

Landslides (2017) 14:1091–1111
 DOI 10.1007/s10346-016-0769-4
 Received: 2 June 2016
 Accepted: 10 October 2016
 Published online: 22 October 2016
 © Springer-Verlag Berlin Heidelberg 2016

Paraskevas Tsangaratos · Ioanna Ilia · Haoyuan Hong · Wei Chen · Chong Xu

Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China

Abstract The main objective of the present study was to produce a landslide susceptibility map by implementing a novel methodology that combines Information Theory and GIS-based methods for the Nancheng County, China, an area with numerous reported landslide events. Specifically, the information coefficient that is estimated from Shannon's entropy index was used to determine the number of classes of each landslide-related variable that maximizes the information coefficient, while three methods, logistic regression, weight of evidence, and random forest algorithm, were implemented to produce the landslide susceptibility map. The comparison of the various models was based on the assessment of a database of 112 past landslide events, which were divided randomly into a training dataset (70%) and a validation dataset (30%). The identification of the areas affected was established by analyzing airborne imagery, extensive field investigation, and the examination of previous research studies, while the morphometric variables were derived using remote sensing technology. The geo-environmental conditions in those locations were analyzed regarding their susceptibility to slide. In particular, 11 variables were analyzed: lithology, altitude, slope, aspect, topographic wetness index, sediment transport index, profile curvature, plan curvature, distance to rivers, distance to faults, and distance to roads. The comparison and validation of the outcomes of each model were achieved using statistical evaluation measures, the receiving operating characteristic, and the area under the success and predictive rate curves. Each model gave similar outcomes; however, the random forest model had a slightly higher predictive performance in terms of area under the curve (0.9220) against the ones estimated for the weight of evidence (0.9090) and the logistic regression model (0.8940). The same pattern of performance was reported when the success power of the models was calculated. Random forest was slightly better than the other two models in terms of area under the curve (0.9350) in comparison with the weight of evidence (0.9255) and logistic regression (0.9097). The predictive performance was estimated by using the validation dataset, while the success power of the models was estimated by using the training dataset. From the visual inspection of the produced landslide susceptibility maps, the most susceptible areas are located at the west and east mountainous areas, while moderate to low susceptibility values characterize the central area.

Keywords Landslide susceptibility · Shannon's entropy index · Logistic regression · Weight of evidence · Random Forest · China

Introduction

Landslides are considered as natural phenomena that are classified as a highly intense threat to human life, property, infrastructure, and natural environment observed mostly in mountainous and hilly regions. According to a report conducted during the FP7

SafeLand project (FP7 SafeLand 2012), China is covered by vast areas that are classified as regions with high landslide risk, while the manifestation of landslide phenomena results in an estimated 700 to 1000 deaths every year and damages of infrastructure and properties that exceed \$10 billion RMB annually. In order to reduce and mitigate the devastated consequences caused by landslides, the Chinese government has taken specific measures since 1999, such as, nationwide landslide investigation and risk zoning; detailed mapping for high risk zones of landslide hazards; stabilization and mitigation on major landslides; weather-based regional landslide hazard warning; geohazard risk assessment on infrastructure construction; and education and training for geohazard mitigation (FP7 SafeLand 2012). Analyzing the first two procedures, the determination of the spatial and temporal extent of landslide hazard requires to identify areas which are, or could be, affected by a landslide and estimate the probability of such landslide occurrence within a specified period of time. However, to specify the precise time frame for future occurrence of a landslide event, it can be a difficult task. As a result, landslide hazard can be represented by landslide susceptibility, if only the predisposing and preparatory landslide variables are considered. A landslide susceptibility map provides the spatial distribution and rating of the terrain according to its propensity to slide, the manifestation of which depends on the topography, geology, geotechnical properties, climate, vegetation, and anthropogenic factors (Fell et al. 2008). According to Guzzetti et al. (2000), a landslide susceptibility map is valuable when the information and data shown are useful, relevant, and fully understood by the user. In this context, the present study produces a landslide susceptibility map for the area of Nancheng County, China, in order to provide vital information concerning landslide phenomena to local authorities and government agencies for implementing appropriate decision-making and land use planning strategies.

In the last three decades, numerous methods and techniques have been utilized for landslide susceptibility and hazard and risk assessments; those methods could be classified into qualitative and quantitative or direct and indirect (Tien Bui et al. 2015). Qualitative methods are considered as methods that are characterized by their subjective nature, which ascertain susceptibility heuristically and mainly involve direct field geomorphological analysis and also the usage of index or parameter maps (Verstappen 1983; Leroi 1996; Soeters and Van Westen 1996). On the other hand, quantitative methods are based on numerical estimates and involve statistical, probabilistic, and data mining methods (Carrara et al. 1991; Van Westen et al. 1997; Castellanos Abella and van Westen 2007; Chowdhury 1976; Baldelli et al. 1996; Van Westen et al. 1997; Lees 1996; Gomez and Kavzoglu 2005). A great number of scientific research can be found that utilize bivariate statistics that has been adopted by many researchers (Magliulo et al. 2008; Yilmaz et al. 2012), as well as multivariate methods that implement

discriminant analysis (Lee et al. 2008) or linear and logistic regression (Dai and Lee 2003; Ayalew and Yamagishi 2005; Tsangaratos and Ilia 2016a), frequency ratio (Lee and Pradhan 2006; Yilmaz 2010; Akinci et al. 2011), certainty factor approach (Lan et al. 2004; Sujatha et al. 2012), and Dempster-Shafer and weight of evidence models (Tangestani 2009; Cervi et al. 2010; Neuhauser et al. 2012; Tien Bui et al. 2012a; Ilia and Tsangaratos 2016; Hong et al. 2016a). In addition, data mining methods have been applied for landslide susceptibility including fuzzy logic (Ercanoglu and Gokceoglu 2004; Pradhan et al. 2010; Akgun et al. 2012), artificial neural network (Pradhan and Lee 2009; Tien Bui et al. 2012b; Tsangaratos and Benardos 2014; Tien Bui et al. 2016), neuro-fuzzy (Vahidnia et al. 2010; Sezer et al. 2011; Pradhan 2013), and decision-tree models (Wan 2009; Yeon et al. 2010; Tien Bui et al. 2012c; Tsangaratos 2012; Pradhan 2013; Tsangaratos and Ilia 2016b). Both methods have been applied worldwide, and their performance is based on the availability and quality of data and the scale of analysis.

During a landslide susceptibility assessment, three important aspects should be successfully addressed in order to enhance the predictive power of landslide susceptibility models (Chacón et al. 2006; Irigaray et al. 2007; Costanzo et al. 2012; Tien Bui et al. 2015; Murillo-García et al. 2015): the preparation of a landslide inventory map, the identification of the variables that significant influence stability in ground surface, and the appropriate reclassification of the variables. The preparation of a landslide inventory map is based on a conceptual frame in which the past and present provide evidence for the future, failures do not occur randomly, failures share common geotechnical characteristics, and similar conditions produce similar patterns of failures (Tsangaratos and Koumantakis 2013). The second essential aspect is the identification of the influence of each variable contributes to the overall susceptibility that is expressed with a weighted coefficient that can be estimated through specific procedures according to different models. Finally, the determination of the classes for each variable is equally essential with the estimation of the coefficients, a procedure that could affect the quality of the outcomes of landslide susceptibility analysis (Chacon et al. 2006; Costanzo et al. 2012).

In this context, the present study attempts to address the above mentioned aspects by following a novel methodology. Specifically, landslide and non-landslide areas were verified by the usage of remote sensing techniques, Google Earth® and the analysis of high resolution digital elevation models, while the significance of each landslide-related variable was estimated by applying statistical and data mining methods that also produced a series of landslide susceptibility maps. However, the main novelty of the study is the determination of the number of classes for each landslide-related variable by estimating the information coefficient that is derived by the Shannon's entropy index. The developed methodology was tested in the Nancheng County, China, by applying three different methods: the Logistic Regression (LR), the Weight of Evidence method (WofE) as representatives of bivariate statistical methods, and Random Forest (RF) as a representative of data mining techniques. The usage of these three methods is considered to be appropriate since they are suitable for regional and semi-regional scale analysis and also they exploit both remote-sensing datasets and field surveys. The computational process was carried out using R Studio, SPSS 16.0 (SPSS 2007), while ArcGIS 10.1 (ESRI 2013) was used for compiling the data and producing the landslide susceptibility maps.

Materials and methods

Study area

The Nancheng County is located in the Eastern of the Jiangxi Province and is under the jurisdiction of the prefecture-level city of Fuzhou. The study area lies between longitudes 440,000 and 490,000 and latitudes 3,020,000 and 3,070,000 (Beijing 1954/3-Degree CM 117E as the reference coordinate system) covering an area of about 1698.3 km², with altitude ranging between 50 to 1180 m above sea level (Fig. 1).

Around 61.57 % of the study area has a slope gradient less than 15° whereas areas with a slope gradient larger than 45° account for only 0.39 %. About 25.38 % of the area is characterized by slope gradient between 15° and 25°, while 10.01 % is characterized by slope gradient between 25° and 35°. Dominant features in the area are the Hongmen reservoir and the Fuhe, Xuijian, and Latin river that flow across the research area. The waters of the Fuhe River reach the Poyang Lake that is located north of the Nanchang prefecture of Jiangxi.

The climate of Nancheng County is classified as humid subtropical (KöppenCfa), with long, humid, very hot summers, and cool and drier winters with occasional cold snaps. According to the Jiangxi Province Meteorological Bureau (<http://www.weather.org.cn>), the mean annual rainfall for the period 1953–2015 ranged between 900.3 and 2866.4 mm. The average annual temperature is 17.8 °C, while the average annual water surface evaporation for the area is estimated to be 1546.7 mm. The rainy season is from April to July accounting for the 55.2 % of the yearly rainfall. In May and June, the average rainfall varies between 270 and 305 mm per month.

Concerning the geological settings, more than 22 geologic groups and units are recognized, data was obtained by the China Geology Survey (<http://www.cgs.gov.cn>). In the present study, the lithology map (scale 1:200,000) was reconstructed by classifying the geological formations into nine classes, based on clay composition, degree of weathering, and physical and strength parameters (Table 1, Fig. 2). The main lithological unit that covers approximately 37 % of the area is granite porphyry of Cretaceous age, tuff, ignimbrite, and sandstone gravel (class E) followed by leptynite, schist, and marbles (class F) that covered 24 % of the area and gray brown granulite, mica schist, and quartz schist (class G) that covered 17 % of the area. The soil profiles of the area are mainly developed due to the action of weathering.

The developed methodology

The methodology followed during the present study could be separated into a four-phase procedure: (a) constructing the inventory map and selecting the appropriate landslide-related variables; (b) the data pre-processing phase; (c) the phase of implementing the various techniques and methods in order to construct the landslide susceptibility map; and (d) the validation and comparison of the models. Figure 3 illustrates the flowchart of the followed methodology, while a brief description of each phase is presented in the paragraphs below.

Constructing inventory map and selecting the landslide related variables

The first phase of the followed methodology was to construct the landslide and non-landslide inventory database. The database included information about the location, type of failure, and other features of landslide incidence and also the locations of non-landslide areas in order to use them during the training and predictive phase.

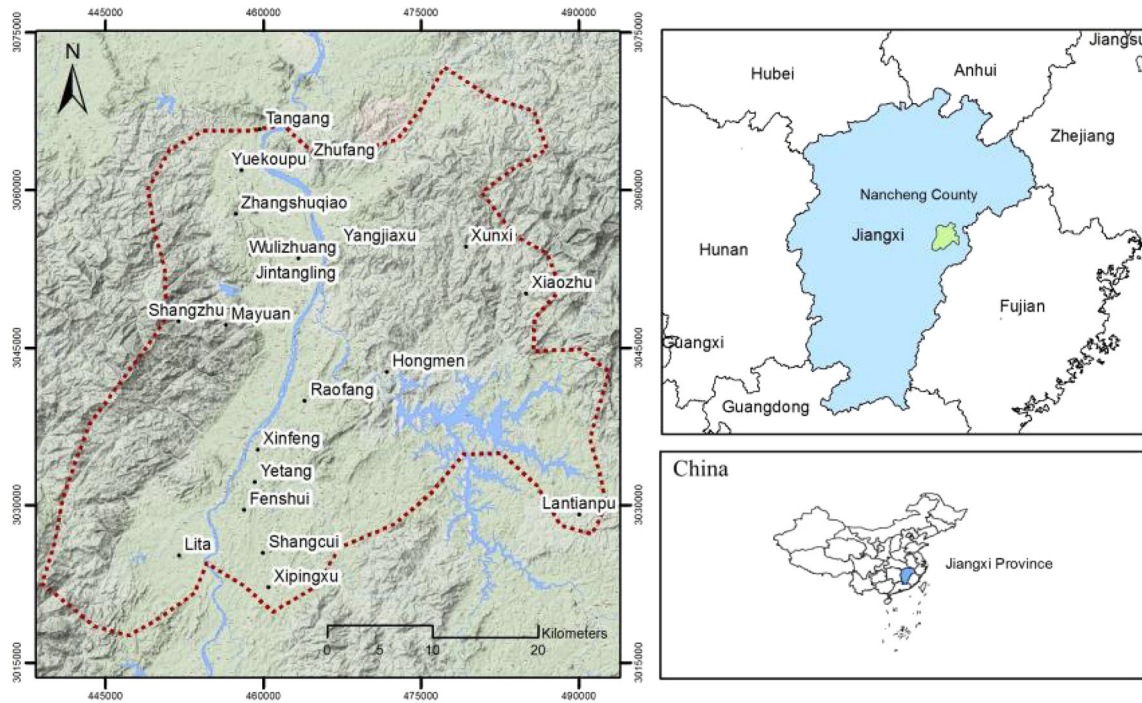


Fig. 1 The study area (Beijing 1954/3-Degree CM 117E as the reference coordinate system, suitable for use in China between 115° 30 E and 118° 30 E)

Specifically, the landslide inventory database which included 112 landslide locations was provided by the Jiangxi Department of Land and Resources (<http://www.jxgtt.gov.cn>) and the Jiangxi Meteorological Bureau (<http://www.weather.org.cn>). The database involved 70 rotational slides and 42 translational slides. The size of the smallest landslide is approximately 15 m², the largest around 18,000 m², and the average is estimated to be 878.7 m². The large-sized landslides (>1000 m²) that occurred in the study area affected 283 people, accounting for only 11.4 % of the total number of landslides. Around 33.1 % of the total landslides are medium-sized (200–1000 m²), and 213 people are affected by these landslides. Small-sized landslides (<200 m²) that affected 374 people are accounted for 55.4 % of the total landslides and are inventoried

on metamorphic rocks (schist, granulite, and marbles), having mass thickness between 3 and 8 m. According to the report in the Nancheng area, landslides occurred during and after incidence of heavy rainfall. Moreover, around 42.7 % of the landslides that were reported occurred when the measured rainfalls were around 95 mm per day whereas the other landslides occurred when the daily rainfall was larger than 110 mm.

