

Spatial pattern and influencing factors of landslide casualty events

WANG Ying^{1,2}, LIN Qigen^{1,2}, SHI Peijun^{1,2}

1. Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing 100875, China;
2. Academy of Disaster Reduction and Emergency Management, Beijing Normal University, Beijing 100875, China

Abstract: Analysis of casualties due to landslides from 2000 to 2012 revealed that their spatial pattern was affected by terrain and other natural environmental factors, which resulted in a higher distribution of landslide casualty events in southern China than in northern China. Hotspots of landslide-generated casualties were in the western Sichuan mountainous area and Yunnan-Guizhou Plateau region, southeast hilly area, northern part of the loess hilly area, and Tianshan and Qilian Mountains. However, local distribution patterns indicated that landslide casualty events were also influenced by economic activity factors. To quantitatively analyse the influence of natural environment and human-economic activity factors, the Probability Model for Landslide Casualty Events in China (LCEC) was built based on logistic regression analysis. The results showed that relative relief, GDP growth rate, mean annual precipitation, fault zones, and population density were positively correlated with casualties caused by landslides. Notably, GDP growth rate ranked only second to relative relief as the primary factors in the probability of casualties due to landslides. The occurrence probability of a landslide casualty event increased 2.706 times with a GDP growth rate increase of 2.72%. In contrast, vegetation coverage was negatively correlated with casualties caused by landslides. The LCEC model was then applied to calculate the occurrence probability of landslide casualty events for each county in China. The results showed that there are 27 counties with high occurrence probability but zero casualty events. The 27 counties were divided into three categories: poverty-stricken counties, mineral-rich counties, and real-estate overexploited counties; these are key areas that should be emphasized in reducing landslide risk.

Keywords: landslide; casualty event; spatial pattern; influencing factors; counties; China

1 Introduction

Landslides are likely to cause loss of lives in China. During the period from 2000 to 2012, China had frequent landslides, 338,964 identifiable individual landslides resulting in 45,381

Received: 2017-06-14 **Accepted:** 2017-07-20

Foundation: National Key Research and Development Program Project, No.2017YFC1502505, No.2016YFA0602403; National Natural Science Foundation of China, No.41271544

Author: Wang Ying, Professor, specialized in disaster risk assessment and post-disaster recovery research.
E-mail: wy@bnu.edu.cn

casualties, approximately 3500 casualties per year (Sheng *et al.*, 2013). In terms of natural disasters, casualties caused by the landslides are only less than those caused by earthquakes and floods. There are various types of landslides in China. According to the Geological Disaster Prevention Regulations promulgated by the State Council (2004), landslides refer to geological hazards, including landslides, rock falls, debris flows, ground fissures, ground subsidence, and ground collapse, which are harmful to people's lives, property, and geological environment. The term "landslides" in this paper encompasses landslide, rock fall, and debris flow, which are the three major land mass wasting movements that cause casualties in China.

The Rockfall and Landslide Disaster Map of China (CIGEM, 2007a) and Debris Flow Disaster Map of China (CIGEM, 2007b) were compiled by the China Institute for Geo-Environment Monitoring based on the environmental geological investigations. They showed that most landslides have occurred in the hilly and mountainous areas of northeast, north and southeast China, Longmen mountain area of northwest, Sichuan, