Arab J Geosci (2017) 10: 187
DOI 10.1007/s12517-017-2961-9

ORIGINAL PAPER

A comparative study on the landslide susceptibility mapping using logistic regression and statistical index models

Zhiyong Wu[1,2] · **Yanli Wu**[1,3] · **Yitian Yang**[2] · **Fuwei Chen**[4] · **Na Zhang**[2] · **Yutian Ke**[5] · **Wenping Li**[1]

Received: 18 April 2016 / Accepted: 17 March 2017 / Published online: 12 April 2017

Abstract The logistic regression and statistical index models are applied and verified for landslide susceptibility mapping in Daguan County, Yunnan Province, China, by means of the geographic information system (GIS). A detailed landslide inventory map was prepared by literatures, aerial photographs, and supported by field works. Fifteen landslide-conditioning factors were considered: slope angle, slope aspect, curvature, plan curvature, profile curvature, altitude, STI, SPI, and TWI were derived from digital elevation model; NDVI was extracted from Landsat ETM7; rainfall was obtained from local rainfall data; distance to faults, distance to roads, and distance to rivers were created from a 1:25,000 scale topographic map; the lithology was extracted from geological map. Using these factors, the landslide susceptibility maps were prepared by LR and SI models. The accuracy of the results was verified by using existing landslide locations. The statistical index model had a predictive rate of 81.02%, which is more accurate prediction in comparison with logistic regression model (80.29%). The models can be used to land-use planning in the study area.

Keywords Landslide · Susceptibility · Logistic regression · Statistical index · China

✉ Yanli Wu
wuyl10@lzu.edu.cn

[1] School of Resources and Geoscience, China University of Mining and Technology, Xuzhou 221116, China

[2] College of Resource and Environmental Sciences, Hebei Normal University for Nationalities, Chengde 067000, China

[3] Exploration & Surveying Division, Northwest Electric Power Design Institute Co., Ltd. of China Power Engineering Consulting Group, Xi'an 710075, China

[4] Science Research Department, Hebei Normal University for Nationalities, Chengde 067000, China

[5] School of Civil Engineering and Mechanics, Lanzhou University, Lanzhou 730000, China

Introduction

In China, the annual loss of lives and property due to natural hazards (e.g., earthquakes and landslides) are significantly high, especially in mountainous regions. In Southwest of China, most of the areas are located in mountainous area, and Daguan County has been recognized as one of the most prone to landslides areas in Yunan province, Southwest of China. In recent years, the occurrence frequency of geological disasters has increased, causing local economic development restricted, which has also caused plenty of casualties and property damage every year. Landslide is considered as one of the most damaging natural disasters that usually occur in mountainous regions (Yalcin et al. 2011). In order to decrease the possible damage caused by landslides, the present study aimed to assess or predict landslides for mountainous regions and Daguan County was selected as a suitable case.

In literature, many attempts have been made to apply a variety of qualitative and quantitative models to predict landslides and produce susceptibility maps based on GIS. Quantitative methods, such as frequency ratio (FR) (Pradhan 2010; Yalcin et al. 2011; Park et al. 2013; Youssef et al. 2014), statistical index (SI) (Pourghasemi et al. 2013; Kavzoglu et al. 2015), and logistic regression (Devkota et al. 2013; Felicísimo et al. 2013; Chen and Wang 2007; Pradhan 2010; Nourani et al. 2014; Pourghasemi et al. 2013), have been applied to landslide susceptibility evaluation. Other quantitative methods, such as certainty factor (CF) (Devkota et al. 2013;

Liu et al. 2014; Zhang et al. 2016; Kanungo et al. 2011), index of entropy (IoE) (Devkota et al. 2013; Youssef et al. 2014; Sharma et al. 2015; Constantin et al. 2011), and weights of evidence (WoE) (Ozdemir and Altural 2013; Sharma and Kumar 2008; Soofastaei et al. 2016; Mohammady et al. 2012), have also been frequently used to landslide susceptibility analysis. In recent years, data mining-based methods have also been proposed and implemented for landslide susceptibility mapping, such as artificial neural network (ANN) (Park et al. 2013; Nourani et al. 2014; Yilmaz 2010; Bui et al. 2015), fuzzy set (logic) (Pradhan 2010; Dragićević et al. 2015; Pourghasemi et al. 2012), support vector machine (SVM) (Yilmaz 2010; Bui et al. 2015; Kavzoglu et al. 2014), and random forests (RF) (Stumpf and Kerle 2011a; Stumpf and Kerle 2011b; Chen et al. 2014). In addition, analytical hierarchy process (AHP) model (Kavzoglu et al. 2014; Pourghasemi et al. 2012; Yalcin et al. 2011), a qualitative method based on expert opinion, is also popular in the landslide susceptibility analysis. In summary, varieties of approaches for landslide susceptibility analysis have been applied in the literature.

In this study, two quantitative methods such as logistic regression and statistical index model based on GIS were used to produce the landslide susceptibility maps for the Daguan County, China. To achieve this aim, a detailed landslide inventory map was prepared by literatures, aerial photographs, and supported by field works. And, 15 landslide-conditioning factors were considered. A total of 136 landslides (70%) were applied for the model construction and 58 landslides (30%) were used for the model validation. Finally, success-rate and prediction-rate curve methods were used to validate the accuracy of the landslide susceptibility maps produced from two models.

The study area

The study area is located in Yunnan Province, Southwest of China (Fig. 1), which lies within between 27°36′ and 28°15′ latitude and 103°43′ and 104°07′ longitude and covers an area of 1692 km^2 with size 43.7 km in the east–west direction and 73.2 km in the south–north direction. It is bounded to the northwest by Yanjin County, to the southeast by Yiliang County, to the south by Shaotong City, and to the west and north by Yongshan County. The altitude varies from 523 to 2773 m and slope angle values range from 0° to 77°, and steep slopes are very common. The yearly average temperature is 15 °C, and the annual precipitation is 992.9 mm. In the study area, the Yili River is the main river system; the main river system and its sub-branches are widely distributed. The study area is one of the frequent occurrences of landslides Counties in Yunnan Province, which contains a total of 194 landslides.

Data preparation

Landslide inventory map

A landslide inventory map was considered to be the base of landslide susceptibility mapping. It contains landslide locations that occurred in the past. And, landslide inventory is related to landslide-conditioning factors, such as geology (lithology), geomorphology (slope and aspect), precipitation, land cover, and distance to roads, faults, and rivers (Shahabi et al. 2014). In the current study, the landslide locations were identified based on a previous inventory map, interpretation of satellite images, and extensive field surveys that were used to check the landslides. Eventually, a detailed and reliable inventory map with a total of 194 landslides was created. Figure 1 shows the distribution of landslide locations in the Daguan County, China. The landslide locations were randomly grouped into two parts. From these, 136 (70%) landslides selected were used for training the model, and the remaining 58 (30%) were used for the model validation in this study area. Three types of landslides are observed in the study area: debris flow, unstable slope, and rock falls. Nevertheless, rock fall types are the dominant ones.

Conditioning factors

Landslide-conditioning factors used in the current study are slope angle, slope aspect, curvature, plan curvature, profile curvature, altitude, NDVI, rainfall, STI, distance to roads, distance to rivers, lithology, distance to faults, TWI, and SPI. From these, slope angle, slope aspect, curvature, plan curvature, profile curvature, altitude, STI, SPI, and TWI were derived from digital elevation model (DEM); NDVI was extracted from Landsat ETM7; rainfall was obtained from local rainfall data; distance to faults, distance to roads, and distance to rivers were created from a 1:25,000 scale topographic map; the lithology was extracted from geological map. Eventually, these 15 factors were selected to map the landslide susceptibility for the study area.

Slope angle

Slope angle is one of the key factors in inducing slope instability and is considered to be one of the important factors in landslide susceptibility mapping (Kanungo et al. 2006). It can be seen that this factor was very popularly used in the literature (Kanungo et al. 2006; Lee et al. 2007; Pourghasemi et al.