The non-landslide areas were identified with the usage of Google Earth® and the analysis of high resolution digital elevation models. Google Earth® provides worldwide coverage of high resolution and very high resolution optical satellite images. Its main advance is that it can present the images into three dimension (3D), providing in that way an excellent tool for exploiting the

Table 1 Types of geological formation of the study area

Class	Unit name	Lithology
A	Zhonghanwu group	Tonalite rocks, granitic gneiss
B	Huang long, Chuanshan, Outangdi	Gravel sandstone, siltstone, limestone, dolomitic limestone
C	Xinyu	Purple siltstone, mudstone and fine sandstone, the central dark gray mudstone containing Glauber's salt, anhydrite, and halite
D	Lin shanqun, Jiuxiantan, Mu bushan, Xia changshan, Huang xie, Xi huashan	Coarse quartz sandstone, shale, coal line, gabbro, diabase
E	Huobashan, Nan keng, Shuijiang, He kou, Liu keng	Igimbrite, tuff, sandstone gravel, granite porphyry
F	Hongshan group,	granulite rock, schist, leptite, marble
G	Wanyuan group	Gray brown granulite, two-mica schist, quartz schist
H	Tan hu, Chenao, Fu fan	tonalite rock, porphyritic granodiorite, porphyritic monzonite granite, porphyritic moyite
I	Anyuan group	Sand, shale, chertconglomerate with coal seams, tuffaceous sandstone, tuff, tuffaceous mudstone, oil shale, olivine basalt

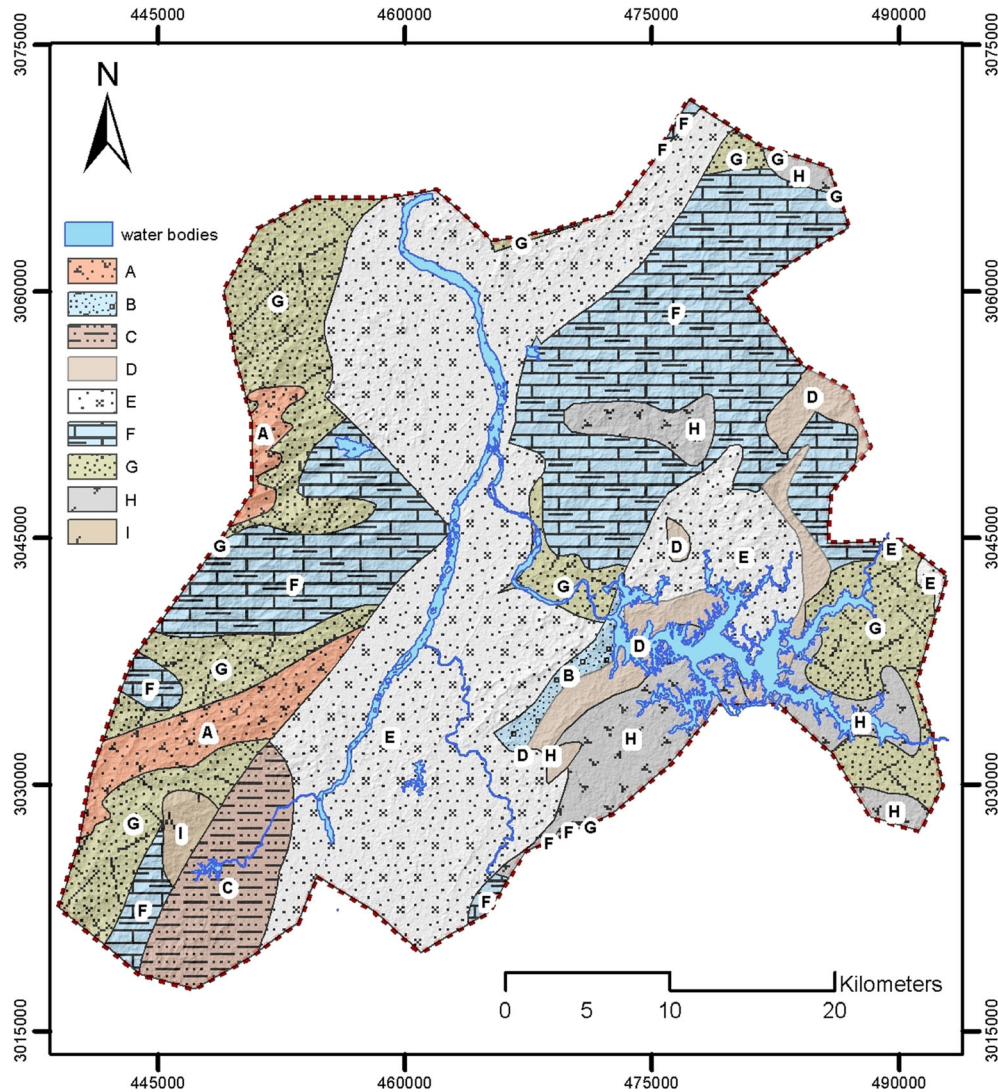


Fig. 2 The lithology map of the study

satellite images and detecting the non-landslide areas. The areas that are potentially classified as non-landslide areas are characterized by gentle and without any changes morphometric characteristic. The height difference, the steepness, and the orientation of slopes and also the absence of concavities and convexities are the main criteria for identifying the non-landslide areas.

Eleven (11) landslide-related variables were selected concerning the experience gained from studying landslide phenomena in the wider area, the local geo-environmental conditions, and the availability of sufficient data, namely, lithology, slope, aspect, altitude, topographic wetness index (TWI), sediment transport index (STI), plan curvature, profile curvature, distance to river network, distance to tectonic features, and distance to road network.

In order to extract the necessary layers that correspond to the morphometric landslide-related variables of slope, aspect, altitude, TWI, STI, plan curvature, and profile curvature, a digital elevation model (DEM) of grid size 25 m was used, generated from the Advanced Space borne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER

GDEM) Version 2 (<http://gdem.ersdac.jspacesystems.or.jp>). The ASTER GDEM Version 2, which was available for the public in 2011, is considered as the highest resolution DEM among the free accessible global DEMs having a spatial resolution of 30 m (Arefi and Reinartz 2011). The spatial product is a joint outcome developed by the Ministry of Economy, Trade, and Industry of Japan and the United States National Aeronautics and Space Administration that covers the entire land surface of the Earth. The road and river networks were digitized from 1:50,000 scale topographic maps.

Constructing the training and validation datasets

As proposed by the methodology training and validating, data sets were randomly produced from the total number of landslide and non-landslide areas. Specifically, by utilizing the subroutine subset wizard that is embodied in the Geostatistic toolbox (ESRI 2013), the first data set contained a number of data that equaled to approximately 70 % of the total number of landslide and non-landslide, while the rest 30 % served as validating data.

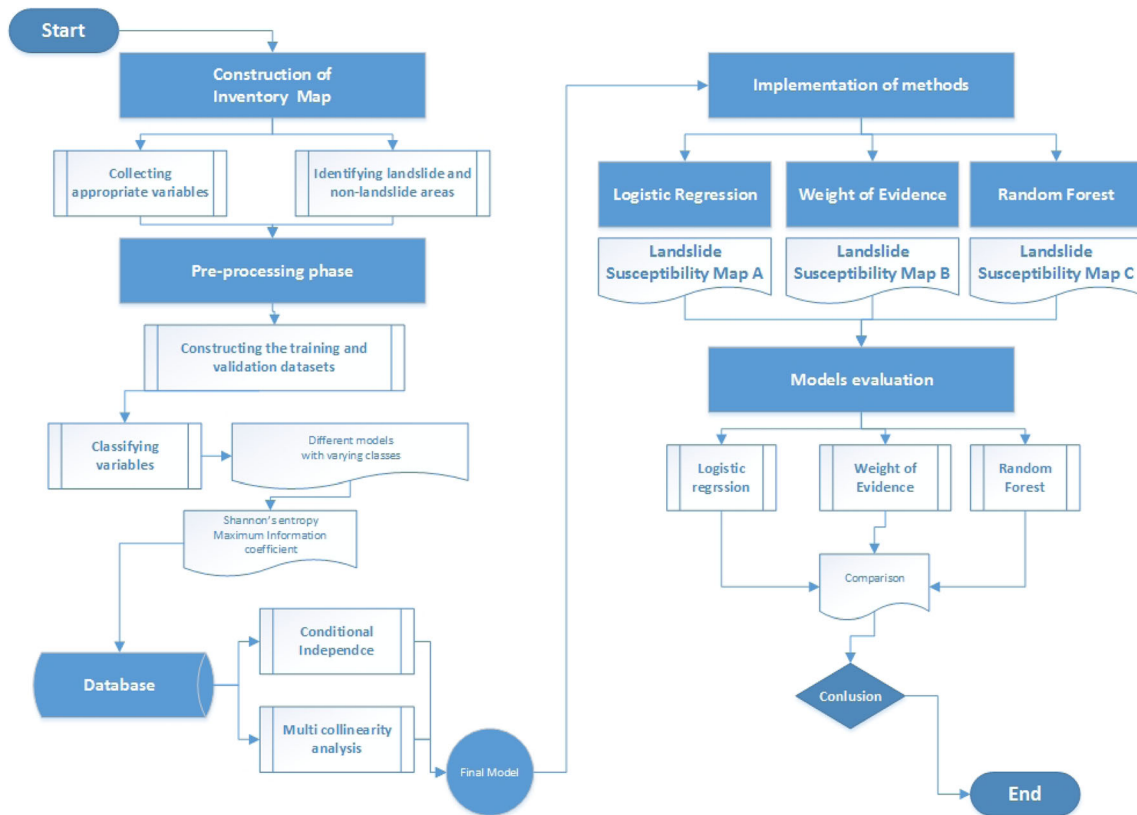


Fig. 3 Flowchart of the developed methodology

Information coefficient—number of classes

The next phase, the phase of pre-processing, involves the estimation of the exact number of classes that maximize the information coefficient of each variable based on Shannon’s entropy index. The Shannon’s entropy index has been used in the Information Theory as a measure originally proposed by Claude Shannon to quantify the entropy, disorder, uncertainty, or information content in strings of text (Shannon 1948). The entropy model introduced by Shannon has been used in several landslide assessments in order to estimate weighting coefficients of landslide-related variables (Yang and Qiao 2009; Yang et al. 2010; Pourghasemi et al. 2012). The model involves the calculation of the density of landslides, as done in bivariate analysis, within each class of each variable. The value of each variable is expressed as an entropy index, which indicates the extent of disorder in the environment. According to Bednarik et al. (2010), the entropy index expresses which variables are the most influential for the evolution of instability. The main difference between the present study and the approaches described in the aforementioned publications is that it estimates the information coefficient each time for a different number of classes and selects the one that maximizes the information coefficient of the variable in question. The information coefficient ranges between 0 and 1, with values closer to 0 indicating less information and values closer to 1 indicating more information. The equations used to calculate the information coefficient are given below (Bednarik et al. 2010; Constantin et al. 2011):

$$P_{ij} = \frac{L_{ij}}{A_{ij}} \tag{1}$$

$$P_{ij} = \frac{P_{ij}}{\sum_{j=1}^{c_j} P_{ij}} \tag{2}$$

$$H_j = -\sum_{i=1}^{c_j} P_{ij} \otimes \log_2 P_{ij} \tag{3}$$

$$H_{jmax} = -\log_2 c_j \tag{4}$$

$$I_j = \frac{H_{jmax} - H_j}{H_{jmax}} \tag{5}$$

where A_{ij} is the area percentage of the i^{th} class of the j^{th} variable, L_{ij} is the landslide percentage of the i^{th} class of the j^{th} variable, P_{ij} is the probability density of the i^{th} class of the j^{th} variable, H_j is the entropy value of the j^{th} variable, H_{jmax} is the entropy of the j^{th} having c_j classes, and I_j is the information coefficient of the j^{th} variable.

The reclassification process was performed by the Reclass subroutine using either Geometrical Intervals or Natural Breaks method (ERSI 2013). The choice of which to use is based on the type of distribution the data have. Specific, Geometrical Intervals was used for visualizing continuous data and providing an alternative to the Natural Breaks classification method. The specific benefit of the Geometrical Intervals method is that it works reasonably well on data that are not distributed normally, particular on data that are heavily skewed. On the other hand, Natural Break method is applied on normally distributed data.

