CrossMark

ORIGINAL PAPER

Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms

Wei Chen[1] · Hamid Reza Pourghasemi[2] · Seyed Amir Naghibi[3]

Abstract The main objective of the current study is to apply a random forest (RF) data-driven model and prioritization of landslide conditioning factors according to this method and its comparison to a multivariate adaptive regression spline (MARS) model for landslide susceptibility mapping in China. For this purpose, at first, landslide locations were identified by earlier reports, aerial photographs, and field surveys and a total of 348 landslides were mapped from various sources in GIS. Then, the landslide inventory was randomly split into a training dataset (70% = 244 landslides) and the remaining (30% = 104 landslides) were used for validation. In this study, 12 landslide conditioning factors were applied to detect the most susceptible areas. These factors were slope aspect, altitude, distance to faults, lithology, normalized difference vegetation index, plan curvature, profile curvature, distance to rivers, distance to roads, slope angle, stream power index, and topographic wetness index. The relationship between each conditioning factor and landslide was finalized using a frequency ration (FR) model. Subsequently, landslide-susceptible areas were mapped using the MARS and RF models. The results revealed that the most important conditioning factors according to the accuracy measure (mean decrease) of the RF model are lithology (23.47%), distance to faults (22.21%), and altitude (19.58%). We also notice that altitude (19.04%), distance to faults (18.83%), and distance to roads (15.29%) have the highest importance according to the Gini measure. Finally, the accuracy of the landslide susceptibility maps produced from the two models was verified using a receiver operating characteristics curve. The results showed that the landslide susceptibility map produced using the MARS model has a higher prediction rate than RF by area under the curve values of 87.51 and 77.32%, respectively. According to the validation results, the map produced by the MARS model exhibits the better accuracy and could be proposed for land-use planning in the study area.

Keywords Landslide susceptibility mapping · Multivariate adaptive regression spline · Random forest · Geographic information system (GIS) · China

Introduction

Landslides are one of the damaging natural hazards occurring in many countries and lead to the loss of life and property. According to García-Rodríguez et al. (2008) and Dou et al. (2014), the annual loss of property caused by landslides is greater than other types of natural hazards, such as earthquakes and floods. Throughout the world, landslides cause hundreds of billions of dollars in damage and thousands of casualties and fatalities each year; additionally, they cause large environmental damages each year (Aleotti and Chowdhury 1999). In Shangnan County, China, landslides are occurring frequently mainly due to anthropogenic activities; thus, they have a great impact on the local economic and environmental developments.

✉ Hamid Reza Pourghasemi
hamidreza.pourghasemi@yahoo.com;
hr.pourghasemi@shirazu.ac.ir

[1] College of Geology and Environment, Xi'an University of Science and Technology, Xi'an 710054, China

[2] Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran

[3] Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, Noor, Mazandaran, Iran

Currently, thousands of people are still threatened by landslides and live under the landslide-prone areas in the study area. Thus, it is essential to prepare landslide susceptibility maps for mitigating of landslide hazards in such areas.

Generally, landslide susceptibility mapping can be prepared as qualitative or quantitative methods, and direct or indirect techniques (Fell et al. 2008; Kayastha et al. 2013; Nourani et al. 2014). The reliability of landslide susceptibility maps depends on the quality and amount of available data, the selection of methodology of modeling, and the working scale (Baeza and Corominas 2001). The landslide susceptibility modeling methodology could be categorized into heuristic, deterministic, and statistical methods (Guillard and Zezere 2012; Regmi et al. 2014a). The heuristic method, which is established based on the assumption that the relationships between landslide susceptibility and the preparatory variables are known and are specified in the models, is a direct or semi-direct mapping methodology (Regmi et al. 2014a). This method was used in many landslide susceptibility researches (Ruff and Czurda 2008; Jaiswal et al. 2013).

In deterministic or physically based analysis, according to Regmi et al. (2014a), the landslide is determined employing slope stability methods in the form of calculating safety factors. These methods need a large number of detailed input factors to build the model; thus, these methods could be appropriate for small watersheds (Regmi et al. 2014b). Geotechnical and groundwater features are two main bases of the deterministic methods. Mathematical methods are employed to define the factor of safety of the unstable slopes (Gökceoglu and Aksoy 1996) and slope stability methods are employed for defining the landslide hazard (Clerici et al. 2006). The most important weakness of geotechnical and safety factor-based methods is that these methods are feasible only for areas where landslide types are simple and geomorphic and geologic features are relatively similar (Van Westen and Terlien 1996). Moreover, the mentioned approaches are difficult to do in regional landslide susceptibility investigations. These models were also used by many researchers (Gökceoglu and Aksoy 1996; Cervi et al. 2010; Jia et al. 2012; Armaş et al. 2014; Akgun and Erkan 2016). Considering the mentioned weak points of deterministic models, in this study, we employed data mining models for landslide susceptibility mapping.

Statistical models, which are the most commonly used approaches worldwide, have also been employed in landslide susceptibility evaluation using GIS, such as the logistic regression (Akgun 2012; Xu et al. 2012b; Devkota et al. 2013; Ozdemir and Altural 2013; Park et al. 2013; Grozavu et al. 2013; Talaei 2014; Chen et al. 2016a; Karimi Sangchini et al. 2016), frequency ratio (Vijith and Madhu 2008; Poudyal et al. 2010; Pourghasemi et al. 2014; Regmi et al. 2014a, b; Chen et al. 2015), certainty factor (Kanungo et al. 2011; Devkota et al. 2013; Pourghasemi et al. 2013d), weights of evidence (Dahal et al. 2008; Ozdemir and Altural 2013; Pourghasemi et al. 2013d; Regmi et al. 2014a), statistical index (Yilmaz et al. 2012; Regmi et al. 2014a), evidential belief function (Ding et al. 2016; Pourghasemi and Kerle 2016), and index of entropy models (Mihaela et al. 2011; Devkota et al. 2013; Jaafari et al. 2014; Pourghasemi et al. 2012c).

In addition, some other widely accepted methods including an artificial neural network (ANN; Pradhan and Lee 2010; Yilmaz 2010a; Tsangaratos and Benardos 2014; Nourani et al. 2014; Tien Bui et al. 2015; Gorsevski et al. 2016), an analytical hierarchy process (Pourghasemi et al. 2013b; Park et al. 2013; Mandal and Maiti 2015; Chen et al. 2016b), fuzzy logic (Pourghasemi et al. 2012b; Guettouche 2013; Zhu et al. 2014; Kumar and Anbalagan 2015), support vector machine (Yilmaz 2010b; Marjanović et al. 2011; Xu et al. 2012a, b; Pradhan 2013; Peng et al. 2014; Tien Bui et al. 2015; Chen et al. 2016d), general additive model (Park and Chi 2008; Petschko et al. 2012), random forest (RF; Youssef et al. 2015b; Chen et al. 2017), and neuro-fuzzy methods (Pradhan 2013; Lee et al. 2015) have also been applied for landslide susceptibility evaluation.

In summary, in order to mitigate landslide hazards, all these models can be used to prepare landslide susceptibility maps. On the other hand, consideration of literature review on landslide susceptibility evaluation shows that there is no paper comparing the performance of multivariate adaptive regression spline (MARS) and RF models in landslide susceptibility mapping so far. Therefore, the main difference between the present study and the methods described in the aforementioned literature is that MARS and RF models were applied for landslide susceptibility mapping in the study area and their performances were compared together. Also, we tried to determine weight of each conditioning factor and its effect on landslide occurrence in the study area by an RF data-driven method.

Study area

The study area is located in Shangnan County, China, between latitudes 33°06′–33°44′N, and longitudes 110°24′–111°01′E (Fig. 1). It covers roughly a surface area of 2307 km². The altitude of the area ranges from 189 to 2050 m a.s.l. The landform can be classified into mountain, hill, and plain. The slope angles of the area range from 0° to 65°. The rivers of the study area belong to the Yangtze River basin. The mean annual rainfall according to a local station in the past 40 years is around 829.8 mm. Also, based on the
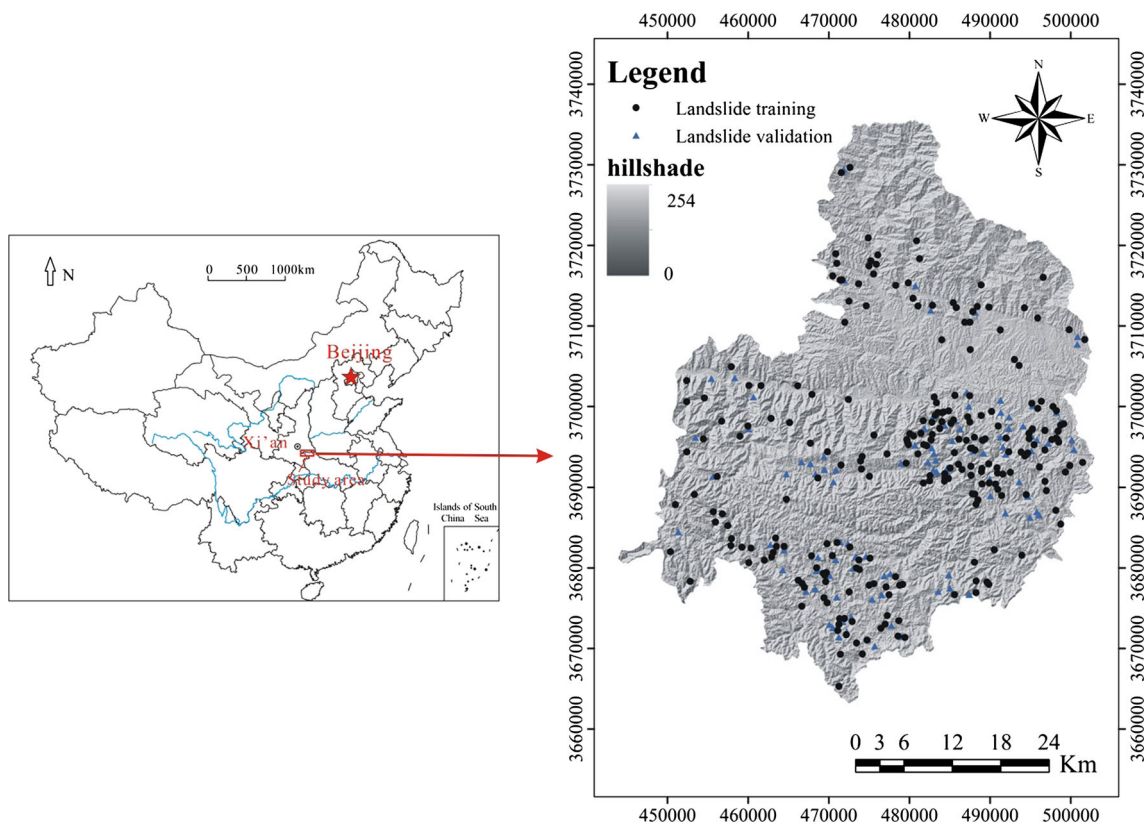
Fig. 1 Location map of the study area

records from China's meteorological department (C.H. of China Meteorological Administration), the minimum and maximum rainfall occurs in December and July, respectively. The average mean annual temperature is 14.6 °C.