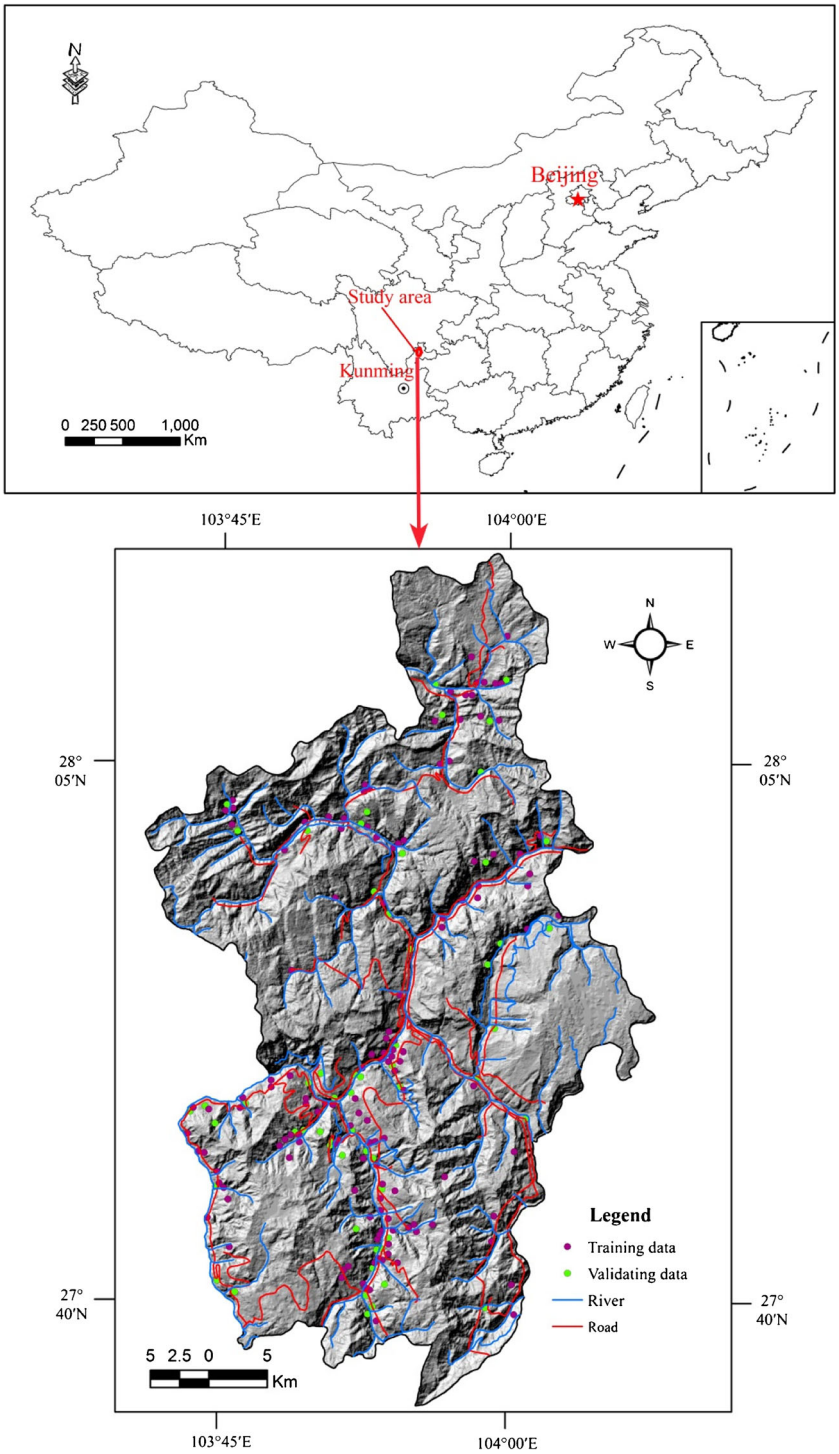


Fig. 1 Location map of study area

2012). In this study, slope angle map was extracted from the digital elevation model (DEM) using ArcGIS 10.0 and grouped into seven classes with an interval of 10°, namely, 0°–10°, 10°–20°, 20°–30°, 30°–40°, 40°–50°, 50°–60°, and >60° (Fig. 2a).

Slope aspect

Slope aspect is also regarded as one of the main landslide-conditioning factors in landslide susceptibility mapping, since generally, landslide occurrences are seriously

affected by slope aspect-related factors (Yalcin et al. 2011). It is also frequently used in landslide susceptibility assessment by different researchers. (Kanungo et al. 2006; Lee et al. 2007; Oh et al. 2010). For the current study, slope aspect was calculated with aid of ArcGIS software, and slope aspect map was categorized into nine classes for the entire map, which is shown in Fig. 2b: flat (−1°): north (337.5°–360°, 0°–22.5°), northeast (22.5°–67.5°), east (67.5°–112.5°), southeast (112.5°–157.5°), south (157.5°–202.5°), southwest (202.5°–247.5°), west (247.5°–292.5°), and northwest (292.5°–337.5°).

Curvature

Curvature is defined as the reciprocal of the radius of a circle that is tangent to the given curve at a point (Ohlmacher 2007). As one of the topographic factors, curvature was selected as a conditioning factor for landslide susceptibility assessment in the current study. The curvature map was also extracted from the digital elevation model (DEM) based on ArcGIS 10.0 and reclassified into three classes (Fig. 2c): <−0.6, −0.6–0.6, and >0.6.

Plan curvature

Plan curvature is described as the curvature of a contour line formed by intersecting a horizontal plane with the surface (Pourghasemi et al. 2012). As a topographic factor and landslide-related factor, plan curvature was divided into three categories in the study area (Fig. 2d), such as <−0.3, −0.3–0.3, and >0.3.

Profile curvature

Similarly, another topographic factor is profile curvature. In particular, profile curvature is defined as the curvature in the vertical plane parallel to the slope direction (Yilmaz et al. 2012). In the same way as for plan curvature, profile curvature was also divided into three categories (Fig. 2e): <−0.5, −0.5–0.5, and >0.5.

Altitude

The altitude map was constructed using the DEM with resolution of 30 × 30 m grid size based on ArcGIS software. In the present study area, the altitude ranges from 523 to 2773 m and is reclassified into seven classes with 300 m interval, such as <800 m, 800–1100 m, 1100–1400 m, 1400–1700 m, 1700–2000 m, 2000–2300 m, and >2300 m (Fig. 2f).

Fig. 2 Thematic maps. **a** Slope angle, **b** slope aspect, **c** curvature, **d** plan curvature, **e** profile curvature, **f** altitude, **g** distance to faults, **h** distance to rivers, **i** distance to roads, **j** lithology, **k** rainfall, **l** NDVI. **m** topographical wetness index (TWI), **n** stream power index (SPI), **o** sediment transport index (STI)

Distance to faults

Fault forms a weak belt or zone which is characterized by heavily fractured rocks (Foumelis et al. 2004; Pourghasemi et al. 2012). In general, closer distance to tectonic structures will result more landslides. In this study, the distance to faults was extracted from the structural geology map of study area at the scale of 1:25,000 and was classified into five classes using an interval of 1000 m based on ArcGIS 10.0 software, and the fault buffer categories were defined as <1000, 1000–2000, 2000–3000, 3000–4000, and >4000 m (Fig. 2j).

Distance to rivers

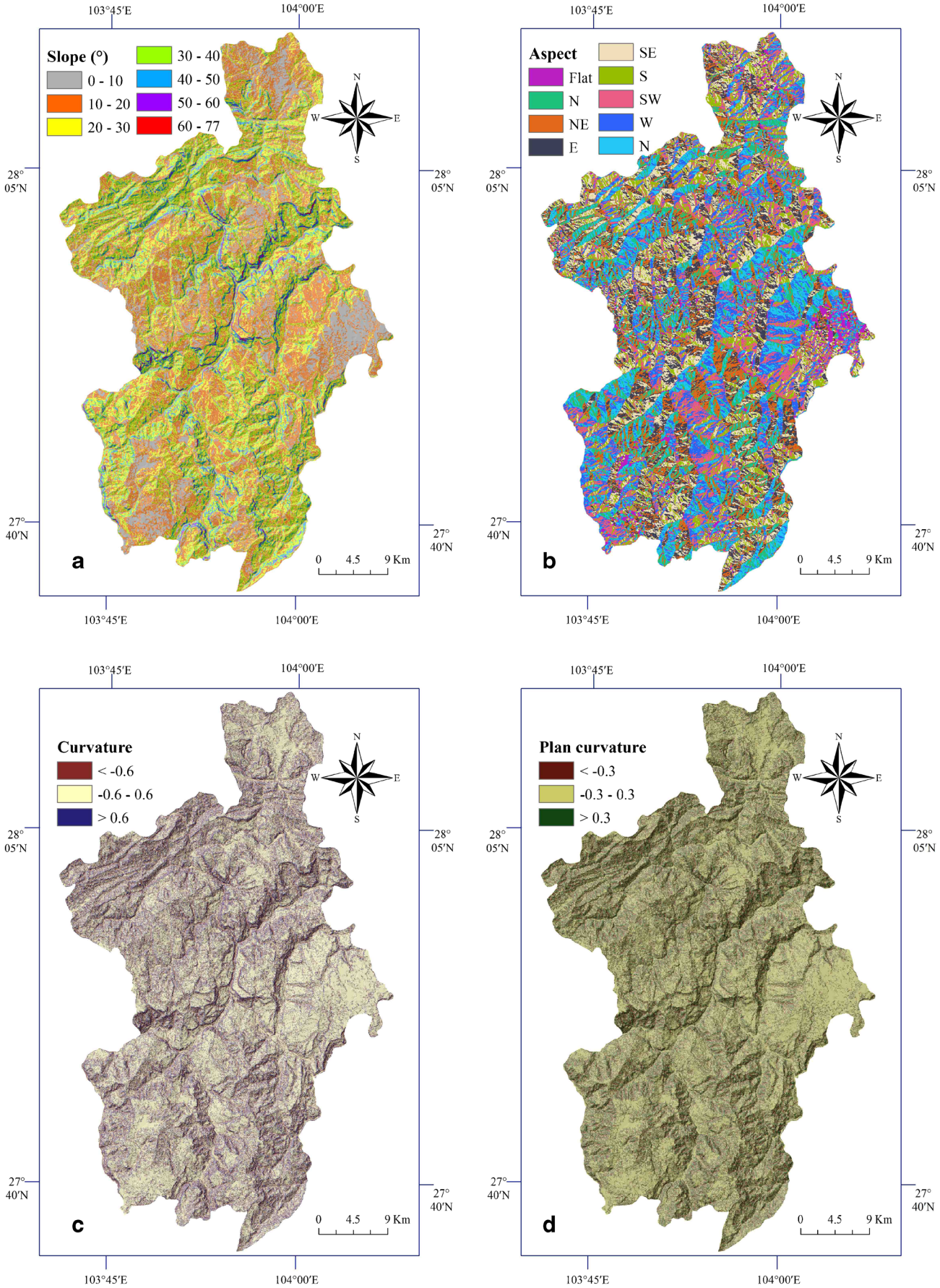
Similar to the distance to roads, the distance to rivers is also very important for landslide susceptibility analysis; generally, the closer the distance from rivers will result more numbers of landslides. In the current study, distance to rivers was classified into five classes using an interval of 500 m, and the river buffer categories were defined as <500, 500–1000, 1000–1500, 1500–2000, and >2000 m (Fig. 2h).