Conditional independence and multicollinearity analysis

In order to implement WofE method, the conditional independency assumption among the landslide-related variables must be valid and the data population of each variable must have a normal distribution. According to Bonham-Carter (1994), this rough assumption may lead to errors and, in order to solve this problem, non-parametric statistics can be used since they are not based on the assumption of normal distribution. To calculate independency when applying non-parametric statistics, the χ^2 (chi-square) method can be used (Ilia and Tsangaratos 2016).

The next step is to implement the multicollinearity analysis in order to estimate the correlation among the predictor features (Dormann et al. 2013; Tien Bui et al. 2015). For this purpose, the proposed methodology uses the variance inflation factor (VIF) and tolerance (TOL) two important indexes for multicollinearity analysis (Marquardt 1970; Weisberg and Fox 2010). Although no rules exist for interpretation of VIF, the most common rule of thumb is using 10 as a threshold for severe multicollinearity, while several authors apply a very strict threshold of 2 or 5, above which variables are considered multicollinear and are excluded from the model (O'Brien 2007; Van Den Eeckhaut et al. 2006, 2010; Guns and Vanacker 2012; Tsangaratos and Ilia 2016a), while a value of TOL smaller than 0.1 indicates serious multicollinearity between independent variables (Menard 2002).

Implementing logistic regression

Logistic Regression is among those statistical methods that have been proved to be highly reliable when performing a landslide susceptibility assessment (Dai et al. 2002; Ayalew and Yamagishi 2005; Yesilnacar and Topal 2005; Gorsevski et al. 2006; Yilmaz 2010; Akgun et al. 2012). The independent variables in this model are considered as predictors of the dependent variable and can be measured on a nominal, ordinal, interval, or ratio scale, while the dependent variable is in a binary format. The relationship between the dependent variable and independent variables is nonlinear (Yesilnacar and Topal 2005).

LR is thought as a special case of a generalized linear model; however, it is based on quite different assumptions concerning the relationship between the dependent and independent variables from those followed by linear regression models. The conditional distribution is a Bernoulli distribution rather than a Gaussian distribution, since the dependent variable has the form of a binary variable (presence or absence of landslides).

In logistic regression analysis, the relationship between the occurrence and its dependency on several variables can be expressed by the following equation (Eq. (6)):

$$p = \frac{1}{1 + e^{-z}} \quad (6)$$

where p is the probability of a landslide occurrence. The probability can take values from 0 to 1 on an S-shaped curve and z is the linear combination of a set of landslide-related variables. Logistic regression involves fitting an equation of the following form to the data (Eq. (7)):

$$z(x_i) = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (7)$$

where b_0 is the intercept of the model, the b_i ($i = 0, 1, 2, \dots, n$) is the slope coefficients of the logistic regression model, and x_i ($i = 0, 1, 2, \dots, n$) are the independent variables. The linear model formed is then a logistic regression of presence or absence of landslides (present conditions) on the independent variables (pre-failure conditions).

Implementing weight of evidence

WofE is a data-driven approach that is based on the Bayes theorem and on the concepts of prior and posterior probability (Bonham-Carter 1994). There are numerous studies of landslide susceptibility analysis that utilize the WofE method (Lee et al. 2002; Lee et al. 2004; van Westen et al. 2003; Mathew et al. 2007; Bettian and Birgit 2007; Neuhäuser and Terhorst 2007; Poli and Sterlacchini 2007; Dahal et al. 2008; Sharma and Kumar 2008; Barbieri and Cambuli 2009; Ghosh et al. 2009; Tangestani 2009; Cervi et al. 2010; Ilia et al. 2010, 2013; Park 2010; Regmi et al. 2010; Armas 2012; Kayastha et al. 2012; Tien Bui et al. 2012a; Thiery et al. 2014; Kouli et al. 2014; Ilia and Tsangaratos 2016), in which the main objective is to estimate if a given set of independent variables could predict the presence of landslide incidence that is considered as the dependent variable. The method investigates the spatial relationship between the distribution of the areas affected by landslides and the distribution of the landslide-related variables (Ilia et al. 2010; Neuhäuser et al. 2012; Ilia and Tsangaratos 2016). A measure of the spatial association between landslide locations and landslide-related variables is provided through the magnitude of contrast (C), which is determined by the difference of positive (W+) and negative (W-) weights. W+ and W- provide information about whether there is a positive or a negative spatial correlation between the landslide-related variables and the landslide locations. When C is positive, it implies positive correlation, and when it is negative, it implies negative spatial association (Bonham-Carter et al. 1989; Agterberg et al. 1990). The studentized value of C is calculated as the ratio of C to its standard deviation stdC , (C/stdC), and serves as a guide to the significance of the spatial association, acting as a measure of the relative certainty of the posterior probability (Bonham-Carter 1994).

Implementing random Forest

Random Forest (RF) is an ensemble learning method, which is based on the generation of several classification trees, which are aggregated to estimate a classification (Breiman et al. 1984; Breiman 2001). The algorithm exploits random binary trees which use a subset of observations through bootstrapping techniques: from the original data set, a random selection of training data is sampled and used to build the model, the data not included are referred to as out-of-bag (OOB) (Breiman 2001). According to Hansen and Salamon (1990), an

ensemble method, such as RF, is more accurate than individual members if only data appear random and are diverse. In the case of RF, diversity is achieved by resampling the data with replacement and randomly changing the predictive factor over the different tree induction processes (Youssef et al. 2015).

One of the main advantages of RF is the ability to avoid over-fitting and growing a large number of random forest trees where it does not create a risk of over-fitting (e.g., each tree is a completely independent random experiment). The RF algorithm data does not need to be rescaled, transformed, or modified. It has resistance to outliers in predictors and automatically handles the missing values (Breiman and Cutler 2004).

Models validation and comparison

For the estimation of the performance of the three methods, two statistical evaluation criteria were utilized by using the training and validation data; the first one is the overall accuracy on the training data, which is an indication of the successful power of the model. The second one is the overall accuracy on the validation data, which is an indication of the predictive power of the model.

Both criteria are calculated as the ratio of the true positives plus the true negatives to the total number of data. The validation processes were achieved by using the receiver operating characteristic (ROC) curve analysis (Fawcett 2006). Using the landslide grid cells in the training dataset, the success-rate results were obtained, while the validation dataset was used for the construction of the prediction-rate curves (Chung and Fabbri 2003). The area under the ROC curve (AUC) has been used as a metric to assess the overall quality of the predictive models by evaluating the models ability to anticipate correctly the occurrence or non-occurrence of predefined events (Hanley and McNeil 1982; Negnevitsky 2002; Fawcett 2006). If AUC is close to 1, the outcomes of the analysis are excellent, while if the AUC is closer to 0.5, the less accurate the result of the analysis is.

In addition, the landslide density ratio was calculated as a measure of sufficiency (Can et al. 2005; Pradhan and Lee 2010). A model is more sufficient and accurate when there is an increase in the landslide density ratio when moving from low susceptible classes to high susceptible classes and when the high susceptibility class covers small extent areas.

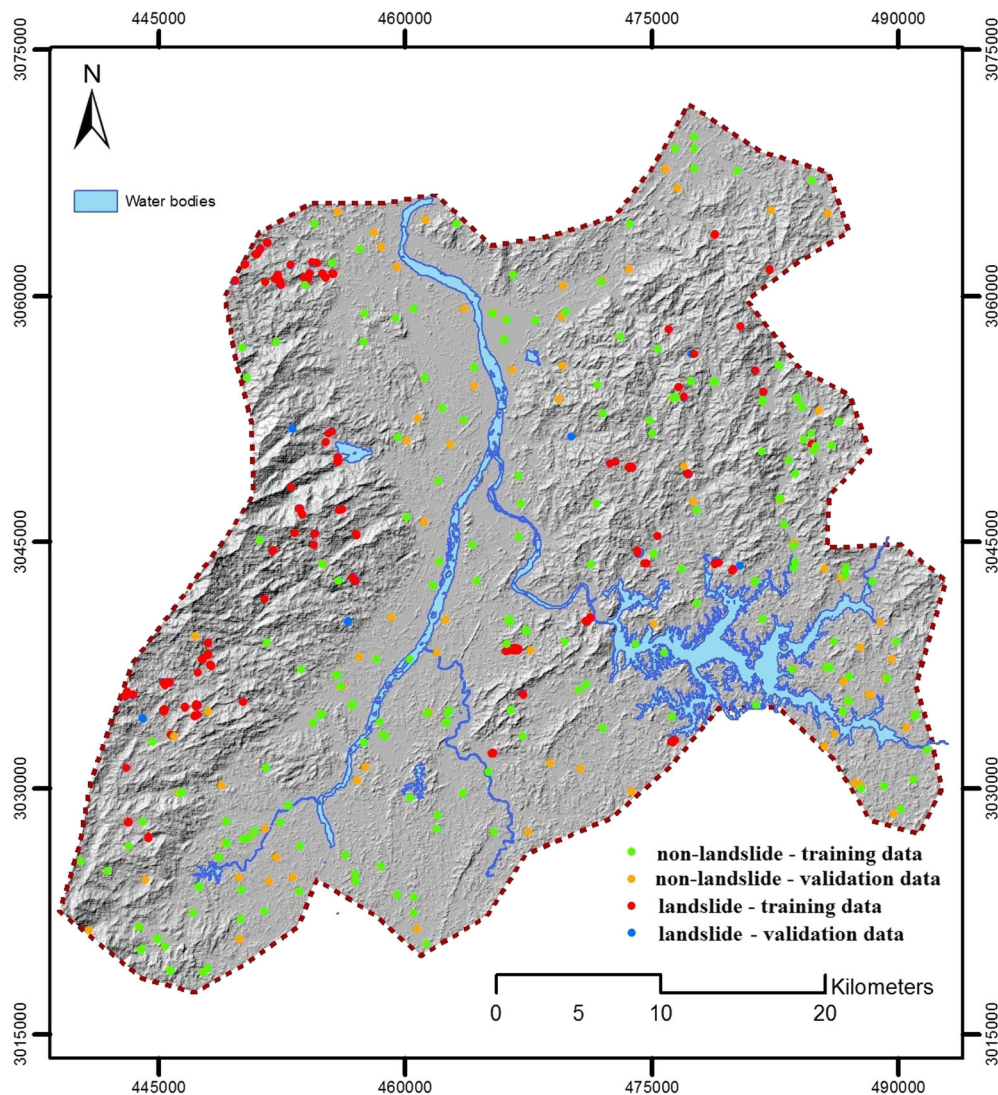


Fig. 4 The spatial distribution of non-landslide and landslide points

Results and discussion

Determining the class numbers of the landslide-related variables

Following the procedure described in the methodology, the landslide and non-landslide inventory database was constructed with the usage of Google Earth® and the analysis of the high resolution DEM. In order to capture representative information concerning the landslide-related variables, about the 112 landslide locations, additional points were introduced when necessary, especially when the landslide area had a large surface coverage, creating a total of 286 points. Equal number of 286 non-landslide points were identified, while by applying the subroutine subset wizard, approximately 70 % of the total number of landslide and non-landslide were used as training data and the rest 30 % served as validating data. Figure 4 illustrates the spatial distribution of the landslide and non-landslide areas.

The next action was to estimate the number of classes of each landslide-related variable that maximize the information coefficient based on Shannon's entropy index. The analysis was performed for two (2) to six (6) classes, for the each variable, except of the variable lithology that is a categorical variable. Table 2 provides the Information coefficient values for each class.

For the first variable, altitude, the information coefficient has the highest value, 0.3543, when classified by the Geometrical Interval classification method into two (2) classes. Similar, slope maximizes the information coefficient when classified into three (3) classes having a value of 0.1432. Aspect, which has been classified with the Natural Break method, also presents the highest information coefficient value (0.1123) when classified into three (3) classes, in comparison with those estimated when classified into a different number of classes. TWI and STI were classified by implementing the Geometrical Interval classification method and maximize the information coefficient when classified into two (2) classes, having values 0.0441 and 0.0466, respectively. Plan curvature, distance from river network, and distance from road network maximize its information coefficient when classified into four (4) classes, with values 0.0115, 0.0627, and 0.0902, while profile curvature when classified into six (6) classes, with information coefficient value, 0.0473. Finally, distance to tectonic features has the

highest value of information coefficient when classified into three (3) classes. Plan curvature and profile curvature were classified by using the Natural Break method, while distance from river network, distance from road network, and distance to tectonic features were classified by using the Geometrical Interval method. NC stands for non-calculable, meaning that for the certain classification, a class of the variable does not contain an incidence. Comparing the information coefficients among the variables, the most informative appears to be altitude, followed by slope and aspect, while the least informative appears to be the plan curvature. Figure 5a-j shows the spatial pattern of the classes that maximize the information coefficient for each of the landslide-related variables used in the analysis.

Multi-collinearity analysis

The VIF's and tolerance values (TOL) were estimated by performing multicollinearity analysis (Table 3). According to the results, there was no serious multicollinearity between the independent variables. The smallest TOL was the one calculated for the plan curvature variable (0.405) which however is higher than 0.100 the theoretical critical value for evidence of collinearity (Menard 2002). Also, the VIP's values for all the variables are less than 5, a similar theoretical threshold of multicollinearity (O'Brien 2007; Van Den Eeckhaut et al. 2006, 2010; Guns and Vanacker 2012; Tsangaratos and Ilia 2016a).

Applying logistic regression method

The training dataset was evaluated using a chi-square of Hosmer-Lemeshow test, Cox and Snell R^2 , and Nagelkerke R^2 , while accuracy percentages of classification for all training sets were also calculated. Hosmer-Lemeshow test showed that the goodness of fit of the equation can be accepted since the significance of the chi-square is larger than 0.05 (Table 4).