In terms of geological background, the geological structure system of the study area is located at the boundary of the North China and the Yangtze plates. Therefore, the geological structure is complex and the tectonic deformation is intense. The faults of the study area have mainly a NW–SE direction. These faults divide the study area into several structural zones that affect the distribution of landslides and geological formations in the study area. There are many geological formations with different ages in the study area, which vary from Archeozoic to Cenozoic. The lithological units are mainly slate, phyllite, schist, gneiss, limestone, dolomite, sandstone, and intrusive rocks. According to Table 1, the softer metamorphic rocks and hard carbonate rocks comprise the highest percentages of the area (29.72 and 28.31%, respectively).

Methodology

As it was mentioned, the main objective of this study was to prepare landslide susceptibility maps in Shangnan County, China, using MARS and RF models. For this

purpose, a flowchart of the applied methodology including the following steps is presented in Fig. 2, including landslide inventory mapping, gathering of the landslide conditioning factors, preparation of thematic layers in the GIS, determination of relationships between each conditioning factor and landslide inventory maps using a frequency ratio (FR), application of MARS and RF models in R statistical packages and landslide susceptibility mapping, evaluation of the prediction accuracy of the above models using receiver operating characteristics (ROC) curves, and selection of the best model.

Data production

Landslide inventory map

Landslide inventories include information related to topics of the regional landslide locations, activity, types, and physical properties that are usually mapped with an associated database (Fell et al. 2008; Demir et al. 2015). Accuracy of the data related to landslide occurrences is very important for landslide susceptibility analysis, hazard, and risk. So, the first step in these analyses is landslide inventory mapping. In the landslide inventory, some

Table 1 Spatial relationship between landslide conditioning factors and landslide by frequency ratio model

Conditioning factors	Classes	Percentage of domain	Percentage of landslide	Frequency ratio
Slope aspect	Flat	0.0064	0.0000	0.0000
	North	11.5951	11.0656	0.9543
	Northeast	12.2792	13.1148	1.0680
	East	13.3268	18.0328	1.3531
	Southeast	13.4759	9.0164	0.6691
	South	12.0249	13.9344	1.1588
	Southwest	12.8589	9.8361	0.7649
	West	12.3422	12.7049	1.0294
	Northwest	12.0906	12.2951	1.0169
Altitude (m)	189–500	13.2934	17.2131	1.2949
	500–800	44.9653	65.9836	1.4674
	800–1100	26.5281	13.1148	0.4944
	1100–1400	13.4948	3.6885	0.2733
	1400–1700	1.5007	0.0000	0.0000
	1700–2050	0.2177	0.0000	0.0000
Distance to faults (m)	0–1000	52.1243	69.6721	1.3367
	1000–2000	22.9780	22.5410	0.9810
	2000–3000	8.4515	5.3279	0.6304
	3000–4000	3.8329	0.8197	0.2138
	>4000	12.6132	1.6393	0.1300
Lithology	Harder metamorphic rocks	19.8345	11.8852	0.5992
	Softer metamorphic rocks	29.7192	57.7869	1.9444
	Hard carbonate rocks	28.3142	17.6230	0.6224
	Hard intrusive rocks	20.1274	9.8361	0.4887
	Soft gravelly soils	2.0047	2.8689	1.4311
NDVI	−0.23 to 0.17	2.4846	3.2787	1.3196
	0.17–0.33	8.9539	6.9672	0.7781
	0.33–0.42	27.4997	27.0492	0.9836
	0.42–0.51	37.1919	31.9672	0.8595
	0.51–0.71	23.8699	30.7377	1.2877
Plan curvature (m)100	−9.57 to −1.09	10.2106	9.4262	0.9232
	−1.09 to −0.11	33.5971	34.0164	1.0125
	−0.11 to 0.88	40.3158	41.3934	1.0267
	0.88–11.42	15.8765	15.1639	0.9551
Profile curvature (m) (100/m)	−11.93 to −1.30	10.9870	8.6066	0.7833
	−1.30 to −0.02	36.6113	36.0656	0.9851
	−0.02 to 1.26	40.3569	42.2131	1.0460
	1.26–11.43	12.0448	13.1148	1.0888
Distance to rivers (m)	0–200	42.3533	56.9672	1.3450
	200–400	29.6864	29.5082	0.9940
	400–600	15.6037	7.7869	0.4990
	600–800	6.3106	2.4590	0.3897
	>800	6.0460	3.2787	0.5423
Distance to roads (m)	0–500	26.8163	46.3115	1.7270
	500–1000	21.1049	17.2131	0.8156
	1000–1500	16.9144	15.9836	0.9450
	1500–2000	12.8858	6.9672	0.5407
	>2000	22.2787	13.5246	0.6071

Table 1 continued

Conditioning factors	Classes	Percentage of domain	Percentage of landslide	Frequency ratio
Slope angle (°)	0–10	13.3904	16.3934	1.2243
	10–20	29.9985	35.6557	1.1886
	20–30	32.1013	27.4590	0.8554
	30–40	18.8384	17.6230	0.9355
	40–50	5.2014	2.4590	0.4728
	50–65	0.4700	0.4098	0.8720
SPI	0–30	48.4231	52.4590	1.0833
	30–60	19.2491	18.0328	0.9368
	60–90	9.3641	9.4262	1.0066
	>90	22.9637	20.0820	0.8745
TWI	<5	42.2010	37.2951	0.8837
	5–7	43.5944	47.9508	1.0999
	7–9	9.9613	11.4754	1.1520
	>9	4.2433	3.2787	0.7727
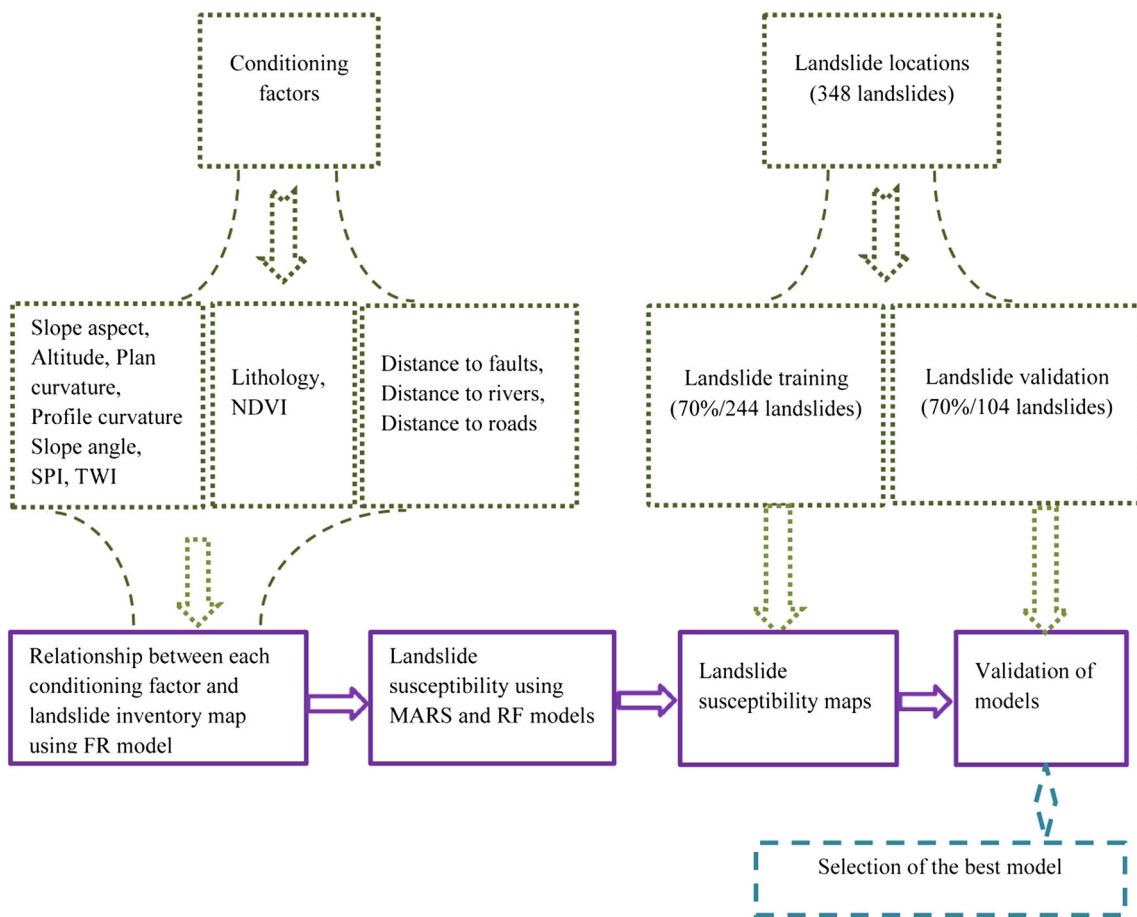
Fig. 2 Flowchart of the used methodology

historic information about landslide occurrences including the frequency, volumes, damage, and types of the landslide phenomena is the backbone of landslide susceptibility studies (van Westen et al. 2006; Youssef et al. 2014). In the present study, landslides were collected from historical reports, interpretation of aerial photographs, and extensive field surveys using the global positioning system (GPS) to locate the position of landslides. A total of 348 landslides were identified and mapped (centroid) in Fig. 1. From these 348 landslides, 244 (70%) locations (grid cells) were randomly selected for training of the models, and 104 (30%) were used for validation of the built models (Maps). 348 non-landslide points (grid cells) were randomly selected from the landslide-free areas, and were also randomly divided into a ratio of 70/30 (244/104) to build training and validation datasets. In the present study, the presence of landslides was assigned a value of 1, while the absence of landslide was assigned a value of 0. Finally, the values of all the landslide conditioning factors were extracted to landslide and non-landslide grid cells to build training and validation datasets for running the mentioned data mining techniques in R statistical packages.