Distance to roads

Distance to road is considered to be one of the causal factors for landslide occurrence, and it is frequently used in landslide susceptibility analysis by many investigators (Nourani et al. 2014; Yilmaz 2010; Pourghasemi et al. 2012). Distribution of landslides along the roads is very common mainly due to the fact that the natural condition of the slope is damaged during the process of road construction; besides, the road cut exposes the joints and fractures that also make the slope unstable. In the present study, distance to roads was taken into account for landslide susceptibility mapping and grouped into five buffer zones using an interval of 500 m (Fig. 2i): <500, 500–1000, 1000–1500, 1500–2000, and >2000 m.

Lithology

The landslide occurrence as a part of geomorphologic studies is related to the lithology (Pourghasemi et al. 2012). It is considered to be a very important conditioning factor because different lithological units have different landslide susceptibility values. In the current study, the lithology map was classified into three classes using ArcGIS 10.0. Various types of lithological units cover the study area. Among them, hard rocks include limestone, dolomite, sandy dolomite, sandstone,


103°45′E
104°00′E
28° 05′N
27° 40′N
Slope (°)
0 - 10
10 - 20
20 - 30
30 - 40
40 - 50
50 - 60
60 - 77
N
W
E
S
0 4.5 9 Km
a
Aspect
Flat
N
NE
E
SE
S
SW
W
N
b
Curvature
< -0.6
-0.6 - 0.6
> 0.6
c
Plan curvature
< -0.3
-0.3 - 0.3
> 0.3
d

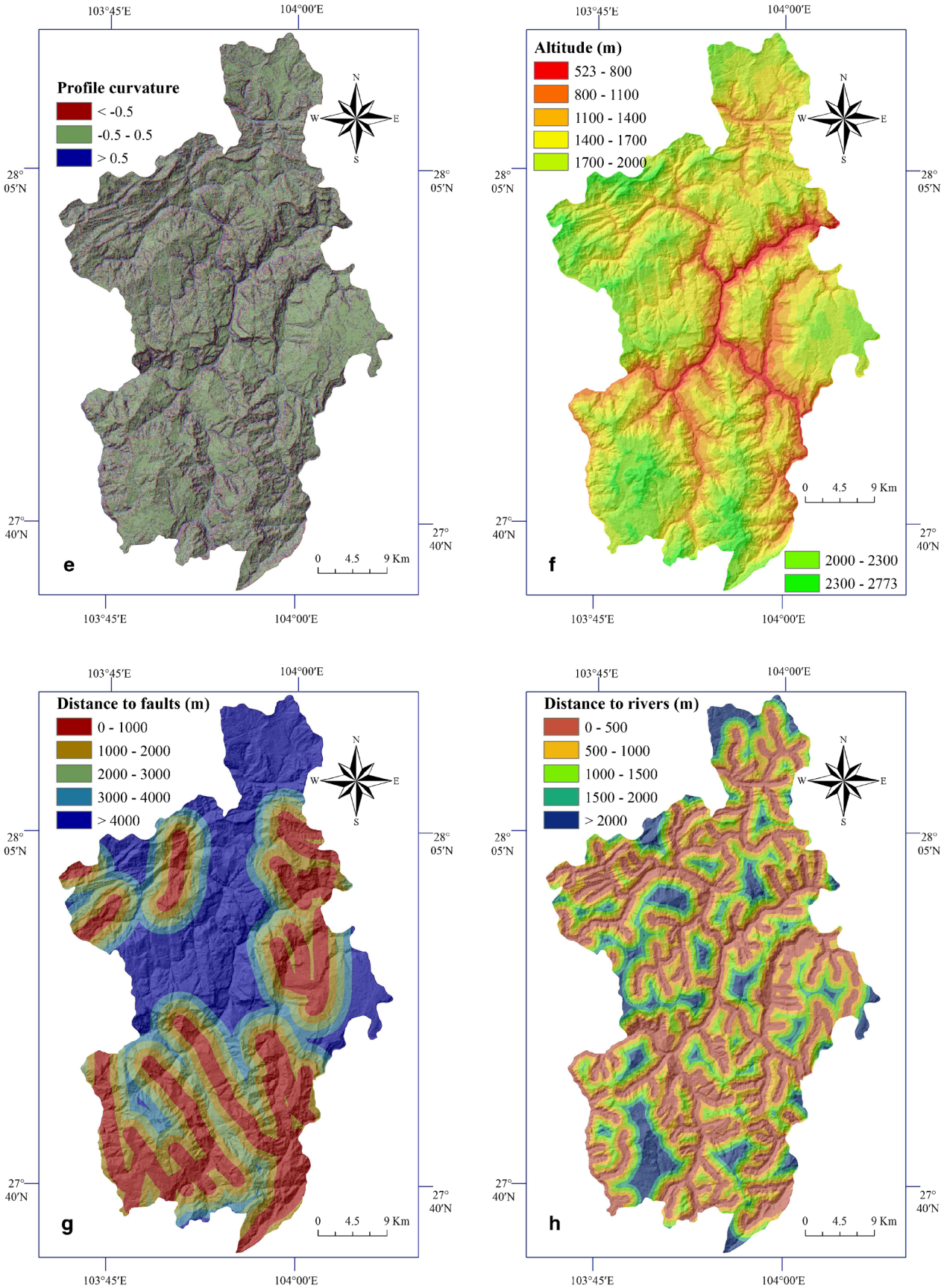


Fig. 2 (continued)