The overall precession and recall index of the classification is 82.5 %, which is quite acceptable. The logit of $f(x)$ function is calculated for all of the grids of the Nancheng County, in which zero (0) corresponds to no susceptibility and one (1) to total susceptibility. Based on constant values that were calculated, the logistic regression is compiled according to Eq. (8) as follows:

Table 2 Information coefficients

Thematic layer (classification method)	2 classes/ information coefficient	3 classes/ information coefficient	4 classes/ information coefficient	5 classes/ information coefficient	6 classes/ information coefficient
Altitude (Geometrical Intervals)	0.3543	0.3028	0.2396	0.1919	NC
Slope (Geometrical Intervals)	0.0996	0.1432	0.0991	0.0842	0.1063
Aspect (Natural Break)	0.0012	0.1123	0.1051	0.1042	0.0700
TWI (Geometrical Intervals)	0.0441	NC	NC	NC	NC
STI (Geometrical Intervals)	0.0466	0.0340	NC	NC	NC
Plan curvature (Natural Break)	0.0018	0.0035	0.0115	0.0080	0.0091
Profile curvature (Natural Break)	0.0037	0.0076	0.0336	0.0355	0.0473
Distance to river network (Geometrical Intervals)	0.0112	0.0002	0.0627	0.0318	0.0207
Distance to tectonic features (Geometrical Intervals)	0.0587	0.0807	NC	NC	NC
Distance to road network (Geometrical Intervals)	0.0793	0.0611	0.0902	0.0543	NC

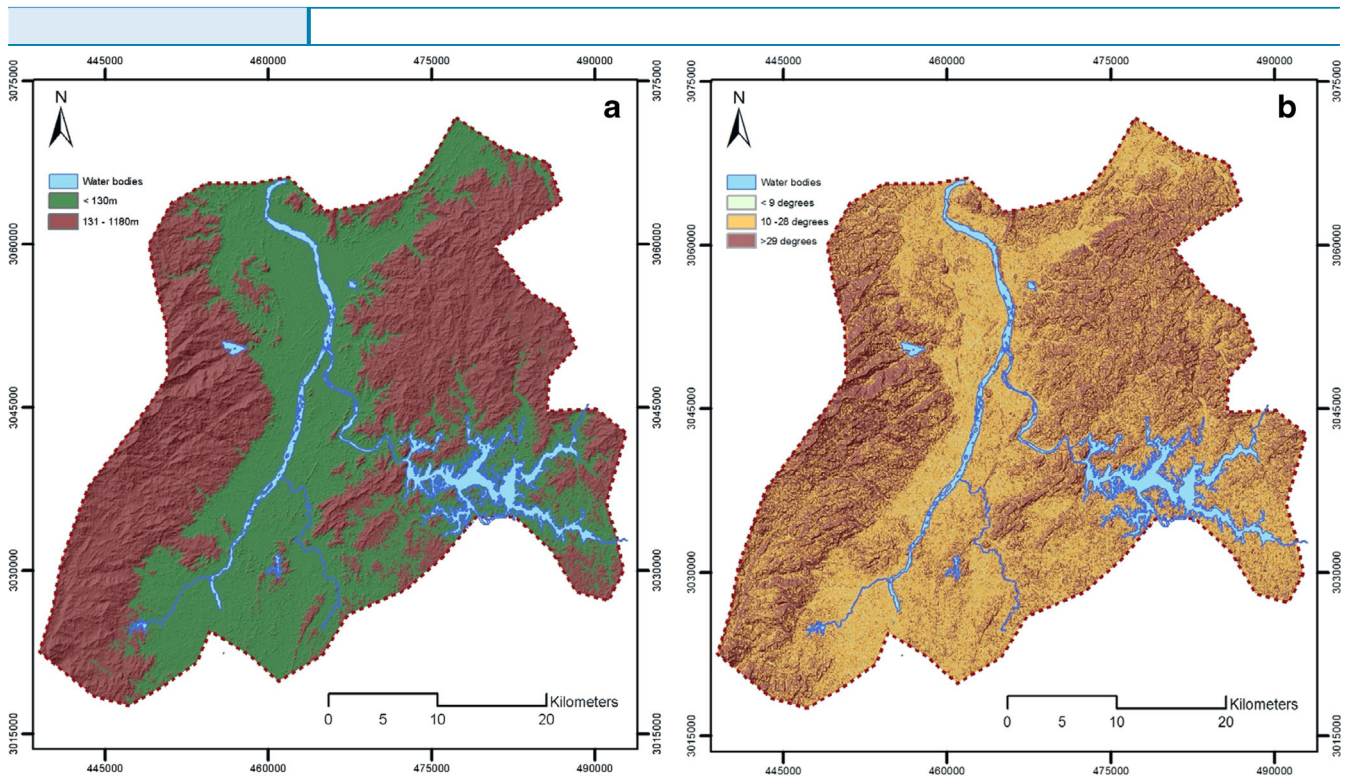


Fig. 5 The landslide-related variables. **a** Altitude, **b** slope angle, **c** aspect, **d** TWI, **e** STI, **f** plan curvature, **g** profile curvature, **h** distance to river network, **i** distance to tectonic features, **j** distance to road network

$$\begin{aligned}
 z = & -5.0789 + [0.3079*(Lithology)] + [2.0617*(Altitude)] \\
 & + [0.8983*(Slope)] + [0.0627*(Aspect)] + [-2.0718*(TWI)] \\
 & + [2.2334*(STI)] + [-0.084*(Profile Curvature)] \\
 & + [-0.2199*(Plan Curvature)] \\
 & + [0.4515*(Distance to River Network)] \\
 & + [0.2376*(Distance to Tectonic Features)] \\
 & + [-0.4703*(Distance to Road Network)] \quad (8)
 \end{aligned}$$

In order to predict the possibility of landslide occurrence in each grid, probability was calculated from Eq. (6) and the landslide susceptibility map was produced (Fig. 6).

The conditional variables lithology, altitude, slope, aspect, STI, distance to river network, and distance to tectonic features affect the LR function positively, while the highest b coefficient according to Eq. (8) is allocated to STI and altitude, which are 2.2334 and 2.0617, respectively. TWI, profile and plan curvature, and distance to road network have a negative effect on the landslide occurrence as they have negative b coefficients. From the visual analysis of the landslide susceptibility map, high and very high susceptible zones are located at the west and east mountainous areas, while the central area is characterized by very low to low susceptibility values. It is clear that the spatial pattern of the landslide susceptibility follows the distribution of the elevation and slope observed in the study area, since lowlands are characterized by very low to low susceptibility values. One can also observe a strong association between the lithological coverage and the landslide susceptibility values.

Applying weight of evidence method

The estimation of the conditional independency among the landslide-related variables was performed by the chi-squared statistic test. Table 5 illustrates the results of the chi-squared test on the observed distribution and expected distribution of the landslide occurrence based on posterior probabilities calculated using the 11 variables. The theoretical χ^2 values are presented in brackets. From the total of 55 pairwise comparisons, 14 conditional dependencies have been identified at a 0.01 significance level and varying degrees of freedom. Specifically, TWI showed six (6) conditional dependencies, while distance to road network showed four (4) conditional dependencies. Despite the observed conditional dependence among some of the variables, it was decided to proceed in the analysis in order to compare the three models under the same settings.

The next phase was to calculate the weights of the landslide-related variables according to the methodology of the WofE method. Table 6 provides the C values that are used to construct the landslide susceptibility map through an aggregated weighted method and also the stdC and C/stdC values. Ranking the positive spatial correlation between the classes of the landslide-related variables and the landslide locations, areas that have elevation greater than 131 m exhibit the highest C value (1.6230), followed by areas that have TWI values less than 5.85 (1.6179). Concerning the lithological formation of the research area, the Wan Yuan group that consists of granulite and mica-quartz schist is found to have the highest C value (1.4038), while areas that have an orientation between 109° and 228° have moderate C values (1.0003).

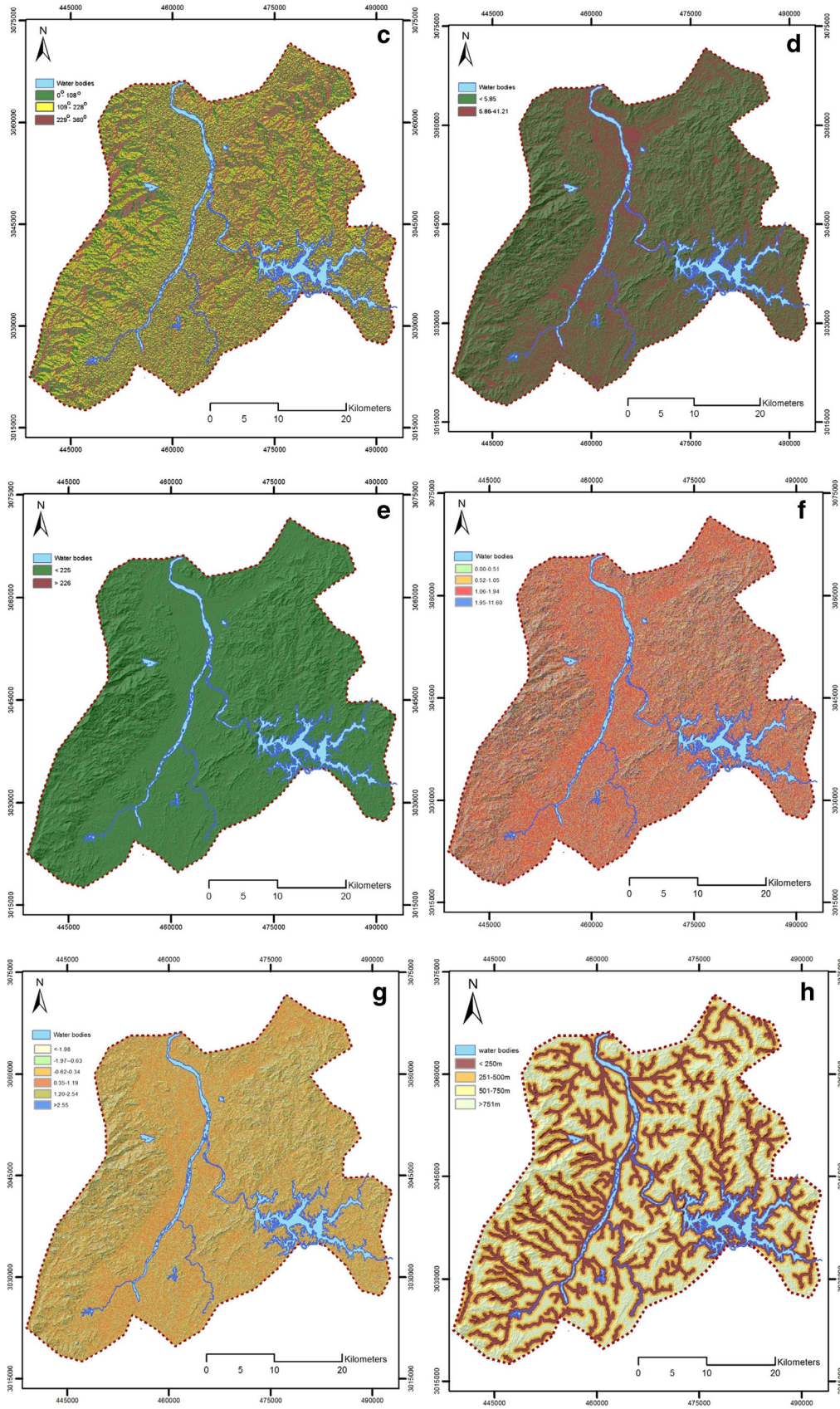


Fig. 5 (continued)

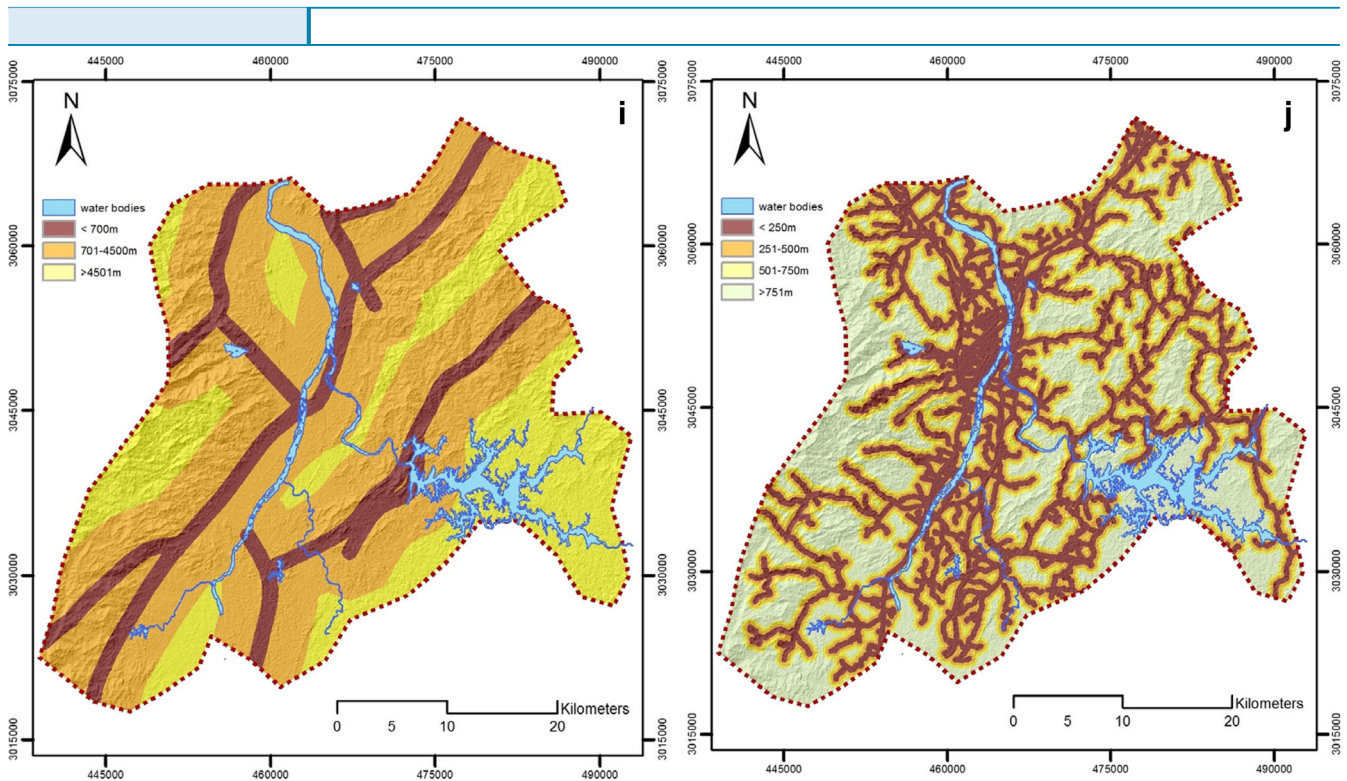


Fig. 5 (continued)

Figure 7 shows the landslide susceptibility map constructed by WoFe method. A similar pattern of landslide susceptibility values to the LR method was observed, with high and very high susceptible areas to be located at the west and east mountainous areas, while the central area is characterized by moderate to low susceptibility values.

