niche modeling in Maxent, the importance of model complexity and the performance of model selection criteria. Ecol Appl 21(2):335–342

Williams RJ (2010) Simple MaxEnt models explain food web degree distributions. Theor Ecol 3(1):45–52

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A (2008) Effects of sample size on the performance of species distribution models. Divers Distrib 14(5):763–773

Wollan AK, Bakkestuen V, Kauserud H, Gulden G, Halvorsen R (2008) Modeling and predicting fungal distribution patterns using herbarium data. J Biogeogr 35(12):2298–2310

Wolmarans R, Robertson MP, van Rensburg BJ (2010) Predicting invasive alien plant distributions, how geographical bias in occurrence records influences model performance. J Biogeogr 37(9):1797–1810

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Campbell Grant EH, Veran S (2013) Presence-only modelling using MAXENT, when can we trust the inferences? Methods Ecol Evol 4(3):236–243

Yu J, Wang C, Wan J, Han S, Wang Q, Nie S (2014) A model-based method to evaluate the ability of nature reserves to protect endangered tree species in the context of climate change. For Ecol Manag 327:48–54

Zabihi M, Pourghasemi HR, Sadat Pourtaghi Z, Behzadfar M (2016) GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. Environ Earth Sci 75:665