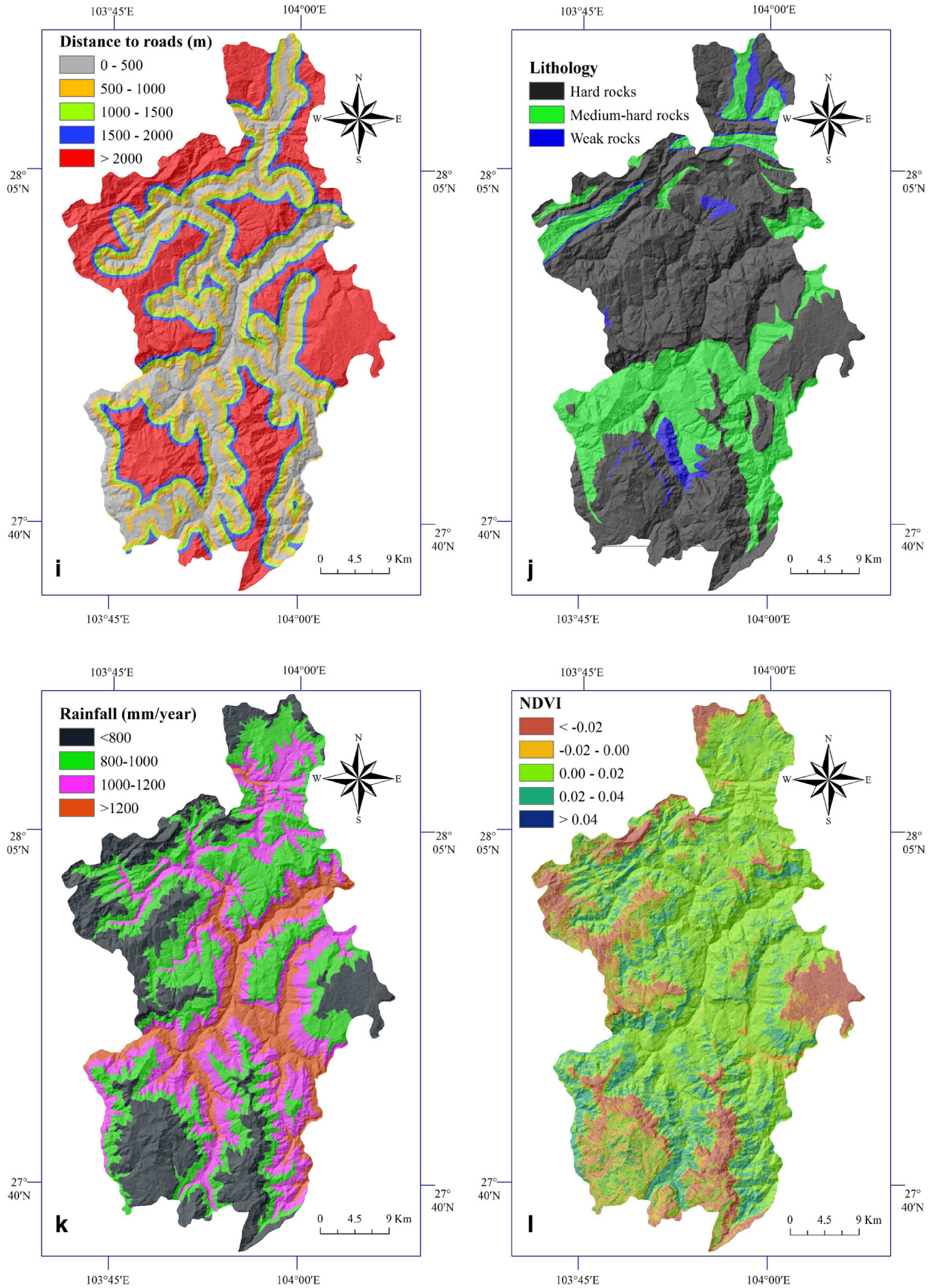


Fig. 2 (continued)

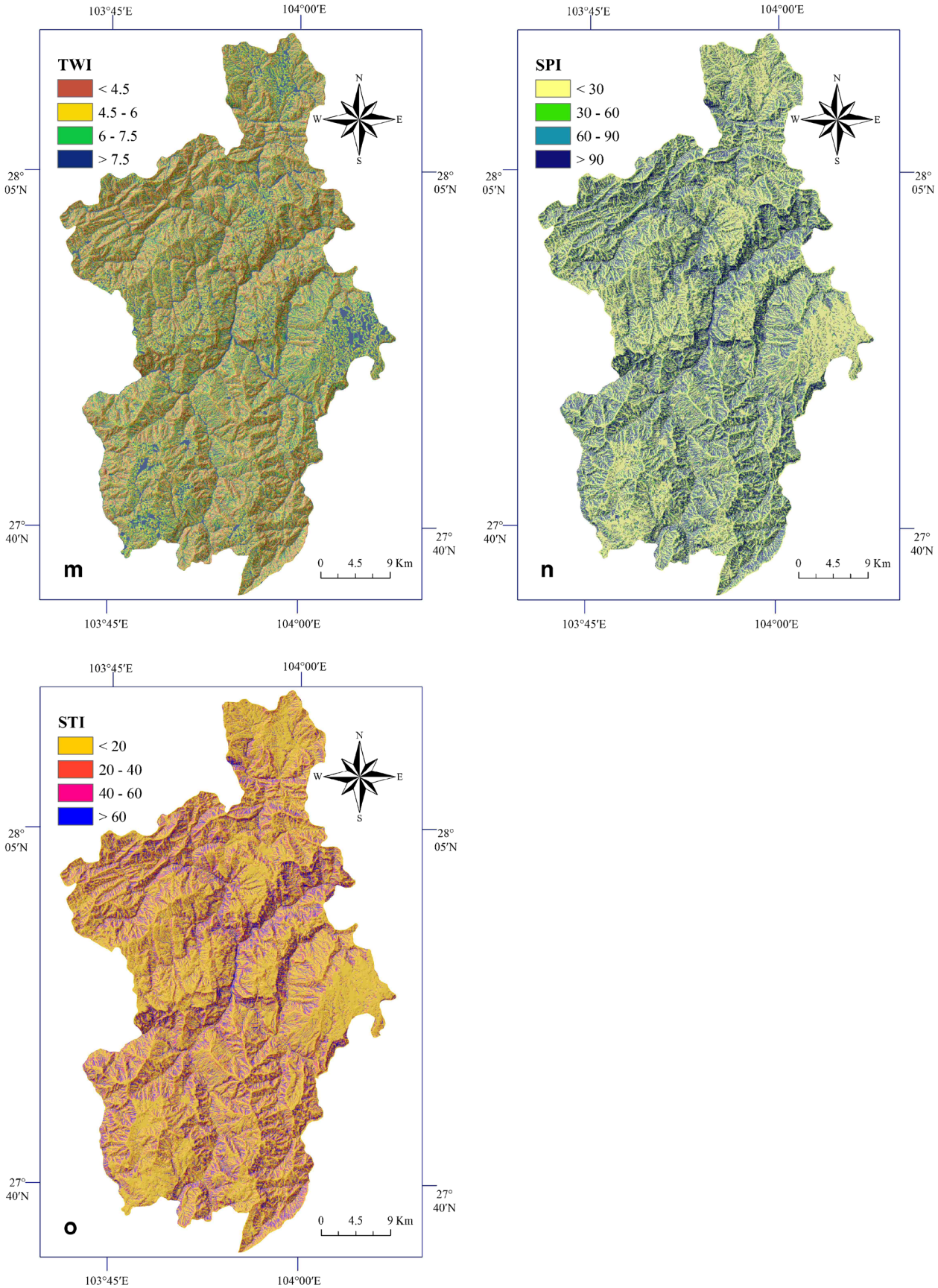


Fig. 2 (continued)

and basalt; the medium-hard rocks consist of argillaceous dolomite and silty sandstone and limestone and dolomite interbedding; the weak rocks are mainly composed of argillaceous dolomite with intercalation of shale, mudstone, siltstone mixed mudstone and shale, shale with intercalation of coal, coal, strong weathering rocks, and soil. It is shown in Fig. 2j.

Rainfall

Rainfall is regarded as one of the most important trigger factors in landslide susceptibility analysis. Rainfall was also considered to produce landslide susceptibility map in this study, and four classes were created using ArcGIS 10.0: <800, 800–1000, 1000–1200, and >1200 mm/year. The average annual precipitation distribution of the study area is shown in Fig. 2k.

NDVI

NDVI, namely, is the normalized difference vegetation index. The NDVI is a measure of surface reflectance and gives a quantitative estimate of vegetation growth and biomass (Hall et al. 1995; Akgun et al. 2012). The NDVI map was extracted from Landsat satellite image. The NDVI value was calculated using ENVI based on the following formula:

$$\mathrm{NDVI} = (IR - R)/(IR + R) \tag{1}$$

where IR is the infrared band of the electromagnetic spectrum and R is the red band of the electromagnetic spectrum.

For the present study, the NDVI value was reclassified into five categories (Fig. 2l): <−0.02, −0.02 to −0.00, −0.00–0.02, 0.02–0.04, and >0.04.

TWI

Topographical wetness index (TWI) is considered to be one of the topographic factors within the runoff model which measures the degree of accumulation of water at a site. It is calculated using the following formula (Beven and Kirkby 1979):

$$TWI = \ln(a/\tan\beta) \tag{2}$$

where a is the specific catchment area and β is the slope gradient (in degrees).

In the study area, TWI value was calculated based on ArcGIS 10.0, which was also divided into four categories (Fig. 2m): <4.5, 4.5–6, 6–7.5, >7.5.

SPI

Stream power index (SPI) is another topographic factor within the runoff model which describes potential flow erosion and related landscape processes (Moore et al. 1991; Akgun et al. 2012). It is defined as follows:

$$SPI = a \times \tan\beta \tag{3}$$

where a is the specific catchment area and β is the slope gradient (in degrees).

The SPI value was calculated with aid of GIS and grouped into four classes: <30, 30–60, 60–90, >90. It is shown in Fig. 2n.

STI

Sediment transport index (STI), as one of the important topographic factors, reflects the erosive power of overland flow (Jaafari et al. 2014). It is defined as follows (Moore and Burch 1986):

$$STI = \left(\frac{a}{22.13}\right)^{0.6} \times \left(\frac{\sin\beta}{0.0896}\right)^{1.3} \tag{4}$$

where a is the specific catchment area and β is the slope gradient (in degrees).