Applying random Forest method

To implement successively the RF method, there is a need to estimate the minimum number of trees required to minimize the

Out-Of-Bag error and also the need to estimate the number of variables randomly sampled as candidates at each split. As illustrated in Fig. 8a, the Out-Of-Bag error (black line) is less fluctuated when the number of trees exceeds 800, while Fig. 8b gives the results of the tuning process concerning the number of variables used in each split. It was decided to train the RF model using two (2) random variables at each split and 1000 trees.

After the training phase ended, some extra information about the influence of each variable has on the overall landslide susceptibility analysis followed by the RF method was gained. Specifically, Fig. 9 illustrates the 11 variables ordered by the mean decrease accuracy and the mean decrease Gini. The mean decrease in Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest model, while the mean decrease in accuracy a variable causes is determined during the Out-Of-Bag error calculation phase. The more the accuracy of the random forest decreases due to the exclusion of a variable, the more important that variable is assumed, thus variables with a large mean decrease in accuracy are more important. According to those two metrics, the most important variable is altitude followed by TWI and lithology.

In Table 7, the variables that are more often used during the training phase are reported. The most used variables were lithology (13.46 %), plan curvature (13.22 %), and distance to river network (12.93 %) followed by distance to road network (12.56 %), aspect (11.27 %), and distance to tectonic features (10.25 %). Profile curvature (9.77 %), TWI (5.61 %), slope (5.02), and altitude (4.91 %) are the least used, while STI only participates in the model 0.89 % of the total number of times each variable was used.

Table 3 Multi collinearity analysis

Variable	Collinearity statistics	
	TOL	VIF
Lithology	0.821	1.218
Altitude	0.634	1.578
Slope	0.665	1.505
Aspect	0.846	1.182
TWI	0.486	2.056
STI	0.879	1.138
Plan curvature	0.405	2.470
Profile curvature	0.558	1.791
Distance to river network	0.889	1.125
Distance to tectonic features	0.763	1.311
Distance to road network	0.740	1.351

Table 4 The overall statistics of the logistic regression

Hosmer-Lemeshow test					
Chi-square	df	sig	-2log likelihood	Cox and Snell R^2	Nagelkerke R^2
10.698	8	0.2194	295.69	0.48	0.64

Figure 10 illustrates the landslide susceptibility map constructed according to the RF method. From the visual analysis of the landslide susceptibility map, it seems that it follows the pattern of altitude, lithology, and the distance to river network. High and very high susceptible zones are located along the road network mainly at the west and east mountainous areas, while the central area is characterized by very low to low susceptibility values.

Insights about the influence the landslide-related variables have in predicting the stability condition of the research area have been obtained by the implementation of the three models. Specifically, RF and WofE considered altitude, lithology, and TWI as the most important variables. LR identifies altitude and lithology as affecting the LR function positively, while TWI affects the LR function negatively. Concerning the altitude of a surface, it could be considered as a variable that indirectly contributes to

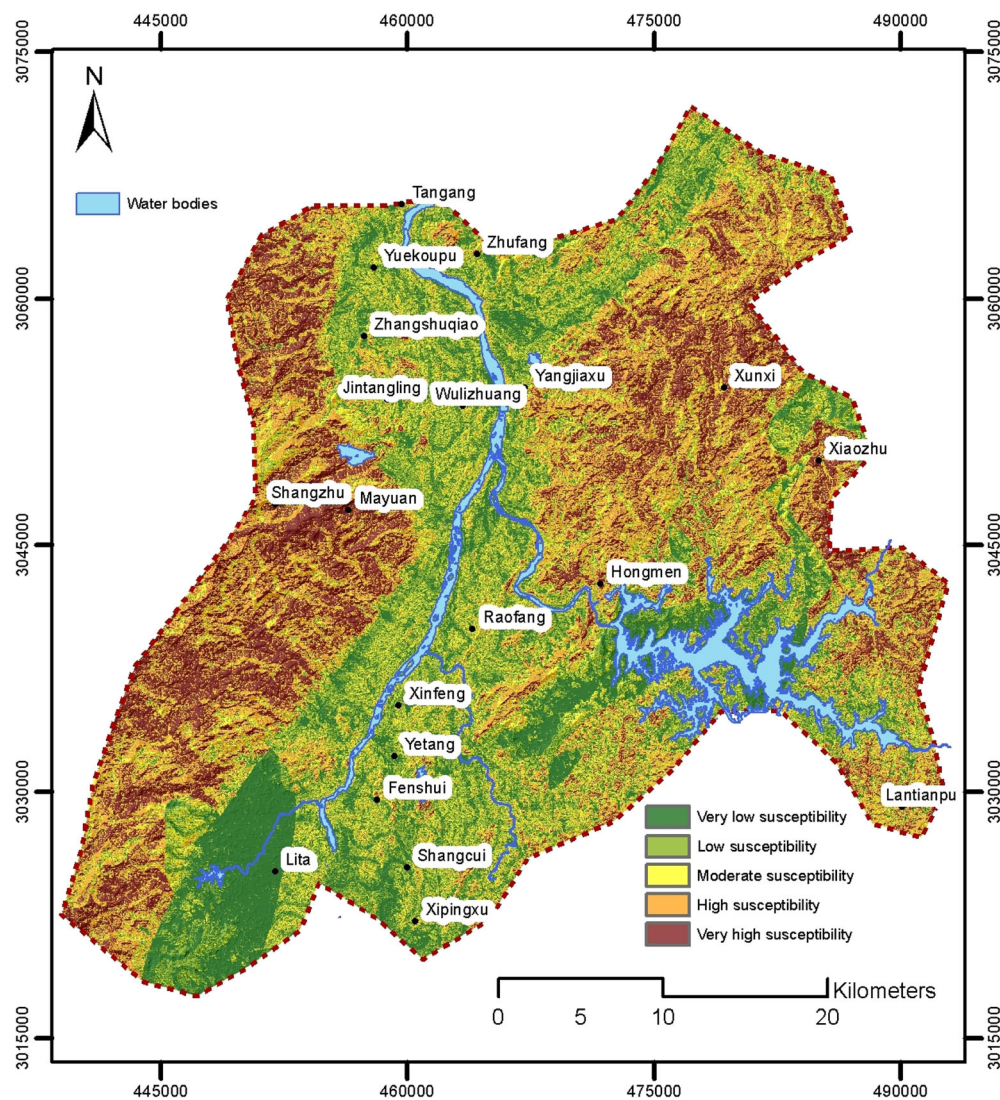
**Fig. 6** Landslide susceptibility map produced by the LR method

Table 5 Test of conditional independence

Variables	A	B	C	D	E	F	G	H	I	J
Lithology	76.78 (15.51)	4.53 (26.30)	10.88 (26.30)	6.67 (15.51)	4.28 (15.51)	12.60 (36.41)	20.98 (55.76)	20.09 (36.41)	25.95 (26.30)	76.14 (36.41)
A	–	17.10 (5.99)	0.65 (3.84)	5.41 (3.84)	0.69 (3.84)	4.07 (7.81)	7.68 (11.07)	4.26 (7.81)	4.54 (5.99)	12.45 (7.81)
B	–	–	1.47 (9.49)	22.37 (5.99)	0.05 (5.99)	4.63 (12.59)	18.00 (18.31)	2.51 (9.49)	1.11 (5.99)	9.20 (12.59)
C	–	–	–	1.68 (5.99)	1.10 (5.99)	8.70 (12.59)	11.91 (18.31)	16.14 (12.59)	2.16 (9.49)	0.98 (12.59)
D	–	–	–	–	8.48 (3.84)	28.41 (7.81)	29.50 (11.07)	8.35 (7.81)	2.23 (5.99)	0.89 (7.81)
E	–	–	–	–	–	2.89 (7.81)	5.38 (11.07)	2.78 (7.81)	0.47 (5.99)	2.18 (7.81)
F	–	–	–	–	–	–	94.46 (24.30)	3.11 (16.91)	2.78 (12.59)	12.68 (16.91)
G	–	–	–	–	–	–	–	10.24 (24.30)	6.55 (18.31)	8.07 (24.30)
H	–	–	–	–	–	–	–	–	15.78 (12.59)	36.40 (16.91)
I	–	–	–	–	–	–	–	–	–	16.35 (12.59)

A altitude, B slope, C aspect, D TWI, E STI, F plan curvature, G profile curvature, H distance to river network, I distance to tectonic feature, J distance to road network.

the slope failure manifestation (Dai et al. 2002). The elevation of a surface is considered to be formed by the combined action of tectonic activity, weathering, and erosion processes and is also related with the action of the climatic conditions through a complex interactive influence. The analysis performed in our study showed that areas with elevations greater than 131 m experience considerable higher chance of landslide occurrence. Regarding lithology, the high percentage of small-sized landslides was observed in areas covered by metamorphic rocks (schists, granulite, and marbles) having mass thickness between 3 and 8 m. However, it should be mentioned that the scale of the available lithological map makes it difficult to distinguish in more detail the overlying lithology. Quaternary deposits that may be present are not mapped and thus the types of landslides observed are not associated with those types of geological formations. This issue should be addressed as key research point in the close future. Finally, concerning the TWI, that is an index to describe the effect of topography on the location and size of saturated source areas of runoff generation (Moore et al. 1991), the analysis revealed that areas less saturated exhibit higher landslide susceptibility.

Validation and comparison

The next phase of the followed methodology was to estimate the relative distribution of the landslide susceptibility zones and the landslide density for each of the three methods. All models showed an increasing landslide density ratio when moving from low susceptible classes to high susceptible classes (Fig. 11).

However, the WofE method showed the highest density (0.7740), followed by the LR method (0.6739) and the RF method (0.4284) in the very high susceptible zone. The percentage of landslides found in the very high susceptible zone for WofE, LR, and RF was estimated to be 77.97, 67.13, and 41.25 %, respectively, while the percentage of the area classified as very high susceptibility according to WofE, LR, and RF was estimated to be 28.64, 19.80, and 18.46 % of the total research area, respectively.

According to the methodology, the validation of the three methods was estimated by calculating the successive and predictive power on the bases of the training and validation dataset. Figure 12 illustrates the area under the ROC curve (AUC) that expresses the models ability to anticipate correctly the occurrence or non-occurrence of landslides for the three models. The highest train AUC value was obtained by the RF method (0.9350) followed by the WofE (0.9255) and the LR method (0.9097). The highest predictive ROC curve with AUC values equal to 0.9220 was again achieved by the RF method followed by the WofE (0.9090) and LR method (0.8940).

Our findings are consistent with the results from similar comparative studies. More specifically, as reported in a landslide susceptibility analysis presented by Esposito et al. (2014) which compared the outcomes of a RF model with a LR model in Rio de Janeiro, Brazil, the RF model showed higher accuracy than the LR model, with AUC values estimated to be 0.81 and 0.72, respectively. Similar results of higher accuracy were also reported in a comparative landslide susceptibility study, indicating the RF