The use of historical landslide inventories that summarize past multiple landslide events may enable a robust landslide susceptibility mapping because it reflects various environmental conditions, and the number of available data tends to be large (Malamud et al. 2004; Paudel et al. 2016). In the present study, detailed landslide data for more than 10 rainfall events occurred in years of 1998, 2000, and 2003–2015 have been used, which crossed a time span of more than 20 years. Besides, with highly developed remote sensing technology, multi-temporal aerial photographs have been used to map landslides after each rainfall event which were verified by field surveys. All these data sources ensure a reliable landslide susceptibility mapping for the study area.

According to our analyses in a GIS environment, the size of the smallest landslide is nearly 15 m^2, whereas, the largest is more than 30,000 m^2, while the average is 9600 m^2. For the volume of the landslides, more than 90% of landslides are less than 100,000 m^3, and more than 85% of the landslides are shallow-seated landslides (<6 m). The landslide area percentage (LAP), which is expressed as a percentage of the area affected by landslide activity, is
$$LAP = (3.34\ km^2/2307\ km^2) \times 100\% = 0.145\% \quad (Xu$$
et al. 2012b).

The quality of a landslide inventory is affected by its accuracy; however, determining the accuracy of the inventory does not have a specific standard (Galli et al. 2008). Some methods have been suggested in order to increase the accuracy of spatial inventories such as: (1) employing very high-resolution digital elevation models (DEMs) for analyzing surface morphology, (2)

interpretation of satellite images, and (3) application of new tools for facilitation of field mapping.

Landslide conditioning factors

The selection of landslide conditioning factors depends on the characteristics of the study area, the landslide type, the scale of the analysis, etc. (Tseng et al. 2015). However, there is no agreement on the universal guidelines for selecting landslide conditioning factors (Xu et al. 2013). In the present study, the landslide conditioning factors were selected from those most commonly used in the literature to evaluate landslide susceptibility (Lee and Talib 2005; Xu et al. 2012a, b; Pradhan 2013; Pourghasemi et al. 2013b; Tien Bui et al. 2015; Hong et al. 2015, 2016; Chen et al. 2014, 2015, 2016c). Particularly, the results of satellite image interpretation and field surveys suggest the following 12 parameters: slope aspect, altitude, distance to faults, lithology, normalized difference vegetation index (NDVI), plan curvature, profile curvature, distance to rivers, distance to roads, slope angle, stream power index (SPI), and topographic wetness index (TWI). Therefore, these 12 landslide conditioning factors were selected in the present study and were standardized to the same size (30 × 30 m) for further analyses.

Slope aspect was selected as a landslide conditioning factor which represents the direction of slope (Ercanoglu et al. 2004; Pourghasemi et al. 2012a, b). The slope aspect of the study area was classified into nine directional classes as flat, north, northeast, east, southeast, south, west, southwest, and northwest (Fig. 3a). Altitude/elevation is another frequently used conditioning factor for landslide susceptibility analysis. Ercanoglu and Gokceoglu (2002) mentioned that altitude influences earth surface and topographic attributes which account for spatial variability of erosion, precipitation, soil thickness, and vegetation. In the present study, the elevation of the study area ranges from 189 to 2050 m (Fig. 3b). The faults are responsible for triggering a huge number of landslides due to the tectonic breaks that usually decrease the rock strength. In the study area, the distance to faults was prepared using the geological map and ranges from 0 to 17,526 m (Fig. 3c). Lithology is one of the most common influence factors in most landslide susceptibility studies. Since different lithological units have various landslide susceptibility values, they are very important in the landslide occurrences. The lithology map of the study area was extracted from existing geological maps. The study area is covered with several lithological units, and classified into five groups of harder metamorphic rocks, softer metamorphic rocks, hard carbonate rocks, hard intrusive rocks, and soft

Fig. 3 Thematic maps of the study area: **a** slope aspect; **b** altitude; **c** distance to faults; **d** lithology; **e** NDVI; **f** plan curvature; **g** profile curvature; **h** distance to rivers; **i** distance to roads; **j** slope angle; **k** SPI; **l** TWI
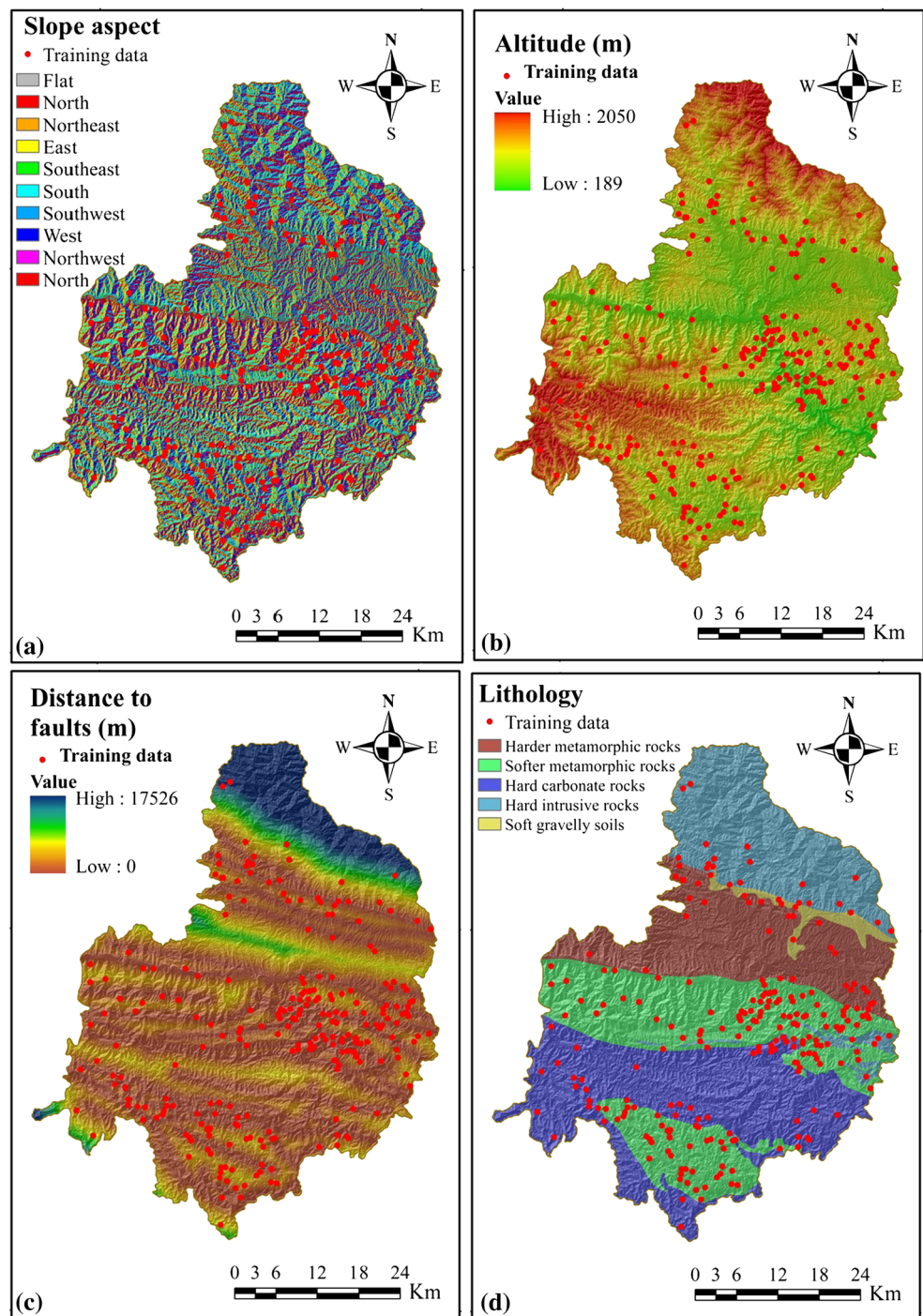
gravelly soils (Fig. 3d). The NDVI, as a conditioning factor, could be considered as a measure of surface reflectance and gives a quantitative estimate of biomass and the vegetation growth (Yilmaz 2009a, b; Pourghasemi et al. 2013a). In this research, an NDVI map was derived from the Landsat 7/ETM$^+$ images of Nov 4, 2014. The NDVI values vary from −0.23 to 0.71 in the present study area (Fig. 3e). Plan curvature

is the curvature of a contour line which is formed by intersecting a horizontal plane with the surface (Conforti et al. 2014). Plan curvature influences the divergence or convergence of water during downhill flow (Yilmaz et al. 2012). In this study, the plan curvature was extracted from the DEM with a spatial resolution of 30 m in ArcGIS 10.0. The values of plan curvature range from −9.57 to 11.42 (Fig. 3f). Profile curvature