STI was also considered to produce landslide susceptibility map for the present study. In this study, STI values were classified into four classes (Fig. 2o), such as <20, 20–40, 40–60, and >60.

Methodology

Logistic regression

Logistic regression method is considered to be one of the statistical methods, which is frequently used in landslide susceptibility assessment by different researchers in different parts of the world (Nourani et al. 2014; Pradhan and Lee 2010; Lee et al. 2007; Oh et al. 2010). As one of the multivariate analysis models, logistic regression model was also selected to quantitatively assess the landslide susceptibility in the present study. Logistic regression allows one to form a multivariate regression relationship between a dependent variable and some independent variables, which is useful for predicting the presence (landslides) or absence (non-landslides) of a characteristic or outcome based on values of a set of predictor variables. The advantage of logistic regression is that the variables may be either continuous or discrete, or any combination of both types, and they do not necessarily have normal distributions, by means of addition of an appropriate link function to the usual linear regression model (Pradhan and Lee 2010). In the present study, the dependent variable is a binary variable (0 or 1), 1 represents presence of a landslide, and 0 for absence of a landslide. The logistic regression model is defined by the following equation: (Atkinson and Massari 1998):

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{5}$$

where Z is the linear combination, β_0 is the intercept of the logistic regression model, the β_1, β_2, …β_n are the coefficients of logistic regression, and X_1, X_2, …X_n are the independent variables.

In order to predict the landslide occurrence, the relationship between the probability of landslide occurrence and its dependency on several variables can be expressed as follows:

$$P = \frac{\exp(z)}{1-\exp(z)} \tag{6}$$

where Z is the linear combination, P is the probability of landslide occurrence.

Statistical index

The statistical index method is another statistical method, which is first proposed by van Westen et al. (1997) in landslide susceptibility assessment. Subsequently, this method has been applied for landslide susceptibility assessment by various researchers in their studies (Pourghasemi et al. 2013; Kavzoglu et al. 2015). In this model, the weight for a parameter class is defined as the natural logarithm of the landslide density in the class divided by the landslide density in the entire study area. The statistical index method is expressed as the following formula (Van Westen et al. 2007):

$$S_{ij} = \ln\left(\frac{D_{ij}}{D}\right) = \ln\left[\left(\frac{N_{ij}}{P_{ij}}\Big/\frac{N}{P}\right)\right] \tag{7}$$

where S_{ij} is the weight given to a certain class i of parameter j, D_{ij} is the landslide density within the class i of parameter j, D is the landslide density within the study area, N_{ij} is the number of landslides in a certain class i of parameter j, P_{ij} is the number of pixels of a certain class i of parameter j, N is the total landslides within the study area, and P is the total pixels within the study area.

Results and discussions

Application of logistic regression

The model coefficients were calculated using the statistical analysis software SPSS. Through the logistic regression model, spatial relationship between each landslide-conditioning factor and landslide is assessed and shown in Table 1. In this analysis, slope angle, curvature, plan curvature, profile curvature, altitude, NDVI, STI, distance to roads, distance to rivers, distance to faults, TWI, and SPI were defined as the "continuous data" which were treated as "scale" in statistical analysis software SPSS, while slope aspect, rainfall, and lithology were "incontinuous data" which were treated as "nominal" data in SPSS.

Using eq. (5), the formula predicting the landslide occurrence in the present study was obtained as follows:

$$\begin{aligned} z = {} & (1.714^{*}\text{slope angle}) + \text{slope aspect}_c \\ & + (0.001^{*}\text{curvature}) + (17.164^{*}\text{plan curvature}) \\ & + (2.918^{*}\text{profile curvature}) + (-6.380^{*}\text{altitude}) \\ & + (-0.563^{*}\text{distance to faults}) \\ & + (-1.317^{*}\text{distance to rivers}) \\ & + (-1.772^{*}\text{distance to roads}) + \text{lithology}_c \\ & + \text{rainfall}_c + (0.030^{*}\text{NDVI}) + (1.947^{*}\text{TWI}) \\ & + (138.314^{*}\text{SPI}) + (-27.343^{*}\text{STI}) - 3.983 \end{aligned} \tag{8}$$

where slope angle is the slope value, curvature is the curvature value, plan curvature is the plan curvature value, profile curvature is the profile curvature value, altitude is the altitude value, distance to faults is the distance to faults value, distance to roads is the distance to roads value, NDVI is the NDVI value, TWI is the TWI value, SPI is the SPI value, STI is the STI value. Logistic regression coefficient values are 1.714, slope aspect$_c$, 0.001, 17.164, 4.19, 2.918, lithology$_c$, rainfall$_c$, −1.317, −1.772, 0.030, 1.947, 138.314, and −27.343 listed in Table 1, and z is the parameter. The model constant or the intercept of model is −3.983.

Sample calculation of LR for few pixels is provided in Table 2. Using eqs. (6) and (8), the probability of landslide occurrence (P) of six samples was calculated. The P is 0 which indicates no occurrence of landslides.

From Eq. (8), it is obvious that slope angle, curvature, plan curvature, profile curvature, distance to rivers, NDVI, and SPI coefficients are positive. This means that these factors are positively related to the landslide occurrence, whereas altitude, distance to faults, distance to roads, TWI, and STI coefficients are negative which indicate a negative relation with the landslide occurrence in the study area.

Using eq. (6), the probability of landslide occurrence (P) was calculated.

Eventually, using eqs. (6) and (8), a landslide susceptibility map was created by logistic regression model based on GIS and reclassified into five classes for visual interpretation using Jenks natural break classification method: very low, low, moderate, high, and very high (Fig. 3).

Application of statistical index

Spatial relationship between each landslide-conditioning factor and landslide by statistical index model is shown in Table 1. From Table 1, it is seen that slope class in 50–60° has highest positive weight value of 0.330, which indicates that the landslide probability is higher in this class. For slope

Table 1 Spatial relationship between each landslide-conditioning factor and landslide by LR and SI model

Factors	Classes	No. of landslide	No. of pixels in domain	Percentage of landslide (%)	Percentage of pixel in the domain (%)	Coefficients of logistic regression	SI
Slope angle (°)	0–10	10	315,376	7.353	15.398	1.714	−0.739
	10–20	35	604,562	25.735	29.517		−0.137
	20–30	50	584,333	36.765	28.530		0.254
	30–40	28	369,683	20.588	18.050		0.132
	40–50	10	137,382	7.353	6.708		0.092
	50–60	3	32,473	2.206	1.585		0.330
	>60	0	4345	0.000	0.212		0
Slope aspect	Flat	0	72,489	0.000	3.539	−22.182	0
	North	12	209,258	8.824	10.217	−0.497	−0.147
	Northeast	13	237,809	9.559	11.611	0.135	−0.194
	East	16	285,248	11.765	13.927	−0.277	−0.169
	Southeast	20	275,745	14.706	13.463	0.742	0.088
	South	11	212,510	8.088	10.376	−0.149	−0.249
	Southwest	21	244,987	15.441	11.961	0.563	0.255
	West	23	257,742	16.912	12.584	0.190	0.296
	Northwest	20	252,366	14.706	12.322	0	0.177
Curvature	< −0.6	27	440,821	19.853	21.523	0.001	−0.081
	−0.6 - 0.6	77	1,167,868	56.618	57.021		−0.007
	> 0.6	32	439,465	23.529	21.457		0.092
Plan curvature	< −0.3	20	454,804	14.706	22.206	17.164	−0.412
	−0.3 - 0.3	79	1,118,913	58.088	54.630		0.061
	> 0.3	37	474,437	27.206	23.164		0.161
Profile curvature	< −0.5	26	320,275	19.118	15.637	2.918	0.201
	−0. 5–0.5	88	1,398,289	64.706	68.271		−0.054
	> 0.5	22	329,590	16.176	16.092		0.005
Altitude (m)	<800	8	57,659	5.882	2.815	−6.380	0.737
	800–1100	32	196,122	23.529	9.576		0.899
	1100–1400	59	348,656	43.382	17.023		0.935
	1400–1700	28	552,115	20.588	26.957		−0.270
	1700–2000	8	493,685	5.882	24.104		−1.410
	2000–2300	1	317,630	0.735	15.508		−3.049
	>2300	0	82,287	0.000	4.018		0
Distance to faults (m)	0–1000	43	546,775	31.618	26.693	−0.563	0.169
	1000–2000	20	395,589	14.706	19.312		−0.272
	2000–3000	29	309,084	21.324	15.089		0.346
	3000–4000	19	219,453	13.971	10.713		0.265
	>4000	25	577,486	18.382	28.192		−0.428
Distance to rivers (m)	0–500	92	770,992	67.647	37.639	−1.317	0.586
	500–1000	31	574,173	22.794	28.030		−0.207
	1000–1500	9	377,803	6.618	18.444		−1.025
	1500–2000	2	190,566	1.471	9.303		−1.845
	>2000	2	134,853	1.471	6.583		−1.499
Distance to roads (m)	0–500	84	467,318	61.765	22.814	−1.772	0.996
	500–1000	25	328,166	18.382	16.021		0.138
	1000–1500	10	271,267	7.353	13.243		−0.588
	1500–2000	9	223,859	6.618	10.929		−0.502
	>2000	8	757,777	5.882	36.994		−1.839
Lithology	Hard rock groups	51	1,392,782	37.500	67.978	−2.413	−0.595