Table 6 The weights of the landslide-related variables based on the WofE method

Thematic Layer	Classes	W+	W−	C	stdC	C/stdC
Lithological Units	A	0.0411	−0.0017	0.0428	0.3609	0.1185
	B	−5.1169	0.0083	−5.1252	10.0003	−0.5125
	C	−6.8026	0.0460	−6.8486	10.0003	−0.6848
	D	−6.6541	0.0395	−6.6937	10.0003	−0.6693
	E	−0.8477	0.2931	−1.1408	0.1929	−5.9146
	F	0.2335	−0.0894	0.3229	0.1529	2.1121
	G	0.9778	−0.4261	1.4038	0.1419	9.8949
	H	−0.8520	0.0425	−0.8945	0.4145	−2.1579
	I	−4.9944	0.0074	−5.0018	10.0003	−0.5002
Altitude (m)	<130 m	−1.1216	0.5013	−1.6230	0.1929	−8.4142
	>131 m	0.5013	−1.1216	1.6230	0.1929	8.4142
Slope Angle (degrees)	0–9	0.2880	−0.0230	0.3110	0.2536	1.2265
	10–28	−0.2751	0.2667	−0.5418	0.1430	−3.7874
	>29	0.2630	−0.2006	0.4636	0.1415	3.2777
Aspect (degrees)	0–108	−0.4396	0.1587	−0.5983	0.1736	−3.4462
	109–228	0.5485	−0.4518	1.0003	0.1430	6.9930
	229–360	−0.4636	0.1762	−0.6398	0.1721	−3.7174
Distance from tectonic features (m)	<700 m	−0.9319	0.1034	−1.0353	0.2978	−3.4770
	701–4500 m	0.0732	−0.1026	0.1758	0.1447	1.2155
	>4501 m	0.1604	−0.0720	0.2325	0.1498	1.5516
Distance from river network (m)	<250 m	0.0979	−0.0568	0.1547	0.1453	1.0644
	251–500 m	0.0941	−0.0327	0.1267	0.1593	0.7956
	501–750 m	−1.0651	0.1418	−1.2069	0.2868	−4.2076
	>751 m	0.2568	−0.0844	0.3413	0.1575	2.1669
Distance from road network (m)	<250 m	0.3880	−0.2378	0.6258	0.1420	4.4072
	251–500 m	0.4033	−0.1390	0.5423	0.1529	3.5466
	501–750 m	−0.5216	0.0657	−0.5873	0.2536	−2.3163
	>751 m	−0.8213	0.2545	−1.0758	0.1980	−5.4324
Plan curvature	0.00–0.51	0.1323	−0.0152	0.1475	0.2260	0.6524
	0.52–1.05	−0.1377	0.0504	−0.1881	0.1633	−1.1518
	1.06–1.94	−0.0497	0.0392	−0.0889	0.1428	−0.6221
	1.95–11.60	0.2422	−0.0556	0.2978	0.1736	1.7152
Profile curvature	<−1.98	0.8230	−0.0540	0.8770	0.2471	3.5492
	−1.97–0.63	−0.0280	0.0052	−0.0332	0.1954	−0.1699
	−0.62–0.34	−0.0576	0.0210	−0.0786	0.1612	−0.4877
	0.35–1.19	−0.0910	0.0378	−0.1288	0.1575	−0.8178
	1.20–2.54	−0.1996	0.0398	−0.2394	0.1980	−1.2088
	2.55–16.23	0.5951	−0.0307	0.6258	0.2868	2.1818
TWI	0–5.85	0.4740	−1.1439	1.6179	0.1980	8.1699
	5.86–41.21	−1.1439	0.4740	−1.6179	0.1980	−8.1699
STI	<225	−0.0101	0.5069	−0.5170	0.45294	−1.14150
	>226	0.5069	−0.0101	0.5170	0.45294	1.14150

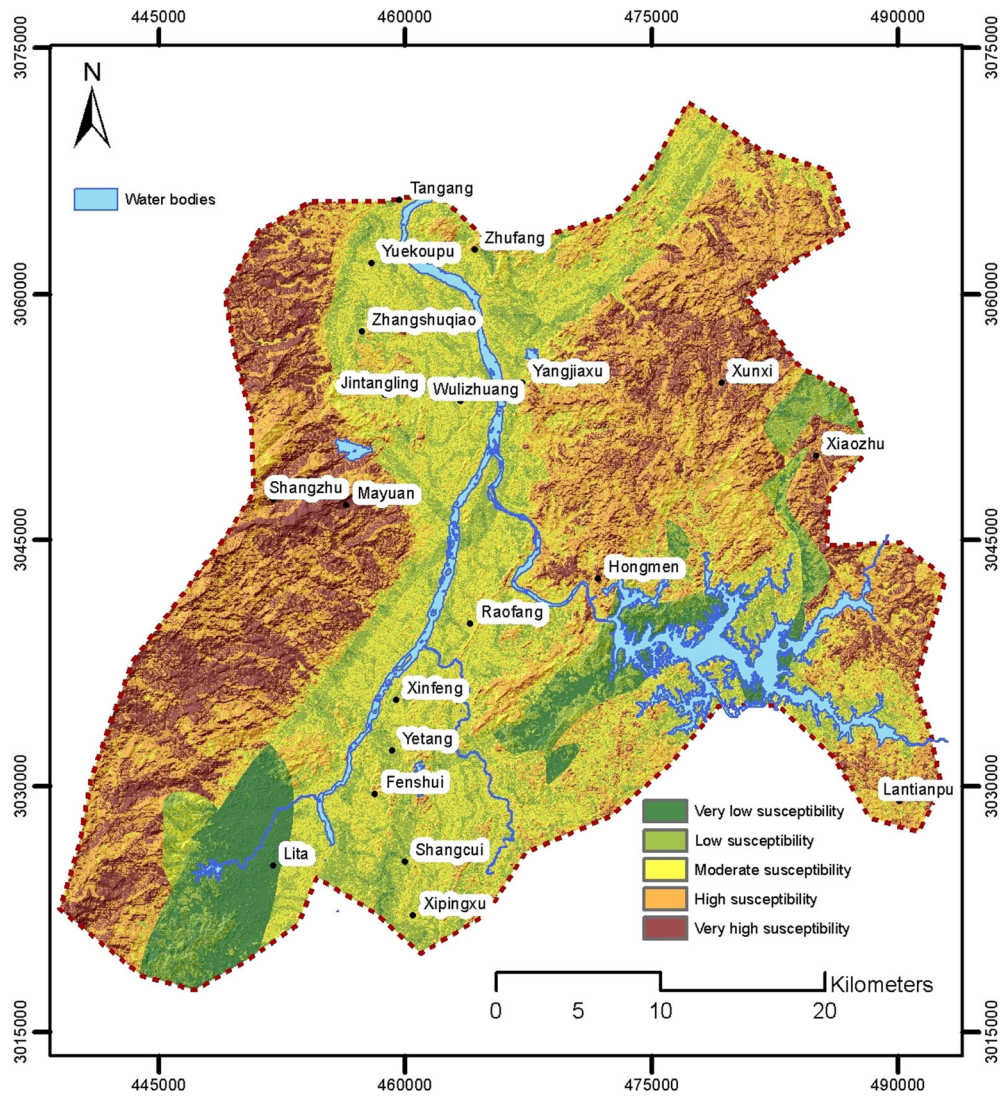


Fig. 7 The landslide susceptibility map produced by the WoFE method

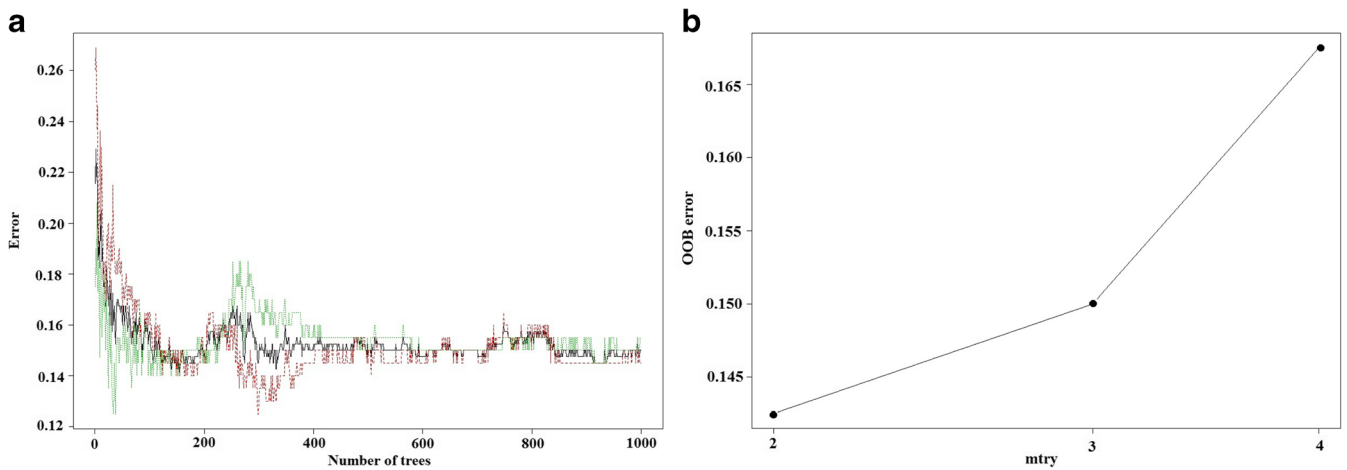


Fig. 8 a Error OOB vs number of trees, b number of variables used in each split (mtry)

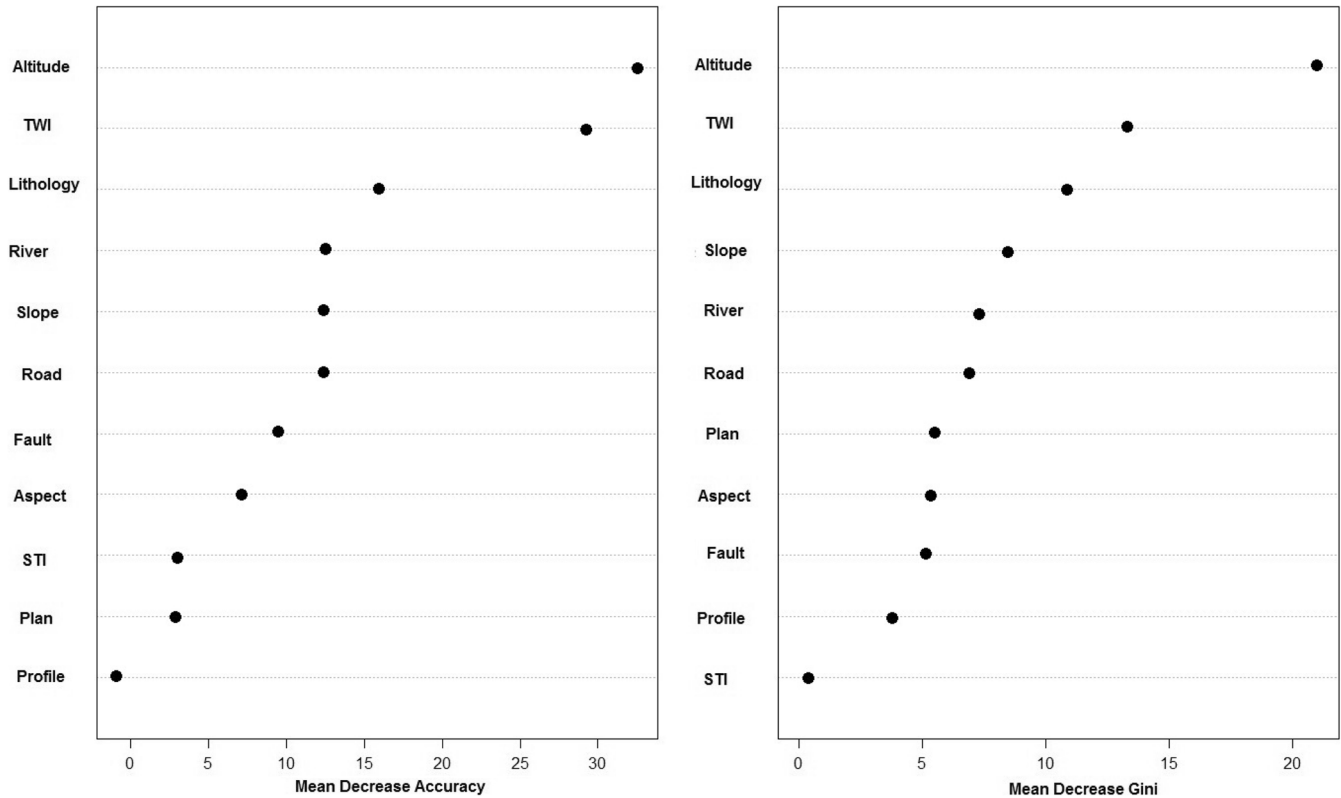


Fig. 9 Mean decrease accuracy and mean decrease Gini

model as the most accurate against a LR model and a frequency ratio model (Trigila et al. 2015). Also, Goetz et al. (2015) reported that RF model had a slightly better performance in a landslide susceptibility assessment conducted in Austria, when compared with LR, WofE, and other advanced data mining techniques. In

contrast to the above studies, findings of a landslide susceptibility analysis held in Lianhua County, China, reported however the poor performance of RF model when compared with a data driven evidential belief function, a frequency ratio and a LR model (Hong et al. 2016b).

In any case, understanding the abilities and limitation of each method remains critical for selecting the most accurate model (Goetz et al. 2015). The b coefficients of the LR function are able to provide an estimate of the importance of each variable plays in explaining the presence of landslide; however, they do not provide information about the relative priorities or importance among the predictive variables. In WofE method, the conditional independency assumption among the landslide-related variables must be valid, while the data population of each variable must have a normal distribution. On the other hand, RF has several advantages; it does not require assumptions on the distribution of explanatory variables, it allows for the use of either categorical or numerical variables, it accounts interactions and nonlinearities among variables, and its ability to provide information about the influence of each variable on the overall result (Catani et al. 2013; Pourghasemi and Kerle 2015).