Fig. 3 continued

(e) NDVI

(f) Plan curvature

(g) Profile curvature

(h) Distance to rivers (m)
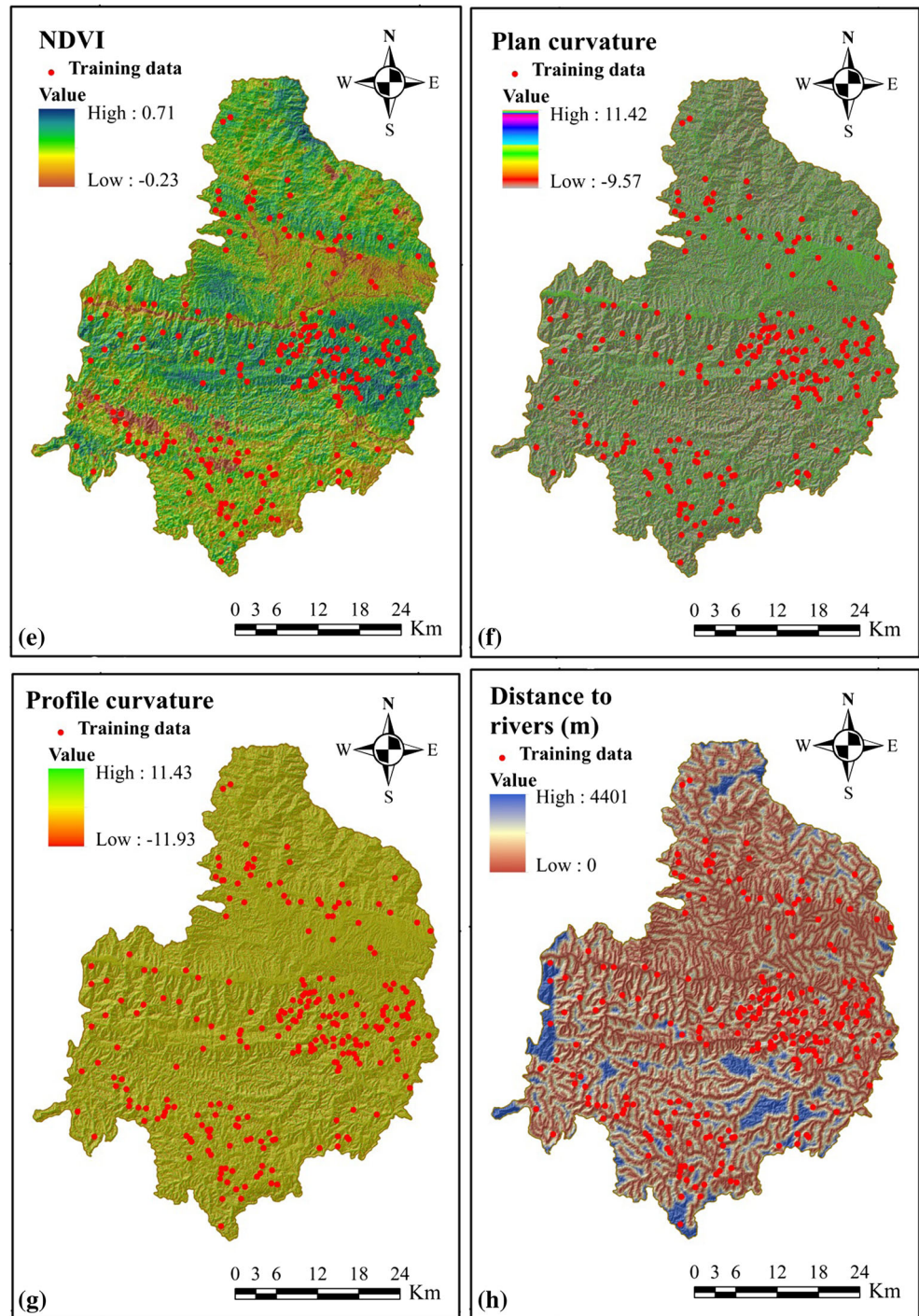
depicts the curvature in the vertical plane parallel to the slope direction. This factor measures the rate of change of slope (Kritikos and Davies 2014). Therefore, it influences the flow velocity of water draining the surface and thus erosion and the resulting down-slope movement of sediment (Yilmaz et al. 2012). In this study, the profile curvature values vary from −11.93 to 11.43 and were also derived from the DEM in ArcGIS

10.0 (Fig. 3g). Runoff plays an important role as a triggering factor for landslides due to rivers being the main mechanisms that contribute to the occurrence of landslides in mountainous regions (Park et al. 2013). For the current study, buffer zones were created to determine the degree to which the streams affected the slopes and range from 0 to 4401 m (Fig. 3h). The distance to roads has been considered as one of the

Fig. 3 continued
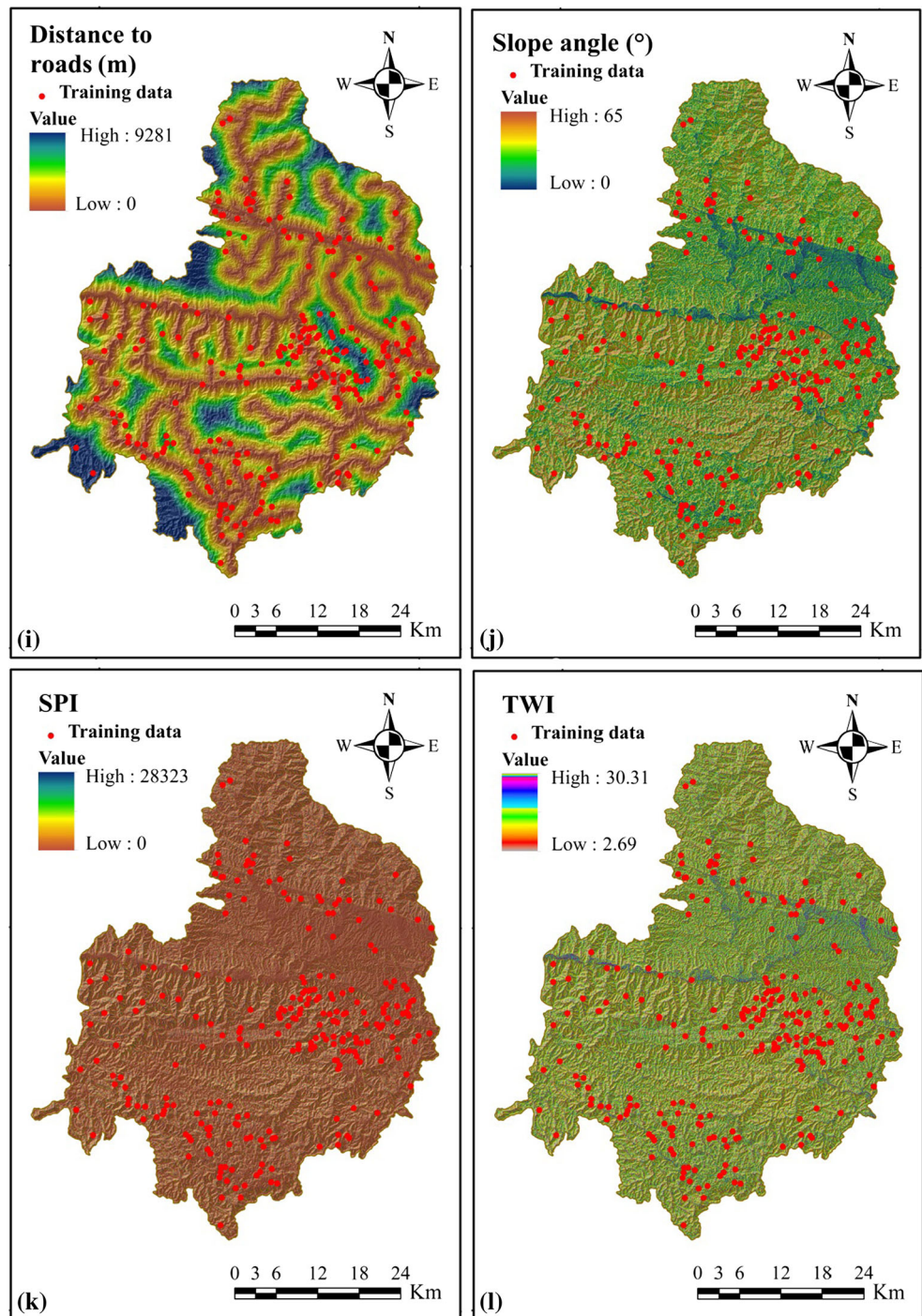
major anthropogenic factors influencing landslides occurrences (Demir et al. 2015). The creation of cut slopes during road construction disturbs the natural topology and affects the stability of the slope. In the study area, buffer zones were used to designate the influence of the road on slope stability and range from 0 to 9281 m (Fig. 3i). Slope angle is an important factor in assessment of slope instability, and it is

frequently used in landslide susceptibility mapping (Lee and Min 2001; Saha et al. 2005; Althuwaynee et al. 2014; Guo et al. 2015). The slope angle map of the study area was prepared from the DEM and ranges from 0° to 65° (Fig. 3j). The SPI, a measure of the erosive power of water flow, assuming that discharge is proportional to the specific catchment area, is a compound topographic attribute (Conforti et al. 2011). The

SPI values of the study area range from 0 to 28,323 (Fig. 3k). The TWI is another secondary topographic factor in the runoff model (Beven and Kirkby 1979; Moore et al. 1991). The TWI values were extracted from the DEM and vary from 2.69 to 30.31 (Fig. 3l).

Models

Frequency ratio (FR)

FR is based on the observed relationships between the distribution of the landslides and each landslide conditioning factor (Tay et al. 2014). The frequency ratios for the class or type of each conditioning factor were calculated by dividing the landslide occurrence ratio by the area ratio (Demir et al. 2013). Each factor's ratio value was calculated using Eq. (1):

$$FR = \frac{\text{Percentage of landslide}}{\text{Percentage of domain}}. \tag{1}$$

In this study, frequency ratio model was used to show the relationship between each factor's subclasses and landslide occurrences. A higher FR value shows that the probability of landslide occurrence is higher in the class.

Multivariate adaptive regression splines (MARS)

MARS could be implemented in order to fit the relationship between output and input variables (Friedman 1991; Naghibi and Moradi Dashtpagerdi 2016). The MARS model applies a non-parametric modeling approach that does not need assumptions about the relationship's form between the independent and dependent factors (Friedman 1991; Balshi et al. 2009). A MARS algorithm operates by splitting the ranges of the interpretive variables into regions and by generating a linear regression equation for each of the mentioned regions. Knots are the break values between regions, while the term bases function (BF) represents each distinct interval of the predictors. BFs are functions of the following form:

$$\max(0, x-k) \text{ or } \max(0, k-x), \tag{2}$$

in which x is an independent variable and k shows a constant corresponding to a knot. The general phrase of MARS can be obtained as follows:

$$\hat{y} = \hat{f}(x) = \beta + \sum_{m=1}^{M} \alpha_m H_m(x), \tag{3}$$

where $\hat{y}$ represents the dependent variable predicted by the function $f(x)$, β is a constant, M shows the number of terms, and each of them formed by a coefficient α_m, and $H_m(x)$ is

an individual basis function or a product of two or more BFs (Conoscenti et al. 2015).

MARS models were developed in two steps. In the first step, the forward algorithm and basis functions are introduced to introduce Eq. 3. Many basis functions are incorporated in Eq. 3 to get better performance. The developed MARS can indicate an over-fitting problem due to a large number of basic functions. Then, in the next step, a backward algorithm, for hindering over-fitting, redundant basis functions are omitted from Eq. 3. MARS uses generalized cross-validation (GCV) to remove the redundant basis functions (Craven and Wahba 1979). The expression of GCV is calculated as Eq. 4:

$$GCV = \frac{\frac{1}{N} \sum_{i=1}^{N} \left[y_i - \hat{f}(x_i) \right]^2}{\left[1 - \frac{C(H)}{N} \right]^2}, \tag{4}$$

where N is the number of data points and $C(K)$ corresponds to a complexity penalty that goes up with the number of BFs in the model and which is introduced as:

$$C(H) = (H + 1) + dH, \tag{5}$$

where d is a penalty for each BF covered in the model and H corresponds to the number of BFs in Eq. 2 (Friedman 1991).