Table 1 (continued)

Factors	Classes	No. of landslide	No. of pixels in domain	Percentage of landslide (%)	Percentage of pixel in the domain (%)	Coefficients of logistic regression	SI
	Medium-hard rock groups	73	572,709	53.676	27.952	−1.186	0.652
	Weak rock groups	12	83,394	8.824	4.070	0	0.774
Rainfall	< 800 mm/yr	43	260,798	31.618	12.732	−0.655	0.910
	800-1000 mm/yr	71	545,461	52.206	26.629	−0.165	0.673
	1000-1200 mm/yr	17	694,437	12.500	33.902	−0.724	−0.998
	> 1200 mm/yr	5	547,691	3.676	26.738	0	-1.984
NDVI	<−0.02.	1	320,960	0.735	15.707	0.030	−3.062
	−0.02--0.00	2	199,906	1.471	9.783		−1.895
	−0.00-0.02	102	1,143,564	75.000	55.963		0.293
	0.02–0.04	31	377,921	22.794	18.495		0.209
	>0.04	0	1060	0.000	0.052		0
TWI	<4.5	40	547,912	29.412	26.752	1.947	0.095
	4.5–6	47	719,154	34.559	35.112		−0.016
	6–7.5	28	410,555	20.588	20.045		0.027
	>7.5	21	370,533	15.441	18.091		-0.158
SPI	<30	49	1,001,467	36.029	48.896	138.314	−0.305
	30–60	25	290,862	18.382	14.201		0.258
	60–90	13	161,048	9.559	7.863		0.195
	>90	49	594,777	36.029	29.040		0.216
STI	<20	71	1,270,415	52.206	62.027	−27.343	−0.172
	20–40	32	394,013	23.529	19.237		0.201
	40–60	12	155,324	8.824	7.584		0.151
	>60	21	228,402	15.441	11.152		0.325

Model constant is −3.983

aspect, Southeast, Southwest, West, and Northwest have positive weight values of 0.088, 0.255, 0.296, and 0.177, respectively. This means that these classes are more prone to landslide occurrence. In case of curvature, the only positive weight value was seen in class of >0.6. For plan curvature, classes of −0.3–0.3 and >0.6 have positive weight values of

Table 2 Sample calculation of LR for few pixels

Factors	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Slope angle (°)	0.580	30.682	4.256	6.932	20.185	27.932
Slope aspect	0.563	0.563	0.135	0.742	0.000	−0.277
Curvature	0.512	2.464	0.625	0.983	−0.353	−0.167
Plan curvature	0.000	1.086	0.331	0.265	−0.400	−0.251
Profile curvature	−0.512	−1.378	−0.294	−0.717	−0.047	−0.084
Altitude (m)	1273.000	2081.420	1392.750	1333.000	2338.820	2334.040
Distance to faults (m)	1539.420	1447.050	8462.430	1727.180	1034.290	4060.800
Distance to rivers (m)	1216.110	3032.050	145.229	1340.200	263.020	1497.190
Distance to roads (m)	775.037	1102.210	116.183	58.092	3516.810	5182.930
Lithology	−1.186	−2.413	0.000	−1.186	−2.413	−1.186
Rainfall	−0.724	−0.655	−0.724	−0.724	−0.655	-0.655
NDVI	0.011	−0.007	0.014	0.019	−0.029	-0.043
TWI	7.65	4.62	5.65	5.16	7.35	6.12
SPI	0.21	35.60	1.58	2.58	210.57	127.25
STI	6.155	6.987	9.298	14.750	1.105	23.378

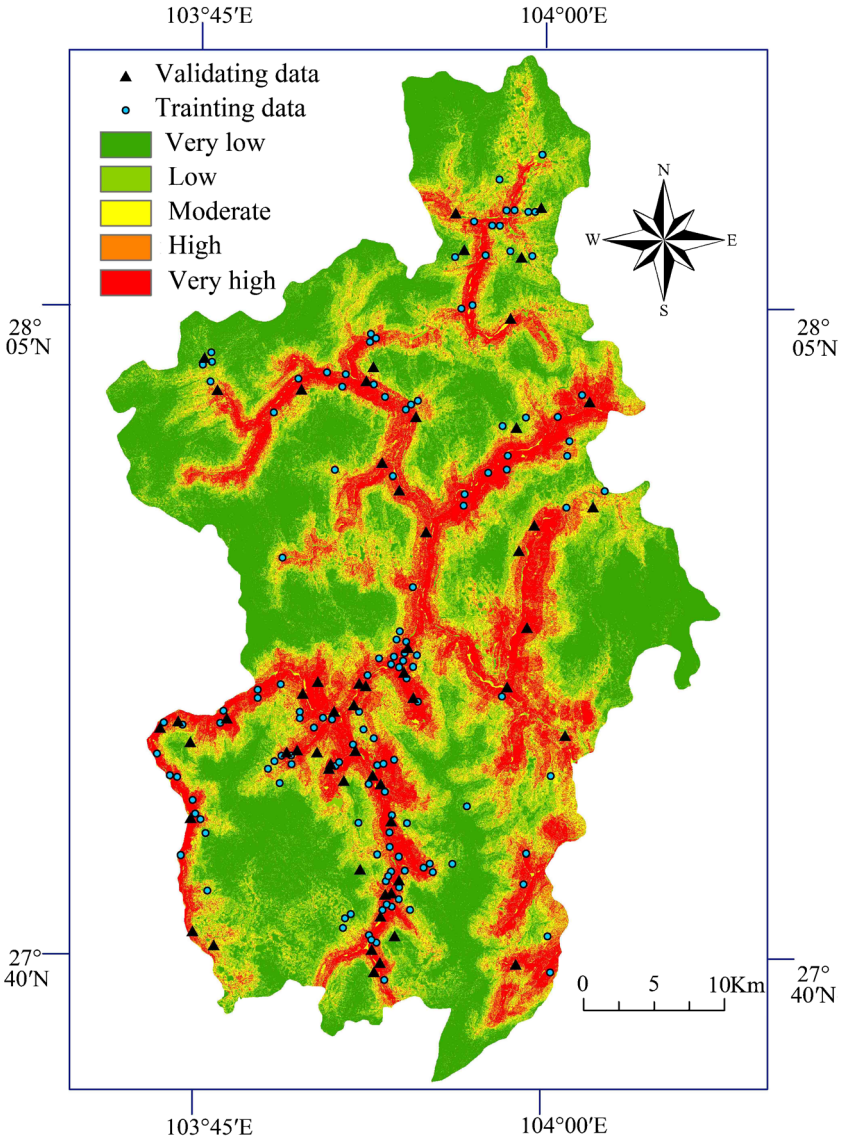


Fig. 3 Landslide susceptibility map using LR model

0.061 and 0.161, respectively. For profile curvature, the SI value is negative (−0.054) in the range between −0.5 and 0.5; other classes have positive value. This means that the range between −0.5 and 0.5 is less prone to landslide occurrence. In case of altitude, classes of <800, 800–1100, 1100–1400, and 1400–1700 m have positive weight values (0.737, 0.899, and 0.935, respectively), whereas classes of 1700–2000 and 2000–2300 m have negative weight values. It indicates that maybe most of people live in the area <1700 m and have less effect in altitude of 1700 m. In case of distance from faults, classes of 0–1000, 2000–3000, and 3000–4000 m have positive weight values (0.169, 0.346, and 0.265, respectively), while class of >4000 m has negative weight value of −0.428. It is clear that as the distance from faults increases, the landslide frequency generally decreases. In case of distance from rivers, the only positive weight value was seen in class of 0–500 m; other classes are −0.207, −1.025, −1.845, and −1.499 for 500–1000, 1000–1500, 1500–2000, and >2000 m, respectively. Similarly, it can be observed that as the distance from rivers increases, the probability of landslide occurrence generally decreases. As for distance from roads, class of 0–500 m has the highest positive weight value of 0.996, which indicates that landslide is prone to occur in this class, and as the distance from rivers decreases, the probability of landslide occurrence generally increases. For lithology, hard rock groups have the negative weight value of −0.595, and weak rock groups have the highest positive weight value of 0.774. This supports the fact that generally hard rock is less prone to landslide whereas weak rock is more prone to landslide. In case of rainfall, class of <800 mm/year has the highest positive weight value of 0.910, followed by 800–1000 mm/year (0.673), 1000–1200 mm/year (−0.998), and >1200 mm/year (−1.984). In case of NDVI, the highest weight value was seen in class of −0.00–0.02 (0.293), and the lowest weight value was seen in class of <−0.02. For TWI, class of <4.5 has the highest SI value of 0.095, followed by 6–7.5 (0.027), 4.5–6