Conclusions

The present study presents a novel methodology in which Shannon's entropy model was used for classifying landslide-

Table 7 Number of times variables were used when applying RF

Variables	Number of times used by the model	Percentage (%)
Lithology	1827	13.46
Altitude	667	4.91
Slope angle	682	5.02
Aspect	1530	11.27
TWI	762	5.61
STI	121	0.89
Profile curvature	1327	9.77
Plan curvature	1808	13.32
Distance to river network	1756	12.93
Distance to tectonic features	1392	10.25
Distance to road network	1705	12.56

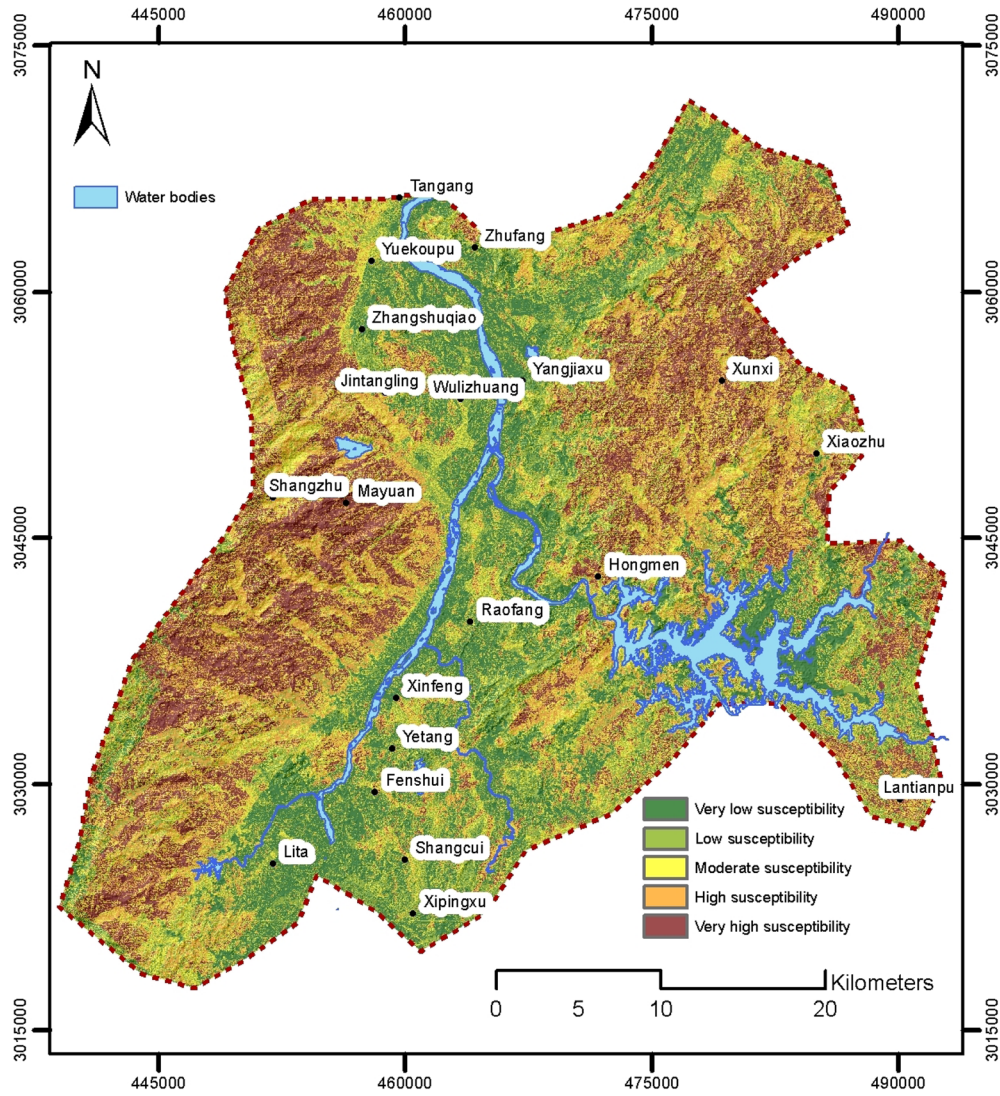


Fig. 10 The landslide susceptibility map produced by the RF method

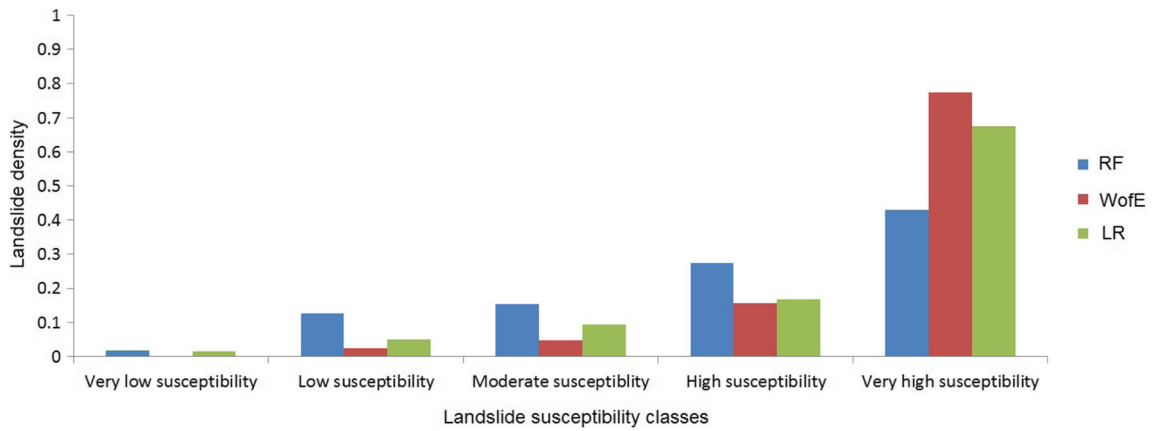


Fig. 11 Bar graphs showing the relative distribution of landslide susceptibility zones and landslide density

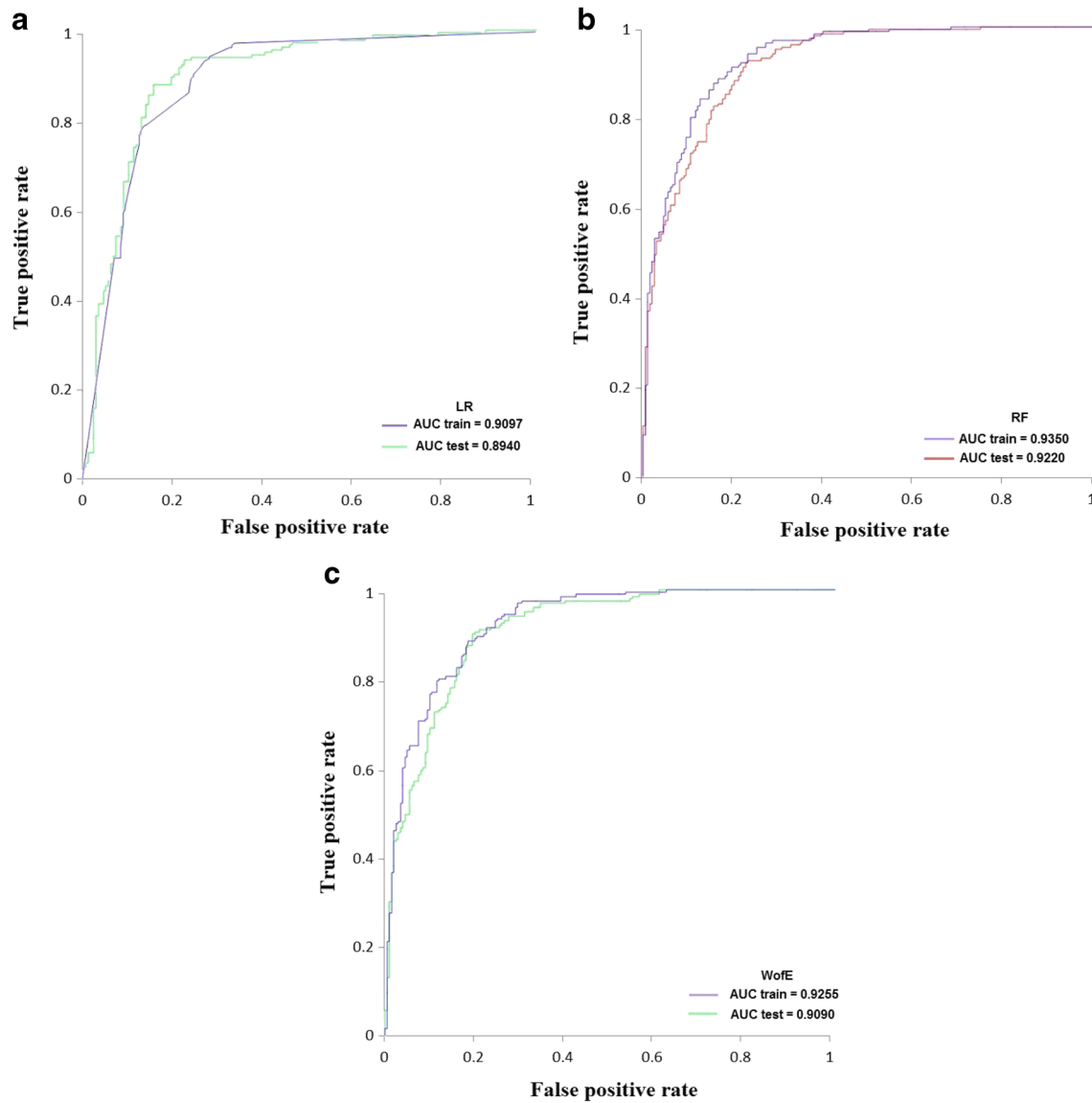


Fig. 12 Successive and prediction rate curve for the LR (a), RF (b), and WofE (c) methods

related variables in order to produce landslide susceptibility maps. Specifically, Shannon's entropy model was utilized for determining the appropriate classes that maximize the information coefficient for each variable. The developed methodology was implemented in the Nancheng County, China, using three quantitative methods, logistic regression, weight of evidence, and random forest, and was based on the analysis of eleven (11) conditional variables, namely, lithology, altitude, slope, aspect, topographic wetness index, sediment transport index, plan curvature, profile curvature, distance to river network, distance to tectonic features, and distance to road network.

According to the results of the research, each model had satisfactory performance, though the RF model had a slightly higher performance in terms of AUC predictive values (0.9220) against the ones estimated for the WofE (0.9090) and the LR model

(0.8940). The same pattern was observed when the success power of the models was calculated. Specifically, RF outperforms LR and WofE, having a higher performance in terms of AUC successive values (0.9350) in comparison with the ones calculated for WofE (0.9255) and LR (0.9097). From the visual inspection of the produced landslide susceptibility maps, the most susceptible areas are located at the west and east mountainous areas, while the central area is characterized by moderate to low susceptibility values.


Acknowledgments

The authors would like to express their gratitude to the Editor-in-Chief and the anonymous reviewers for their helpful comments on the manuscript. This research was supported by the National Natural Science Foundation of China (41472202, 41401032) and by General Program of Jiangxi Meteorological Bureau.

References

- Agterberg FP, Bonham-Carter GF, Wright DF (1990) In: Gaal GG, Merriam DF (eds) Statistical pattern integration for mineral exploration. Computer applications in resource estimation: prediction and assessment for metals and petroleum. Pergamon, Oxford, pp. 1–21
- 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 Mamdani fuzzy algorithm. *Comput Geosci* 38(1):23–34
- Akinci H, Dogan S, Kiligoclu C, Temiz MS (2011) Production of landslide susceptibility map of Samsun (Turkey) city Centre by using frequency ration model. *International Journal of Physical Science* 6(5):1015–1025
- Arefi H, Reinartz P (2011) Accuracy enhancement of ASTER global digital elevation models using ICESat data. *Remote Sens* 3(7):1323–1343
- Armas I (2012) Weights of evidence method for landslide susceptibility mapping; PrahovaSubcarpathians. Romania; *Natur Hazards* 60(3):937–950
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65(1–2):15–31
- Baldelli P, Aleotti P, Polloni G (1996) Landsliding susceptibility numerical map at the Messina Strait crossing site. In: K Senneset (ed), *Proc. 7th Int Symposium on Landslides*, Trondheim, Balkema, Rotterdam pp 1153–1158
- Barbieri G, Cambuli P (2009) The weight of evidence statistical method in landslide susceptibility mapping 424 of the Rio Pardu Valley (Sardinia, Italy). 18th World IMACS/MODSIM Congress, Cairns, Australia
- Bednarik M, Magulová B, Matys M, Marschalko M (2010) Landslide susceptibility assessment of the Kralůvany–Liptovský Mikuláš railway case study. *Phys Chem Earth* 35:162–171
- Bettian N, Birgit T (2007) Landslide susceptibility assessment using “weights-of-evidence” applied to a study area at the Jurassic escarpment (SW-Germany). *Geomorphology* 86:24
- Bonham-Carter GF (1994) *Geographic information Systems for Geoscientists: modeling with GIS*. Pergamon/Elsevier, London
- Bonham-Carter GF, Agterberg FP, Wright DF (1989) Weights of evidence modeling: a new approach to mapping mineral potential. In: Agterberg FP, Bonham-Carter GF (eds) *Statistical applications in the earth science. Geological survey of Canada Paper* (89–9):171–183
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Breiman L, Cutler A (2004) *Interface Workshop–April 2004*. Available online: <http://stat-www.berkeley.edu/users/breiman/RandomForests/interface04.pdf>. Accessed 25 Sept 2016
- Breiman L, Freidman JH, Olshen RA, Stone CJ (1984) *Classification and regression trees*. Wadsworth
- Can T, Nefeslioglu HA, Gokceoglu C, Sonmez H, Duman TY (2005) Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analysis. *Geomorphology* 72(1–4):250–271
- Carrara A, Cardinali M, Detti R, Guzzetti F, Pasqui V, Reichenbach P (1991) GIS techniques and statistical models in evaluating landslide hazard. *Earth Surf Process Landforms* 16:427–445
- Castellanos Abella EA, van Westen CJ (2007) Generation of a landslide risk index map for Cuba using spatial multi-criteria evaluation. *Landslides* 4:311–325
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Exploring model sensitivity issues across different scales in landslide susceptibility. *Natur Hazards Earth Syst Sci* 13:2815–2831
- Cervi F, Bert 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(4):433–444
- Chacón J, Irigaray C, Fernández T, El Hamdouni R (2006) Engineering geology maps: landslides and geographical information systems. *Bull EngGeol Environ* 65:341–411
- Chowdhury R (1976) Initial stresses in natural slope analysis, rock engineering for foundations and slopes. Geotechnical Engineering Division Specialty Conference, American Society of Civil Engineers (ASCE), pp. 404–415
- Chung CJ, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Natur Hazards* 30(3):451–472
- 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). *Environ Earth Sci* 63:397–406
- Costanzo D, Rotigliano E, Irigaray C, Jiménez-Perálvarez JD, Chacón J (2012) Factors selection in landslide susceptibility modelling on large scale following the GIS matrix method: application to the river Beiro Basin (Spain). *Natur Hazards Earth Syst Sci* 12:327–340
- 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* 54:311–324
- Dai FC, Lee CF (2003) A spatiotemporal probabilistic modeling of storm-induced shallow landsliding using aerial photographs and logistic regression. *Earth Surf Process Landforms* 28:527–545
- Dai FC, Lee CF, Ngai YY (2002) Landslide risk assessment and management: an overview. *Eng Geol* 64(1):65–87
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46
- Ercanoglu M, Gokceoglu C (2004) Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area west Black Sea region, Turkey. *Eng Geol* 75(3–4):229–250
- Esposito C, Barra A, Evans SG, Mugnozza GS, Delaney K (2014) Landslide susceptibility analysis by the comparison and integration of random forest and logistic regression methods; application to the disaster of Nova Friburgo-Rio de Janeiro, Brasil. *Geophys Res Abstr* 16:11407
- ESRI (2013) *ArcGIS desktop: release 10.1 Redlands, CA: Environmental Systems Research Institute*
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27:861–874
- Fell R, Corominas J, Bonnard C, Cascini L, Leroy E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol* 102:85–98
- Ghosh S, van Westen CJ, Carranza EJM, Ghoshal T, Sarkar N, Surendranath M (2009) A quantitative approach for improving the BIS (Indian) method of medium scale landslide susceptibility. *J Geol Soc India* 74:625–638
- Goetz HN, Brenning A, Petschko H, Leopold R (2015) Evaluating machine learning and statistical prediction techniques for landslide susceptibility modelling. *Comput Geosci* 81:1–11
- Gomez H, Kavzoglu T (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River basin, Venezuela. *EngGeol* 78(1–2):11–27
- Gorsevski PV, Paul EG, Jan Boll WJE, Randy BF (2006) Spatially and temporally distributed modelling for landslide susceptibility. *Geomorphology* 80:178–198
- Guns M, Vanacker V (2012) Logistic regression applied to natural hazards: rare event logistic regression with replications. *Natur Hazards Earth Syst Sci* 12:1937–1947
- Guzzetti FM, Cardinali PR, Carrara A (2000) Comparing landslides maps: a case study in the upper tiber basin, Central Italy. *Environ Manag* 25:247–263
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1):29–36
- Hansen L, Salamon P (1990) Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell* 12:993–1001
- Hong H, Naghibi SA, Pourghasemi HR, Pradhan B (2016a) GIS-based landslide spatial modelling in Ganzhou city. *China Arabian J Geosci*. doi:10.1007/s12517-015-2094-y
- Hong H, Pourghasemi HR, Pourtaghi ZS (2016b) Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* 259:105–118
- Ilija I, Tsangaratos P (2016) Applying weight of evidence method and sensitivity analysis to produce a landslide susceptibility map. *Landslides* 13(2):379–397
- Ilija I, Tsangaratos P, Koumantakis I, Rozos D (2010) Application of a Bayesian approach in GIS-based model for evaluating landslide susceptibility. Case study Kimi area, Euboea, Greece. *Bull Geol Soc Greece* 3:1590–1600
- Ilija I, Rozos D, Koumantakis I (2013) Landform classification using GIS techniques, the case study of Kimi municipality area, Euboea Island, Greece. *Bulletin of the Geological Society of Greece, Proceedings of the 13th International Congress, Chania XLVII*
- Irigaray C, Fernández T, El Hamdouni R, Chacón J (2007) Evaluation and validation of landslide-susceptibility maps obtained by a GIS matrix method: examples from the Betic cordillera (southern Spain). *Natur Hazards* 41:61–79
- Kayastha P, Dhital MR, De Smedt F (2012) Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. *Natur Hazards* 63:479–498