Random forest (RF)

Classification trees are a non-linear technique for predicting a response by a set of binary decision rules that establish class assignment based on the predictors (Breiman et al. 1984). The RF multivariate statistical technique is a variation of the Bayesian tree or binary classification tree, because it considers a group (forest) of n trees to multiply the efficiency and predictive capability (Breiman 2001; Cutler et al. 2007; Trigila et al. 2015). The model was performed with the R statistical package (randomForest) which is suitable to cope with mixed variables, both categorical and numerical. Bootstrapping is a way of assessing the accuracy of a parameter estimate or a prediction. The RF uses the bagging technique (bootstrap aggregation) to choose, at each node of the tree, random samples of variables and observations as the training dataset for model training (Trigila et al. 2015). The random selection of the training dataset in an RF model may have effects on the results of the model; so, a set of numerous trees are used to guarantee the stability of the model. Unselected cases [out of the bag (OOB)] are used to compute the error of the model (OOB error), equal to the standard deviation error between predicted and observed values, and to establish a ranking of importance of the variables. Also, RF calculates two specific importance measures including mean decrease accuracy and mean

decrease Gini to show the importance and contribution of different factors in the modeling process. Because of the randomness of the RF technique at the extraction of independent variables and observations in each node, it is appropriate to preliminarily validate the effect of the number of trees and the number of runs on the stability of the model (Catani et al. 2013; Trigila et al. 2015).

Results

Application of the FR model

The results of spatial relationship between landslide locations and landslide conditioning factors using the FR model are shown in Table 1. From the FR values, it is seen that east-facing slopes followed by south-facing, northeast-facing, west-facing, and northeast-facing slopes are more susceptible to landsliding. In the case of altitude, results showed that there was no trend between landslide occurrence and altitude increase. The class of 500–800 m had the highest FR value of 1.46, followed by the classes of 189–500 m (1.29), and 800–1100 m (0.49). In the case of distance to faults, it can be seen that there is a reverse relationship between this factor and landslide occurrence. The distance from faults of 0–1000 m and 1000–2000 m had the highest FR values of 1.33 and 0.98, respectively. For the lithology, it is seen that most landslides occurred in lithological units of softer metamorphic rocks with an FR value of 1.94, followed by the lithological unit of soft gravelly soils with an FR value of 1.43. In the case of NDVI, the class −0.23 to 0.17 had the highest FR value of 1.31, indicating that this class is the most susceptible to landsliding. In the case of plan curvature, classes of −0.11 to 0.88 and −1.09 to −0.11 had the highest FR values of 1.02 and 1.01, respectively. For the profile curvature, the classes of 1.26–11.43 had the highest FR value (1.08), followed by the classes of −0.02 to 1.26 (1.04) and −1.30 to −0.02 (0.98). From the analysis of distance to rivers, it can be seen that there is a reverse relationship between this factor and landslide occurrence. In the case of distance to roads, distance to roads of 0–500 had the highest FR value of 1.72, indicating that the roads have a high influence on landslide occurrence. In the case of slope angle, the classes of 0–10 and 10–20 had the highest FR values of 1.22, and 1.18, respectively. While other slope classes have FR values lower than 1. For the SPI, the results showed that the class of 0–30 had the highest FR value (1.08), followed by 60–90 (1.00), 30–60 (0.93), and >90 (0.87). In the case of TWI, the results showed that the class of 7–9 had the highest FR value (1.15), followed by 5–7 (1.09), <5 (0.88), and >9 (0.77), which indicates that there is no clear relationship between this factor and landslide occurrence.

Table 2 The distribution of the landslide values and areas with respect to the landslide occurrence potential zones

| Factor	nsubset	GCV	RSS
Slope aspect 5	10	22.60	48.70
Slope aspect 7	11	28.30	52.60
Slope aspect 8	5	15.90	34.10
Altitude (m)	17	100.00	100.00
Distance to faults (m)	17	100.00	100.00
Lithology 2	16	72.40	81.80
Lithology 3	6	17.80	37.50
Lithology 5	9	18.20	45.30
NDVI	15	54.50	70.80
Plan curvature (100/m)	4	12.80	30.10
Profile curvature (100/m)	3	9.50	25.70
Distance to rivers (m)	14	46.60	65.40
Distance to roads (m)	14	46.60	65.40
Slope angle (degree)	12	32.90	56.20
SPI	16	72.4	81.80
TWI	8	21.30	43.70

Application of the MARS model

The optimal model presents 18 terms and includes 18 BF (the terms created during the forward pass were 86), with a GCV of 0.19. Only 12 of the 17 independent variables were used in the optimal model (Table 2), because MARS only uses the essential independent variables (Gutiérrez et al. 2009). Table 2 presents the nsubset is the index vector specifying which cases to implement, i.e., which rows in x to implement (default is NULL, meaning all), the GCV of the model (summed over all responses; the GCV is calculated using the penalty argument) and the residual sum-of-squares (RSS) of the model (summed over all responses if y has multiple columns). So, based on Table 2, regarding the importance of independent variables to support the model, the most significant are altitude and distance to faults. Other important factors to describe the spatial distribution of landslides in the study area are lithology, SPI, and NDVI. In this kind of model, the importance of the independent variables should be perceived with caution (Donati and Turrini 2002; Gutiérrez et al. 2009). Finally, the landslide potential map produced by the MARS model is presented in Fig. 6.

Application of the RF model

Aggregation of the OOB prediction is presented in Table 3 (confusion matrix). The OOB represents that when the resulting model is applied to new data, the answer will be in error 25% of the time. A 75% accuracy is indicated, which is a reasonably good model. The overall measure of

Table 3 Confusion matrix from the RF model (0 = no landslide, 1 = landslide)

	0	1	Class error
0	143	101	0.41
1	67	177	0.27
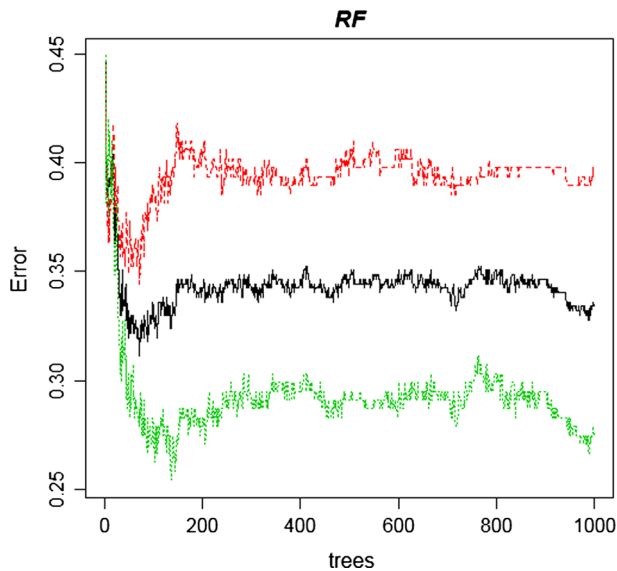
Fig. 4 The error rate of the overall RF model [OOB: out of bag (*black line*), 0: absent landslide (*red line*) and 1: present landslide (*green line*)]

accuracy is then followed by a confusion matrix that records the conflict between that final model's predictions and the present outcomes of the training observations. The present observations are the rows of Table 3, whereas the columns corresponded to what the model predicts for an observations and the cell count the number of observations in each variable (Williams 2011). The results of the confusion matrix showed that the model predicted 143 non-

landslides as non-landslides and 101 non-landslides as landslides. On the other hand, RF models predicted 177 landslides as landslide and 67 landslides as non-landslides. Moreover, the OOB estimate of the error rate for this model was 34.43% which indicates that the model is 65.57% accurate, which is a reasonably good model (Fig. 4). Results from variable selection in the RF model are represented in Table 4 and Fig. 5. The higher values show that the variable is relatively more important (Williams 2011). The accuracy measure (mean decrease) then lists lithology (23.47), distance to faults (22.21), altitude (19.58) as next most important. We also notice that altitude (19.04), distance to faults (18.83), and distance to roads (15.29) have higher importance according to the Gini measure than with the accuracy measure (Fig. 6).

Finally, landslide susceptibility mapping by the RF model was prepared in ArcGIS 10.0 and grouped using a natural break classification (Pourghasemi et al. 2013c; Zare et al. 2013; Naghibi et al. 2015, 2016) into low, moderate, high, and very high potential groups (Fig. 7).

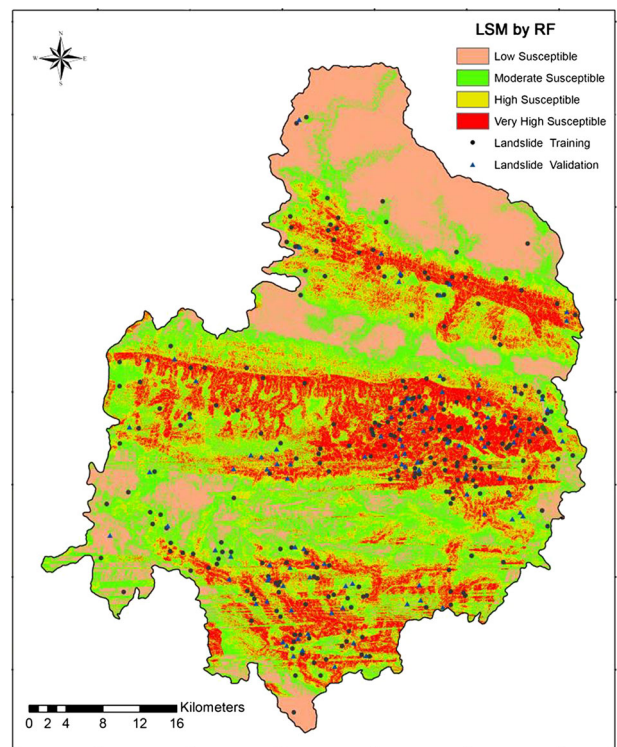## Validation of the landslide susceptibility maps

A proper method is required to verify the landslide susceptibility maps. In the present study, the produced landslide susceptibility maps were validated using the ROC curve. This method has been widely used in landslide susceptibility mapping to compare the different models (Guo et al. 2015; Youssef et al. 2015a). In the ROC curve, the true positive rate of the model (the percentage of existing landslides correctly predicted by the model) is plotted against the false positive rate (the percentage of predicted landslides out of the total actual negatives). The area under the curve (AUC) reflects the predictive ability of the models. The AUC values range from 0.5 to 1. The ideal model has an AUC value close to 1.0 (perfect fit), while a value close to 0.5 indicates a random fit.