Table 3 Sample calculation of SI for few pixels

Factors	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
Slope angle (°)	−0.739	0.132	−0.739	−0.739	0.254	0.254
Slope aspect	0.255	0.255	−0.194	0.088	0.177	−0.169
Curvature	−0.128	−0.128	−0.128	−0.128	−0.128	−0.128
Plan curvature	0.061	0.161	0.161	0.061	−0.412	0.061
Profile curvature	0.201	0.201	−0.054	0.201	−0.054	−0.054
Altitude (m)	0.935	−0.304	0.935	0.935	0.000	0.000
Distance to faults (m)	−0.272	−0.272	−0.428	−0.272	−0.272	−0.428
Distance to rivers (m)	−0.102	−0.149	0.586	−0.102	0.586	−0.102
Distance to roads (m)	0.138	−0.588	0.996	0.996	−0.183	−0.183
Lithology	0.652	−0.595	0.774	0.652	−0.595	0.652
Rainfall	0.673	−0.198	0.673	0.673	−0.198	-0.198
NDVI	0.293	−0.189	0.293	0.293	−0.306	-0.306
TWI	-0.158	−0.016	−0.016	−0.016	0.027	0.027
SPI	-0.305	0.258	−0.305	−0.305	0.216	0.216
STI	-0.172	−0.172	−0.172	−0.172	0.151	0.201

(−0.016), and >7.5 (−0.158). For SPI, the 30–60 range has the highest SI value (0.258), and class of <30 has the lowest SI value (−0.305). For STI, class of >60 has the highest positive weight value of 0.325, followed by 20–40 (0.201), 40–60

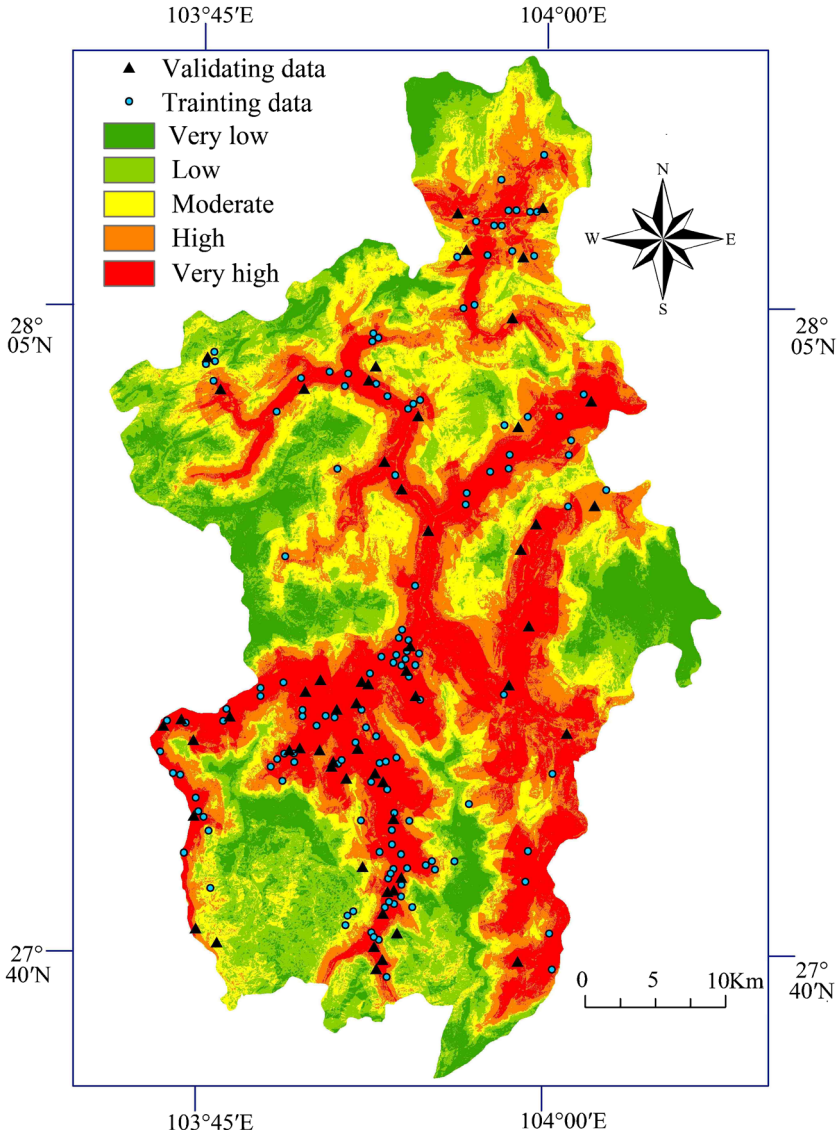


Fig. 4 Landslide susceptibility map using SI model

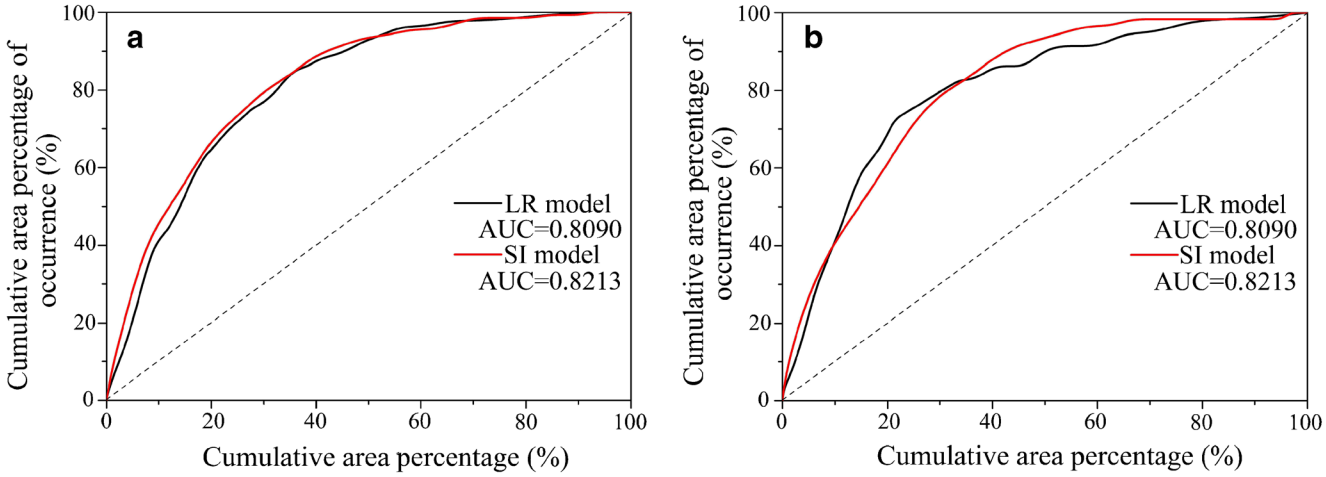


Fig. 5 **a** Success rate of the landslide susceptibility map, **b** prediction rate of the landslide susceptibility map

(0.151), and <20 (−0.172). This means that landside is more prone to occur in the class of >60, while landside is less prone to occur in the class of <20.

Finally, the landslide susceptibility map was produced using the following equation:

$$\begin{aligned} \mathrm{LSI} = {} & (\text{slope angle})_{\mathrm{SI}} + (\text{slope aspect})_{\mathrm{SI}} + (\text{curvature})_{\mathrm{SI}} \\ & + (\text{plan curvature})_{\mathrm{SI}} + (\text{profile curvature})_{\mathrm{SI}} \\ & + (\text{altitude})_{\mathrm{SI}} + (\text{distance to faults})_{\mathrm{SI}} \\ & + (\text{distance to rivers})_{\mathrm{SI}} + (\text{distance to roads})_{\mathrm{SI}} \\ & + (\text{lithology})_{\mathrm{SI}} + (\text{rainfall})_{\mathrm{SI}} + (\text{NDVI})_{\mathrm{SI}} \\ & + (\text{TWI})_{\mathrm{SI}} + (\text{SPI})_{\mathrm{SI}} + (\text{STI})_{\mathrm{SI}} \end{aligned} \quad (9)$$

Sample calculation of SI for few pixels is provided in Table 3. Using eq. (9), LSI calculated was 1.332, −1.604, 2.382, 2.165, −0.737, and −0.157 for samples 1, 2, 3, 4, 5, and 6, respectively.