- Kouli M, Loupasakis C, Soupios P, Rozos D, Vallianatos F (2014) Landslide susceptibility mapping by comparing the WLC and WoFE multi-criteria methods in the West Crete Island, Greece. *Environ Earth Sci* 72(12):5197–5219
- Lan HX, Zhou CH, Wang LJ, Zhang HY, Li RH (2004) Landslide hazard spatial analysis and prediction using GIS in the Xiaojiang watershed, Yunnan, China. *Eng Geol* 76:109–128
- Lee S, Pradhan B (2006) Probabilistic landslide hazard and risk mapping on Penang Island, Malaysia. *J Earth Syst Sci* 115(6):661–672
- Lee HJ, Syvitski JPM, Parker G, Orange D, Locat J, Hutton JWH, Imran J (2002) Distinguishing sediment waves from slope failure deposits: field examples, including the “Humboldt slide” and modeling results. *Mar Geol* 192:79–104
- Lee S, Choi J, Min K (2004) Probabilistic landslide hazard mapping using GIS and remote sensing data at Boeun, Korea. *Int J Remote Sens* 25:2037–2052
- Lee CT, Huang CC, Lee JF, Pan KL, Lin ML, Dong JJ (2008) Statistical approach to earthquake-induced landslide susceptibility. *EngGeol* 100:43–58
- Lees BG (1996) Neural networks applications in the geosciences: an introduction. *Comput Geosci* 22:955–957
- Leroi E (1996) Landslide hazard-risk maps at different scales: objectives, tools and developments. In: Sennekes K (ed) *Landslides*. Proc. Int. Symp. on Landslides, Trondheim, pp. 35–52
- Magliulo P, Di Liso A, Russo F, Zelano A (2008) Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: a case study in southern Italy. *Natur Hazards* 47:411–435
- Marquardt D (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 12:605–607
- Mathew J, Jha VK, Rawat GS (2007) Weights of evidence modeling for landslide hazard zonation mapping of Bhagirathi Valley, Uttarakhand. *Curr Sci* 92(5):628–638
- Menard SW (2002) *Applied logistic regression analysis*, 2nd edn. Sage, Thousand Oaks CA, p. 111
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modeling: a review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5:3–30
- Murillo-García FG, Alcántara-Ayala I, Ardizzone F, Cardinali M, Fiouucci F, Guzzetti F (2015) Satellite stereoscopic pair images of very high resolution: a step forward for the development of landslide inventories. *Landslides* 12:277–291. doi:10.1007/s10346-014-0473-1
- Negnevitsky M (2002) *Artificial intelligence: a guide to intelligent systems*. Addison-Wesley/Pearson Education Harlow England, p. 394
- Neuhäuser B, Terhorst B (2007) Landslide susceptibility assessment using weights-of-evidence applied on a study site at the Jurassic escarpment of the Swabian Alb (SW Germany). *Geomorphology* 86:12–24
- Neuhäuser B, Damm B, Terhorst B (2012) GIS-based assessment of landslide susceptibility on the base of the weights-of-evidence model. *Landslides* 9(4):511–528
- O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Quality Quant* 41:673–690
- Park NW (2010) Application of Dempster–Shafer theory of evidence to GIS-based landslide susceptibility analysis. *Environ Earth Sci* 62(2):367–376
- Poli S, Sterlacchini S (2007) Landslide representation strategies in susceptibility studies using weights-of-evidence modeling technique. *Natur Resour Res* 16(2):121–134
- Pourghasemi HR, Kerle N (2015) Random forests and evidential belief function-based landslide susceptibility assessment in western Mazandaran Province. *Iran Environ Earth Sci*. doi:10.1007/s12665-015-4950-1
- Pourghasemi HR, Mohammady M, Pradhan B (2012) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* 97:71–84. doi:10.1016/j.catena.2012.05.005
- 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 (2009) Landslide risk analysis using artificial neural model focusing on different training sites. *Int J Phys Sci* 3(11):1–15
- 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* 25:747–759
- Pradhan B, Sezer E, Gokceoglu C, Buchroithner MF (2010) Landslide susceptibility mapping by neuro-fuzzy approach in a landslide prone area (Cameron highland, Malaysia). *IEEE Trans Geosci Remote Sens* 48(12):4164–4177
- Regmi NR, Giardino JR, Vitek JD (2010) Modeling susceptibility to landslides using the weight of evidence approach: western Colorado, USA. *Geomorphology* 115:172–187
- Safe Land-PP7 (2012) Deliverable D2.2. Examples of international practice in landslide hazard and risk mapping. Assessing the state of art of landslide hazard and risk assessment in the P.R. of China pp 223
- Sezer AE, Pradhan B, Gokceoglu C (2011) Manifestation of an adaptive neuro - fuzzy model on landslide susceptibility mapping: Klangvalley, Malaysia. *Expert Syst Appl* 38(7):8208–8219
- Shannon CE (1948) A mathematical theory of communication. *Bull Syst Technol J* 27:379–423
- Sharma M, Kumar R (2008) GIS-based landslide hazard zonation: a case study from the Parwanoo area, lesser and outer Himalaya, H.P., India. *Bull Eng Geol Environ* 67:129–137
- Soeters R, Van Westen CJ (1996) Slope instability recognition, analysis, and zonation. Special Report - National Research Council, Transportation Research Board 247:129–177
- SPSS Inc. (2007) *SPSS for Windows, Version 16.0*. Chicago
- Sujatha ER, Rajamanickam GV, Kumaravel P (2012) Landslide susceptibility analysis using probabilistic certainty factor approach: a case study on Tevankarai stream watershed, India. *J Earth Syst Sci* 121(5):1337–1350
- Tangestani MH (2009) A comparative study of Dempster-Shafer and fuzzy models for landslide susceptibility mapping using a GIS: an experience from Zagros Mountains, SW Iran. *Asian J Earth Sci* 35:66–73
- Thiery Y, Maquaire O, Fressard M (2014) Application of expert rules in indirect approaches for landslide susceptibility assessment. *Landslides* 11(3):411–424
- Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB (2012a) Spatial prediction of landslide hazards in Hoabinh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* 96:28–40
- Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB (2012b) Landslide susceptibility assessment in the Hoabinh province of Vietnam: a comparison of the Levenberg–Marquardt and Bayesian regularized neural networks. *Geomorphology* (171–172):12–19
- Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012c) Landslide susceptibility assessment in Vietnam using support vector machines, decision tree and Naïve Bayes models. *Mathematical Problems in Engineering*. pp 1–26
- Tien Bui D, Anh Tuan T, 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
- Tien Bui D, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) 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* 13:361–378
- Trigila A, Carla I, Carlo E, Gabriele SM (2015) Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology* 249:119–136
- Tsangaratos P (2012) Research on the engineering geological behaviour of the geological formations by the use of Information Systems. Phd Thesis, Athens, Greece (In Greek). p 363
- Tsangaratos P, Benardos A (2014) Estimating landslide susceptibility through an artificial neural network classifier. *Natur Hazards* 74(3):1489–1516
- Tsangaratos P, Ilia I (2016a) Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. *Catena* 145:164–179
- Tsangaratos P, Ilia I (2016b) Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi perfection, Greece. *Landslides* 13(2):305–320
- Tsangaratos P, Koumantakis I (2013) The value of geological data, information and knowledge in producing landslide susceptibility maps. *Bulletin of the Geological Society of Greece, Proceedings of the 13th International Congress Chania XLVII*
- Vahidnia MH, Alesheikh AA, Alimohammadi A, Hosseinali F (2010) A GIS-based neuro-fuzzy procedure for integrating knowledge and data in landslide susceptibility mapping. *Comput Geosci* vol 36:1101–1114
- Van Den Eeckhaut M, Vanwalleghem T, Poesen J, Govers G, Verstraeten G, Vandekerckhove L (2006) Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). *Geomorphology* 76:392–410
- Van Den Eeckhaut M, Marre A, Poesen J (2010) Comparison of two landslide susceptibility assessments in the champagne–Ardenne region (France). *Geomorphology* 115:141–155
- Van Westen CJ, Rengers N, Terlien M (1997) Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation. *Geol Rundsch* 86:4004–4414
- Van Westen CJ, Rengers N, Soeters R (2003) Use of geomorphological information in indirect landslide susceptibility assessment. *Natur Hazards* 30:399–419

- Verstappen HT (1983) Applied geomorphology: geomorphological surveys for environmental development. Elsevier xi +, Amsterdam, p. 437
- Wan S (2009) A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map. *EngGeol* 108(3–4):237–251
- Weisberg S, Fox J (2010) An R companion to applied regression. Sage Publications Incorporated, Los Angeles, London, New Delhi, Singapore, Washington DC
- Yang Z, Qiao J (2009) Entropy-based hazard degree assessment for typical landslides in the three Gorges Area, China. *Environ Sci Eng* 519–529
- Yang Z, Qiao J, Zhang X (2010) Regional Landslide Zonation based on entropy method in three Gorges Area, China. 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010), pp 1336–1339
- Yeon YK, Han JG, Ryu KH (2010) Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng Geol* 16(3–4):274–283
- Yesilnacar E, Topal T (2005) Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Eng Geol* 79:251–266
- Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ Earth Sci* 61(4):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(7):2161–2178
- Youssef AM, Pourghasemi HR, Pourtaghi Z, Al-Katheeri MM (2015) 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
-
- P. Tsangaratos · H. Hong**  · **C. Xu**
Key Laboratory of Active Tectonics and Volcano, Institute of Geology, China Earthquake Administration,
#1 Huayanli, Chaoyang District, PO Box 9803, Beijing, 100029, People's Republic of China
e-mail: hong_haoyuan@outlook.com
- P. Tsangaratos · I. Ilija**
School of Mining and Metallurgical Engineering, Department of Geological Studies, National Technical University of Athens,
Zografou Campus: HeronPolytechniou 9, 15780, Zografou, Greece
- H. Hong**
Jiangxi Provincial Meteorological Observatory,
Jiangxi Meteorological Bureau,
No.109 ShengfuBeier Road, Nanchang, 330046, China
- W. Chen**
College of Geology & Environment,
Xi'an University of Science and Technology,
Xi'an, 710054, China