Table 4 Relative influence of effective conditioning factors in the RF model (0 = no landslide, 1 = landslide)
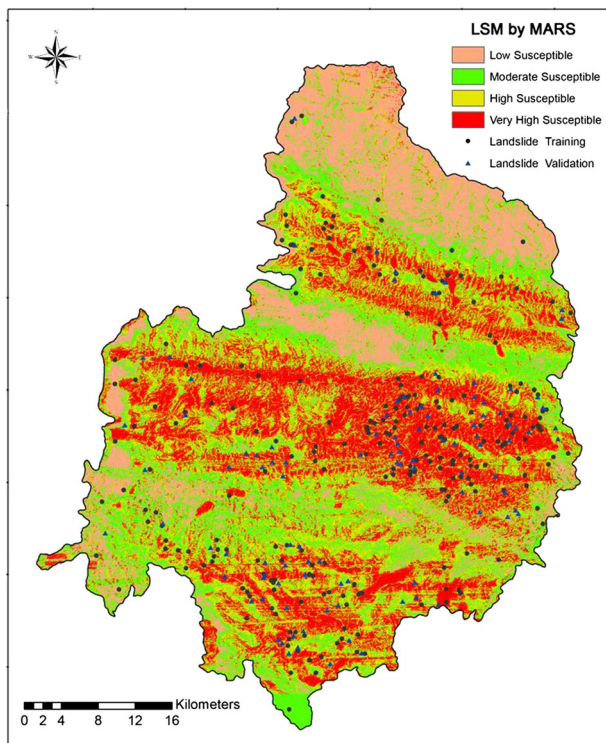
Factor	0	1	Mean decrease accuracy	Mean decrease Gini
Slope aspect	−1.76	−0.26	−1.47	11.50
Altitude (m)	10.07	18.54	19.58	19.04
Distance to faults (m)	11.85	19.37	22.21	18.83
Lithology	19.24	14.58	23.47	11.16
NDVI	−3.44	−1.19	−3.35	10.24
Plan curvature (100/m)	−3.85	1.55	−1.55	10.02
Profile curvature (100/m)	−1.59	4.31	2.29	11.09
Distance to rivers (m)	5.06	7.06	8.77	13.67
Distance to roads (m)	16.52	−0.09	12.00	15.29
Slope angle (degree)	0.51	3.42	2.78	11.48
SPI	2.07	4.91	5.23	12.04
TWI	−1.41	3.17	1.49	9.96
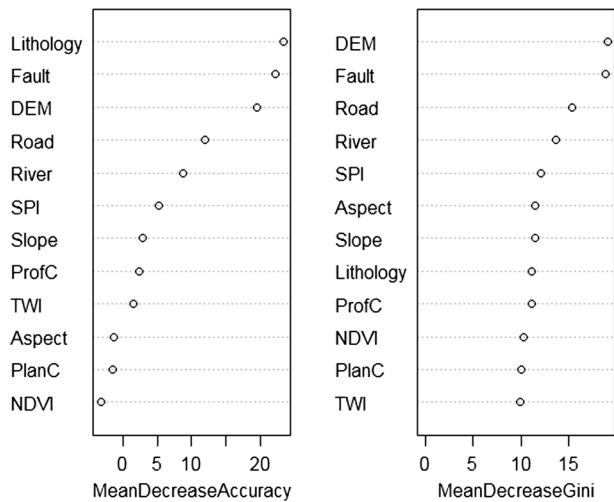
Fig. 5 The error rate of the overall RF model (OOB: out of bag, 0: absent landslide and 1: present landslide)

Fig. 6 Landslide potential map produced by multivariate adaptive regression spline (MARS) model

Generally, there are two methods used for evaluating a model, such as success rate and prediction rate curves. Pourghasemi et al. (2012a, b) depicted that the success rate method used the training landslide points that have been used for building the landslide models, so it is not a proper method for assessing the prediction capability of the models. Thus, in the present study, the validation analysis

Fig. 7 Landslide potential map produced by the random forest (RF) model

was performed using the remaining 104 landslide data points (validation dataset). Figure 8 shows the ROC curves of the MARS and RF models for the validation data set. The ROC plot showed that in the landslide susceptibility map prepared using the MARS method, the AUC value is 87.51%, while in the RF model, the AUC value is 77.31%. So, the results indicate that the landslide susceptibility map obtained by the MARS model has higher prediction capability than the RF model.
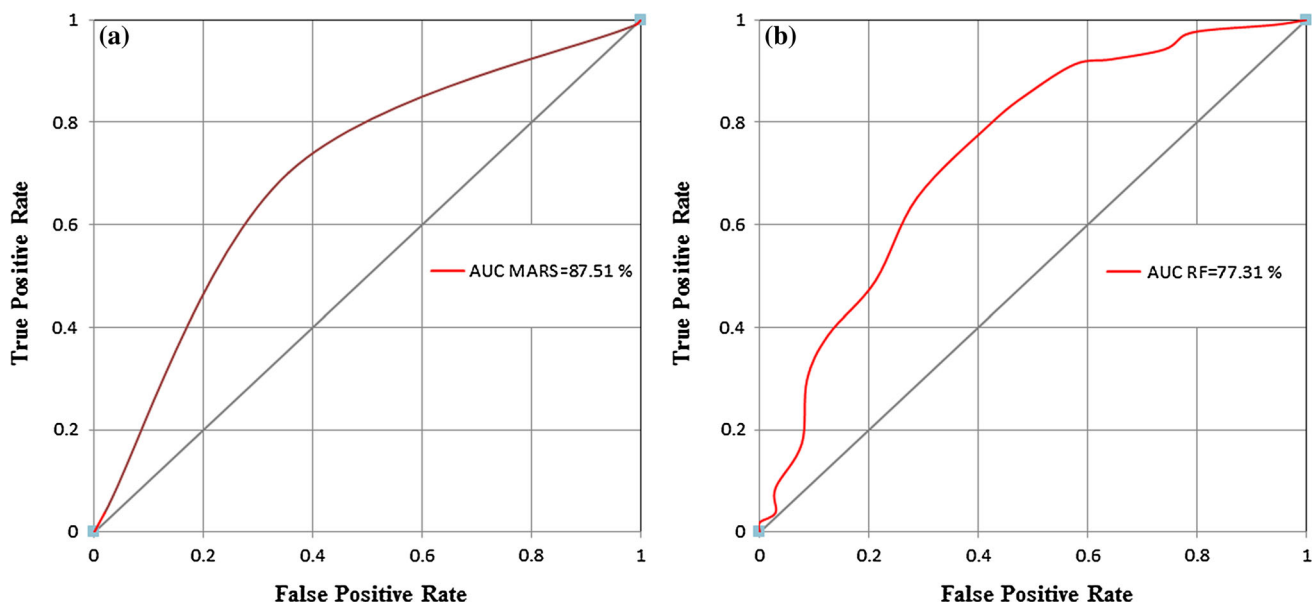
Discussion

In the current study, the results are represented and discussed in two important sections, including the performance of the models and the importance of the conditioning factors in landslide susceptibility mapping.

The performance of the models

Application and comparison of the MARS and RF models is relatively new in landslide susceptibility mapping studies. The results of the current study showed that the MARS model had a relatively higher prediction performance than the RF model. According to the literature review, these models had acceptable performance in different fields of study, including landslide, gully, and groundwater studies.

Fig. 8 Prediction rate curves for the landslide potential maps produced by **a** RF and **b** MARS models

In research, Conoscenti et al. (2015) determined terrain susceptibility to earth-flow using logistic regression and MARS models. They found that the MARS model performed better than logistic regression considering both the goodness-of-fit and the predictive power according to ROC values. They also indicated that MARS reproduces non-linear relationships using several linear regressions, which allows MARS to create models with a better fit to the training data while maintaining high predictive power. In other research, Adoko et al. (2013) concluded that MARS can constitute a reliable alternative to an ANN in modeling geo-engineering problems such as tunnel convergence. Also, Gutiérrez et al. (2009) compared classification and regression tree (CART) and MARS models, non-parametric models, to map the potential distribution of gullies. They found that MARS exhibit a better performance for predicting gullying with areas under the ROC curve values of 0.98 and 0.97 for the validation datasets, while CART presented values of 0.96 and 0.66. Trigila et al. (2015) carried out a comparison study of shallow landslide susceptibility mapping using LR and RF models. They concluded that the LR and RF methods are fully comparable in terms of identification of the most significant variables and predictive capabilities, and the RF technique is better for the minimization of false positives. Youssef et al. (2015b) produced landslide susceptibility maps using four data mining models, namely RF, CART, a boosted regression tree (BRT), and a general linear model (GLM). The success and prediction rate curves showed that the RF model had satisfactory performance. Goetz et al. (2015) applied logistic regression, generalized additive models, weights-

of-evidence, support vector machine, RF, and bootstrap aggregated classification trees with penalized discriminant analysis in landslide susceptibility modeling for three areas in the Province of Lower Austria, Austria. They concluded that in terms of pure prediction performance, the RF and bootstrap aggregated classification trees with penalized discriminant analysis modeling techniques were the best.

Also, other researchers used RF in the field of groundwater potential mapping and got acceptable results. Rahmati et al. (2016) applied maximum entropy (ME) and RF models for groundwater potential mapping in the Mehran Region, Iran. They found that the AUC for prediction performance of the RF and ME methods was calculated as 83.1 and 87.7%, respectively. In other research, Naghibi and Pourghasemi (2015) evaluated the capability of three machine learning models such as BRT, CART, and RF, and comparison of their performance by multivariate GLM, and bivariate [evidential belief function (EBF)] statistical methods in the groundwater potential mapping. They concluded that CART, BRT, and RF machine learning techniques showed acceptable performance in mapping groundwater potential with the AUC values of 86.39, 86.12, and 86.05%, respectively. Rahmati et al. (2016) applied RF and maximum entropy (ME) models for groundwater potential mapping in the Mehran Region, Iran. Their results depicted that RF and ME models had AUC for success rates of 86.5 and 91%, respectively, while the AUC for prediction rates of RF and ME methods were 83.1 and 87.7%, respectively. Comparing the results of this study and other studies showed that RF and MARS with AUC values of 77.32 and 87.51%, respectively, have

outperformed other models such as EBF (AUC = 0.73) and FR (AUC = 0.77) models in Hong et al. (2016), weights of evidence (AUC = 0.74) in Mohammady et al. (2012), logistic regression in Chen et al. (2016c), naïve Bayes (AUC = 0.81 which is lower than RF in the current study) in Tien Bui et al. (2012), and Dempster Shafer (AUC = 0.72) in Pourghasemi et al. (2013a). In Naghibi et al. (2016), RF (AUC = 0.71) showed weak performance in the modeling process as well as the results of the current study depicted. Although RF performed weaker than MARS in this study, it can be concluded that RF and MARS models could provide better landslide susceptibility maps than the mentioned models.