Using ArcGIS 10.0 software, the landslide susceptibility map was divided into five regions by Jenks natural break classification method: very low, low, moderate, high, and very high (Fig. 4).

Validation of landslide susceptibility maps

In order to evaluate the performances of landslide susceptibility maps by logistic regression and statistical index models, both the success and predication rate curves were applied by comparing them with the existing landslide data. The area under the curve (AUC) was used to evaluate the accuracy of landslide susceptibility maps quantitatively. The success-rate curve was obtained by comparing the training dataset (136 landslides) with the landslide susceptibility map (Fig. 5a). Similarly, the prediction rate curve was drawn by a comparison of the validating dataset (58 landslides) with the landslide susceptibility map (Fig. 5b). The success-rate curve result shows that the area under the curve (AUC) value was 0.8090 for LR model and 0.8213 for SI model, which indicates that the success rate was 80.90% for LR model and 82.13% for SI model. This means that the capability for correctly classifying the areas with existing landslides is higher in this study. The prediction rate curve shows that the area under the curve (AUC) value was 0.8029 for LR model and 0.8102 for SI model, which means that the success rate was 80.29% for LR model and 81.02% for SI model. It also indicates that both LR model and SI model with the higher prediction accuracy and predict landslides very well in the study area. Using the validation of landslide susceptibility maps by the success-rate and prediction-rate methods, it is seen that both the success rate and the prediction rate have the same results and perform very well in landslide susceptibility mapping.

Conclusions

GIS-based techniques have been widely used in landslide susceptibility mapping throughout the world. In this study, GIS-based logistic regression (LR) and statistical index (SI) models were applied for landslide susceptibility mapping in Daguan County, Southwest of China, and their performances were compared and tested. From the validation of the results, both the logistic regression (LR) and statistical index (SI) models have been successfully applied to produce the reliable landslide susceptibility maps in the current study area.

In process of producing landslide susceptibility maps of the study area, a landslide inventory map of study area was compiled in the first stage, and a total of 194 landslides were detected by literatures, aerial photographs, and supported by field works. Landslides were randomly classified into two for modeling (70%) and model validation (30%). Fifteen landslide-related factors were selected from spatial database: slope angle, slope aspect, curvature, plan curvature, profile curvature, altitude, distance from faults, distance from rivers, distance from roads,

lithology, rainfall, NDVI, TWI, SPI, and STI. Through the analysis by two quantitative methods (LR and SI models) in ArcGIS 10.0, the landslide susceptibility maps of the study area were constructed. Finally, to validate the results, both the success-rate and prediction-rate methods were applied, and the area under the curve (AUC) was calculated. The validation results show that the landslide susceptibility map produced by logistic regression (LR) and statistical index (SI) model with the prediction rate of 80.29 and 81.02%, respectively. Similarly, the success rates were 80.90 and 82.13% for LR and SI models, respectively. It is concluded that statistical index (SI) model performs slightly better than logistic regression (LR) model, and two models with higher prediction accuracy can be used for land use planning and slope management in near future.

Acknowledgments The authors thank the National Basic Research Program of China "973" (No. 2015CB251601), the People's Livelihood Research Project of Hebei Province (201301211), and the Research Fund for the Scientific Studies in Higher Education Institutions of Hebei Province (QN2015306) for the financial support. Also, the authors would like to express their gratitude to the anonymous reviewers for their constructive comments and suggestions, which highly increased the quality of the paper.

References

Akgun A, Sezer EA, Nefeslioglu HA, Gokceoglu C, Pradhan B (2012) An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Comput Geosci 38(1):23–34

Atkinson PM, Massari R (1998) Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. Comput Geosci 24(4):373–385

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. Hydrol Sci J 24(1):43–69

Bui DT, 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 1–18

Chen Z, Wang J (2007) Landslide hazard mapping using logistic regression model in Mackenzie Valley, Canada. Nat Hazards 42(1):75–89

Chen W, Li X, Wang Y, Chen G, Liu S (2014) Forested landslide detection using LiDAR data and the random forest algorithm: a case study of the three gorges, China. Remote Sens Environ 152:291–301

Constantin M, Bednarik M, Jurchescu MC, Vlaicu M (2011) Landslide susceptibility assessment using the bivariate statistical analysis and the index of entropy in the Sibiciu Basin (Romania). Environmental Earth Sciences 63(2):397–406

Devkota KC, Regmi AD, Pourghasemi HR, Yoshida K, Pradhan B, Ryu IC et al (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

Dragićević S, Lai T, Balram S (2015) GIS-based multicriteria evaluation with multiscale analysis to characterize urban landslide susceptibility in data-scarce environments. Habitat International 45:114–125

Felicísimo ÁM, Cuartero A, Remondo J, Quirós E (2013) Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10(2):175–189

Foumelis M, Lekkas E, Parcharidis I (2004) Landslide susceptibility mapping by GIS-based qualitative weighting procedure in Corinth area. Bulletin of the Geological Society of Greece XXXVI, 904–912. Proceedings of the 10th international congress, Thessaloniki, April 2004

Hall FG, Townshend JR, Engman ET (1995) Status of remote sensing algorithms for estimation of land surface state parameters. Remote Sens Environ 51(1):138–156

Jaafari A, Najafi A, Pourghasemi HR, 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(4):909–926

Kanungo DP, Arora MK, Sarkar S, Gupta RP (2006) A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Eng Geol 85(3):347–366

Kanungo DP, Sarkar S, Sharma S (2011) Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. Nat Hazards 59(3):1491–1512

Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. Landslides 11(3):425–439

Kavzoglu T, Sahin EK, Colkesen I (2015) An assessment of multivariate and bivariate approaches in landslide susceptibility mapping: a case study of Duzkoy district. Nat Hazards 76(1):471–496

Lee S, Ryu JH, Kim IS (2007) Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. Landslides 4(4):327–338

Liu M, Chen X, Yang S (2014) Collapse Landslide and Mudslides Hazard Zonation. In Landslide Science for a Safer Geoenvironment (pp. 457–462). Springer International Publishing

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, Burch GJ (1986) Physical basis of the length-slope factor in the universal soil loss equation. Soil Sci Soc Am J 50(5):1294–1298

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5(1):3–30

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

Oh HJ, Lee S, Soedradjat GM (2010) Quantitative landslide susceptibility mapping at Pemalang area, Indonesia. Environmental Earth Sciences 60(6):1317–1328

Ohlmacher GC (2007) Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Eng Geol 91(2):117–134

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 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. Environmental earth sciences 68(5):1443–1464

Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. Nat Hazards 63(2):965–996

Pourghasemi HR, Moradi HR, Aghda SF (2013) Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. Nat Hazards 69(1):749–779

Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. Journal of the Indian Society of Remote Sensing 38(2):301–320

Pradhan B, Lee S (2010) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. Environmental Earth Sciences 60(5):1037–1054

Shahabi H, Khezri S, Ahmad BB, Hashim M (2014) Landslide susceptibility mapping at central Zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models. Catena 115:55–70

Sharma M, Kumar R (2008) GIS-based landslide hazard zonation: a case study from the Parwanoo area, Lesser and Outer Himalaya, HP, India. B Eng Geol Environ 67(1): 129–137

Sharma LP, Patel N, Ghose MK, Debnath P (2015) Development and application of Shannon's entropy integrated information value model for landslide susceptibility assessment and zonation in Sikkim Himalayas in India. Nat Hazards 75(2):1555–1576

Soofastaei A, Aminossadati SM, Arefi MM, Kizil MS (2016) Development of a multi-layer perceptron artificial neural network model to determine haul trucks energy consumption. Int J Min Sci Technol 26(2): 285–293

Stumpf A, Kerle N (2011a) Object-oriented mapping of landslides using random forests. Remote Sens Environ 115(10):2564–2577

Stumpf A, Kerle N (2011b) Combining random forests and object-oriented analysis for landslide mapping from very high resolution imagery. Procedia Environmental Sciences 3:123–129

Van Westen CJ (1997) Statistical landslide hazard analysis. ILWIS 2:73–84

Van Westen CJ, Rengers N, Terlien MTJ, Soeters R (1997) Prediction of the occurrence of slope instability phenomenal through GIS-based hazard zonation. Geol Rundsch 86(2):404–414

Yalcin A, Reis S, Aydinoglu AC, Yomralioglu T (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. Catena 85(3):274–287

Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environmental Earth Sciences 61(4):821–836

Yilmaz C, Topal T, Süzen ML (2012) GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak-Turkey). Environmental earth sciences 65(7):2161–2178

Youssef AM, Al-Kathery M, Pradhan B (2014) Landslide susceptibility mapping at al-hasher area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models. Geosci J 19(1):113–134

Zhang P, Peterson S, Neilans D, Wade S, McGrady R, Pugh J (2016) Geotechnical risk management to prevent coal outburst in room-and-pillar mining. Int J Min Sci Technol 26(1):9–18