In the case of landslide inventory accuracy, two methods were considered to increase the accuracy of the inventory, field work and using satellite images. It is mentioned by Malamud et al. (2004) that determining the level of completeness of an inventory map is not a simple task. Therefore, work on this subject could be suggested for future works as inventory accuracy highly affects the results of the models and landslide susceptibility maps.

The importance of the conditioning factors in landslide susceptibility mapping

Determining the importance of conditioning factors in landslide susceptibility mapping is important in landslide studies. The study revealed that the most important conditioning factors according to the accuracy measure (mean decrease) of an RF model are lithology (23.47%), distance to faults (22.21%), and altitude (19.58%). We also notice that altitude (19.04%), distance to faults (18.83%), and distance to roads (15.29%) have higher importance according to the Gini measure than with the accuracy measure. Also, Kawabata and Bandibas (2009) introduced geology as the most important factor in landslide susceptibility mapping. In other research, Youssef et al. (2015a) concluded that slope angle, land use, and altitude have higher importance in landslide occurrence, which is consistent with the results of current study. In other research, lithology and altitude were the most important factors in landslide susceptibility mapping (Devkota et al. 2013).

Conclusions

Landslide spatial modeling with more advanced models such as MARS and RF can give handy landslide susceptibility maps, and consideration of these maps by different organizations can lessen severe damages caused by landslides. This study was conducted to compare the spatial capability of the MARS and RF models to produce landslide susceptibility maps and determine areas prone to landslide occurrence in Shangnan County, Shangluo City, China. For this purpose, 12 landslide effective factors were selected to be used as model inputs: slope aspect, altitude, distance to faults, lithology, NDVI, plan curvature, TWI, SPI, profile curvature, distance to rivers, distance to roads, and slope angle. The location of the landslides was detected from historical reports, interpretation of aerial photographs, and extensive field surveys by GPS. From these locations, 70% (244) and 30% (104) of the landslides were used for training and validation of the MARS and RF models, respectively. Finally, the ROC curve was used to validate these results. The AUC results showed that the prediction rate of the MARS model is higher than the RF model. Further, it was inferred that the most important conditioning factors according to the accuracy measure (mean decrease) of the RF model are lithology, distance to faults, and altitude. We also noticed that altitude, distance to faults, and distance to roads have higher importance according to the Gini measure than with the accuracy measure. The produced landslide susceptibility maps could be appreciated as a useful tool for local authorities and government agencies in choosing suitable locations for future land-use planning. As a final conclusion, MARS was applied successfully and its application in landslide-prone areas could be proposed in other areas and regional scale.

References

Adoko AC, Jiao YY, Wu L, Wang H, Wang ZH (2013) Predicting tunnel convergence using multivariate adaptive regression spline and artificial neural network. Tunn Undergr Space Technol 38:368–376

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 9:93–106

Akgun A, Erkan O (2016) Landslide susceptibility mapping by geographical information system-based multivariate statistical and deterministic models: in an artificial reservoir area at Northern Turkey. Arab J Geosci 9:1–15

Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bull Eng Geol Environ 58(1):21–44

Althuwaynee OF, Pradhan B, Park HJ, Lee JH (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

Armaş I, Vartolomei F, Stroia F, Braşoveanu L (2014) Landslide susceptibility deterministic approach using geographic information systems: application to Breaza town, Romania. Nat Hazards 70:995–1017

Baeza C, Corominas J (2001) Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. Earth Surf Proc Land 26:1251–1263

Balshi MS, Mcguire AD, Duffy P, Flannigan M, Walsh J, Melillo J (2009) Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. Glob Change Biol 15:578–600

Beven K, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci Bull 24:43–69

Breiman L (2001) Random forests. Mach Learn 45:5–32

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees Belmont. Wadsworth International Group, CA

Catani F, Lagomarsino S, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Nat Hazards Earth Syst Sci 13:2815–2831

Cervi F, Berti M, Borgatti L, Ronchetti F, Manenti F, Corsini A (2010) Comparing predictive capability of statistical and deterministic methods for landslide susceptibility mapping: a case study in the northern Apennines (Reggio Emilia Province, Italy). Landslides 7:433–444

Chen W, Li W, Hou E, Zhao Z, Deng N, Bai H, Wang D (2014) Landslide susceptibility mapping based on GIS and information value model for the Chencang District of Baoji, China. Arab J Geosci 7:4499–4511

Chen W, Li W, Hou E, Bai H, Chai H, Wang D, Cui X, Wang Q (2015) Application of frequency ratio, statistical index, and index of entropy models and their comparison in landslide susceptibility mapping for the Baozhong Region of Baoji, China. Arab J Geosci 8:1829–1841

Chen T, Niu R, Jia X (2016a) A comparison of information value and logistic regression models in landslide susceptibility mapping by using GIS. Environ Earth Sci 75(10):867

Chen W, Li W, Chai H, Hou E, Li X, Ding X (2016b) GIS-based landslide susceptibility mapping using analytical hierarchy process (AHP) and certainty factor (CF) models for the Baozhong region of Baoji City, China. Environ Earth Sci 75:1–14

Chen W, Pourghasemi HR, Zhao Z (2016c) A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. Geocarto Int. doi:10.1080/10106049.2016.1140824

Chen W, Wang J, Xie X, Hong H, Trung NV, Bui DT, Wang G, Li X (2016d) Spatial prediction of landslide susceptibility using integrated frequency ratio with entropy and support vector machines by different kernel functions. Environ Earth Sci 75:1344. doi:10.1007/s12665-016-6162-8

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. CATENA 151:147–160

Clerici A, Perego S, Tellini C, Vescovi P (2006) A GIS-based automated procedure for landslide susceptibility mapping by the conditional analysis method: the Baganza valley case study (Italian Northern Apennines). Environ Geol 50:941–961

Conforti M, Aucelli PP, Robustelli G, Scarciglia F (2011) Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). Nat Hazards 56(3):881–898

Conforti M, Pascale S, Robustelli G, Sdao F (2014) Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena 113:236–250

Conoscenti C, Ciaccio M, Caraballo-Arias NA, Gómez-Gutiérrez Á, Rotigliano E, Agnesi V (2015) Assessment of susceptibility to

earth-flow landslide using logistic regression and multivariate adaptive regression splines: a case of the Belice River basin (western Sicily, Italy). Geomorphology 242:49–64

Craven P, Wahba G (1979) Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized crossvalidation. Numer Math 31:377–403

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88:2783

Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K (2008) GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. Environ Geol 54(2):311–324

Demir G, Aytekin M, Akgün A, Ikizler SB, Tatar O (2013) A comparison of landslide susceptibility mapping of the eastern part of the North Anatolian Fault Zone (Turkey) by likelihood-frequency ratio and analytic hierarchy process methods. Nat Hazards 65(3):1481–1506

Demir G, Aytekin M, Akgun A (2015) Landslide susceptibility mapping by frequency ratio and logistic regression methods: an example from Niksar-Resadiye (Tokat, Turkey). Arab J Geosci 8(3):1801–1812

Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC, Dhital MR, Althuwaynee OF (2013) Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya. Nat Hazards 65(1):135–165

Ding Q, Chen W, Hong H (2016) Application of frequency ratio, weights of evidence and evidential belief function models in landslide susceptibility mapping. Geocarto Int. doi:10.1080/10106049.2016.1165294

Donati L, Turrini MC (2002) An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy). Eng Geol 63:277–289

Dou J, Oguchi T, Hayakawa YS, Uchiyama S, Saito H, Paudel U (2014) GIS-based landslide susceptibility mapping using a certainty factor model and its validation in the Chuetsu Area, Central Japan. Landslide Science for a Safer Geoenvironment. Springer, Switzerland, pp 419–424

Ercanoglu M, Gokceoglu C (2002) Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach. Environ Geol 41(6):720–730

Ercanoglu M, Gokceoglu C, Van Asch TW (2004) Landslide susceptibility zoning north of Yenice (NW Turkey) by multivariate statistical techniques. Nat Hazards 32:1–23

Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage W (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Eng Geol 102:85–98

Friedman JH (1991) Multivariate adaptive regression spline. Ann Stat 19:1–67

Galli M, Ardizzone F, Cardinali M, Guzzetti F, Reichenbach P (2008) Comparing landslide inventory maps. Geomorphology 94(3–4):268–289. doi:10.1016/j.geomorph.2006.09.023

García-Rodríguez MJ, Malpica JA, Benito B, Díaz M (2008) Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. Geomorphology 95:172–191

Goetz JN, Brenning A, Petschko H, Leopold P (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput Geosci 81:1–11

Gökceoglu C, Aksoy H (1996) Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. Eng Geol 44:147–161

Gorsevski PV, Brown MK, Panter K, Onasch CM (2016) Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: a case study in the Cuyahoga Valley National Park, Ohio. Landslides 13:467–484

Grozavu A, Plescan S, Patriche CV, Margarint MC, Rosca B (2013) Landslide susceptibility assessment: GIS application to a complex mountainous environment. Integrating Nature and Society Towards Sustainability, Environ Science and Engineering, The Carpathians, pp 31–44

Guettouche MS (2013) Modeling and risk assessment of landslides using fuzzy logic. Application on the slopes of the Algerian Tell (Algeria). Arab J Geosci 6:3163–3173

Guillard C, Zezere J (2012) Landslide susceptibility assessment and validation in the framework of municipal planning in Portugal: the case of Loures Municipality. Environ Manag 50:721–735

Guo C, Montgomery DR, Zhang Y, Wang K, Yang Z (2015) Quantitative assessment of landslide susceptibility along the Xianshuihe fault zone, Tibetan Plateau, China. Geomorph 248:93–110

Gutiérrez ÁG, Schnabel S, Contador JFL (2009) Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. Ecol Modell 220(24):3630–3637

Hong H, Pradhan B, Xu C, Tien Bui D (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena 133:266–281

Hong H, Naghibi SA, Pourghasemi HR, Pradhan B (2016) GIS-based landslide spatial modeling in Ganzhou City, China. Arab J Geosci 9(2):112. doi:10.1007/s12517-015-2094-y

Jaafari A, Najafi A, Pourghasemi H, Rezaeian J, 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 11:909–926

Jaiswal P, Srinivasan P, Venkatraman NV (2013) A data-guided heuristic approach for landslide susceptibility mapping along a transportation corridor in the Nilgiri Hills, Nilgiri District, Tamil Nadu. Indian J Geosci 67(3):273–288

Jia N, Mitani Y, Xie M, Djamaluddin I (2012) Shallow landslide hazard assessment using a three-dimensional deterministic model in a mountainous area. Comput Geotech 45:1–10

Kanungo D, Sarkar S, Sharma S (2011) Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. Nat Hazards 59:1491–1512

Karimi Sangchini EK, Emami SN, Tahmasebipour N, Pourghasemi HR, Naghibi SA, Arami SA, Pradhan B (2016) Assessment and comparison of combined bivariate and AHP models with logistic regression for landslide susceptibility mapping in the Chaharmahal-e-Bakhtiari Province, Iran. Arab J Geosci 9(3):201

Kawabata D, Bandibas J (2009) Landslide susceptibility mapping using geological data, a DEM from ASTER images and an artificial neural network (ANN). Geomorphology 113:97–109

Kayastha P, Dhital MR, De Smedt F (2013) Application of analytical hierarchy process (AHP) for landslidesusceptibility mapping: a case study from the Tinau watershed, west Nepal. Comput Geosci 52:398–408

Kritikos T, Davies T (2014) Assessment of rainfall-generated shallow landslide/debris-flow susceptibility and runout using a GIS-based approach: application to western Southern Alps of New Zealand. Landslides 12(6):1051–1075

Kumar R, Anbalagan R (2015) Landslide susceptibility zonation in part of Tehri reservoir region using frequency ratio, fuzzy logic and GIS. J Earth Syst Sci 124(2):431–448

Lee S, Min K (2001) Statistical analysis of landslide susceptibility at Youngin, Korea. Environ Geol 40:1095–1113

Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. Environ Geol 47:982–990

Lee MJ, Park I, Lee S (2015) Forecasting and validation of landslide susceptibility using an integration of frequency ratio and neuro-fuzzy models: a case study of Seorak mountain area in Korea. Environ Earth Sci 74:413–429

Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P (2004) Landslide inventories and their statistical properties. Earth Surf Process Landf 29:687–711

Mandal S, Maiti R (2015) Application of analytical hierarchy process (AHP) and frequency ratio (FR) model in assessing landslide susceptibility and risk. Springer, Singapore

Marjanović M, Kovačević M, Bajat B, Voženílek V (2011) Landslide susceptibility assessment using SVM machine learning algorithm. Eng Geol 123:225–234

Mihaela C, Martin B, Marta CJ, Marius V (2011) Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). Environ Earth Sci 63:397–406

Mohammady M, Pourghasemi HR, 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

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

Naghibi SA, Moradi Dashtpagerdi M (2016) Evaluation of four supervised learning methods for groundwater spring potential mapping in Khalkhal region (Iran) using GIS-based features. Hydrogeol J. doi:10.1007/s10040-016-1466-z

Naghibi SA, Pourghasemi HR (2015) A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. Water Resour Manag 29(14):5217–5236

Naghibi SA, Pourghasemi HR, Pourtaghi ZS, Rezaei A (2015) Ground water qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran. Earth Sci Inform 8(1):171–186

Naghibi SA, Pourghasemi HR, Dixon B (2016) GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ Monit Assess 188:1–27

Nourani V, Pradhan B, Ghaffari H, Sharifi SS (2014) Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. Nat Hazards 71(1):523–547

Ozdemir A, Altural T (2013) A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. J Asian Earth Sci 64:180–197

Park NW, Chi KH (2008) Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model. Int J Remote Sens 29:247–264

Park S, Choi C, Kim B, Kim J (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. Environ Earth Sci 68:1443–1464

Paudel U, Oguchi T, Hayakawa Y (2016) Multi-resolution landslide susceptibility analysis using a DEM and random forest. Int J Geosci 07:726–743

Peng L, Niu R, Huang B, Wu X, Zhao Y, Ye R (2014) Landslide susceptibility mapping based on rough set theory and support vector machines: a case of the three Gorges area, China. Geomorph 204:287–301

Petschko H, Bell R, Brenning A, Glade T (2012) Landslide susceptibility modeling with generalized additive models–facing the heterogeneity of large regions. Landslides Eng Slopes Prot Soc Through Imp Underst 1:769–777

Poudyal CP, Chang C, Oh HJ, Lee S (2010) Landslide susceptibility maps comparing frequency ratio and artificial neural networks: a case study from the Nepal Himalaya. Environ Earth Sci 61(5):1049–1064

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

Pourghasemi HR, Mohammady M, Pradhan B (2012a) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin. Iran. Catena 97:71–84

Pourghasemi HR, Pradhan B, Gokceoglu C (2012b) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat Hazards 63:965–996

Pourghasemi HR, Pradhan B, Gokceoglu C (2012c) Remote sensing data derived parameters and its use in landslide susceptibility assessment using Shannon's Entropy and GIS. Appl Mech Mater 225:486–491

Pourghasemi HR, Goli Jirandeh A, Pradhan B, Xu C, Gokceoglu C (2013a) Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. J Earth Syst Sci 122(2):349–369

Pourghasemi HR, Moradi HR, Fatemi Aghda SM (2013b) Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69:749–779

Pourghasemi HR, Pradhan B, Gokceoglu C, Moezzi KD (2013c) A comparative assessment of prediction capabilities of Dempster-Shafer and weights-of-evidence models in landslide susceptibility mapping using GIS. Geomat Nat Hazards Risk 4(2):93–118

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

Pourghasemi HR, Moradi HR, Fatemi Aghda SM, Gokceoglu C, Pradhan B (2014) GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). Arab J Geosci 7(5):1857–1878

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci 51:350–365

Pradhan B, Lee S (2010) Landslide susceptibility assessment and factor effect analysis: back-propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling. Environ Model Softw 25(6):747–759

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

Regmi AD, Devkota KC, Yoshida K, Pradhan B, Pourghasemi HR, Kumamoto T, Akgun A (2014a) 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

Regmi AD, Yoshida K, Pourghasemi HR, Dhital MR, Pradhan B (2014b) Landslide susceptibility mapping along Bhalubang-Shiwapur area of mid-western Nepal using frequency ratio and conditional probability models. J Mt Sci 11(5):1266–1285

Ruff M, Czurda K (2008) Landslide susceptibility analysis with a heuristic approach in the Eastern Alps (Vorarlberg, Austria). Geomorphology 94:314–324

Saha AK, Gupta RP, Sarkar I, Arora MK, Csaplovics E (2005) An approach for GIS-based statistical landslide susceptibility zonation with a case study in the Himalayas. Landslides 2:61–69

Talaei R (2014) Landslide susceptibility zonation mapping using logistic regression and its validation in Hashtchin Region, northwest of Iran. J Geol Soc India 84:68–86

Tay LT, Lateh H, Hossain MK, Kamil AA (2014) Landslide hazard mapping using a poisson distribution: a case study in Penang Island, Malaysia. Landslide science for a safer geoenvironment. Springer, Switzerland, pp 521–525

Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and naive bayes models. Math Probl Eng. doi:10.1155/2012/974638

Tien Bui D, Tuan TA, Klempe H, Pradhan B, Revhaug I (2015) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides. doi:10.1007/s10346-015-0557-6

Trigila A, Iadanza C, Esposito C, Scarascia-Mugnozza G (2015) Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). Geomorph 249:119–136

Tsangaratos P, Benardos A (2014) Estimating landslide susceptibility through a artificial neural network classifier. Natural Hazards 74:1489–1516

Tseng C, Lin C, Hsieh W (2015) Landslide susceptibility analysis by means of event-based multi-temporal landslide inventories. Nat Hazards Earth Syst Sci Discuss 3(2):1137–1173

Van Westen C, Terlien M (1996) An approach towards deterministic landslide hazard analysis in GIS. A case study from Manizales (Colombia). Earth Surf Process Landf 21:853–868

Van Westen CJ, Van Asch TW, Soeter R (2006) Landslide hazard and risk zonation—why is it still so difficult? Bull Eng Geol Environ 65(2):167–184

Vijith H, Madhu G (2008) Estimating potential landslide sites of anupland sub-atershed in Western Ghat's of Kerala (India) through frequency ratio and GIS. Environ Geol 55:1397–1405

Williams G (2011) Random forests. Springer, New York

Xu C, Dai FC, Xu X, Lee YH (2012a) GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. Geomorph 145–146:70–80

Xu C, Xu XW, Dai FC, Saraf AK (2012b) Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. Comput Geosci 46:317–329

Xu C, Xu X, Dai F, Wu Z, He H, Shi F, Wu X, Xu S (2013) Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake of China. Nat Hazards 68:883–900

Yilmaz I (2009a) A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks. Bull Eng Geol Environ 68(3):297–306

Yilmaz I (2009b) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat-Turkey). Comput Geosci 35(6):1125–1138

Yilmaz I (2010a) The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability (CP)

and artificial neural network (ANN). Environ Earth Sci 60:505–519

Yilmaz I (2010b) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environ Earth Sci 61:821–836

Yilmaz C, Topal T, Suzen ML (2012) GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak-Turkey). Environ Earth Sci 65:2161–2178

Youssef AM, Pradhan B, Jebur MN, El-Harbi HM (2014) Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia. Enviro Earth Sci 73(7):3745–3761

Youssef AM, Pradhan B, Pourghasemi HR, Abdullahi S (2015a) Landslide susceptibility assessment at Wadi Jawrah Basin, Jizan region, Saudi Arabia using two bivariate models in GIS. Geosci J 19(3):449–469

Youssef AM, Pourghasemi HR, Pourtaghi ZS, Al-Katheeri MM (2015b) Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides. doi:10.1007/s10346-015-0614-1

Zare M, Pourghasemi HR, Vafakhah M, Pradhan B (2013) Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multi-layer perceptron (MLP) and radial basic function (RBF) algorithms. Arab J Geosci 6:2873–2888

Zhu AX, Wang RX, Qiao JP, Qin CZ, Chen YB, Liu J, Du F, Lin Y, Zhu TX (2014) An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic. Geomorphology 214:128–138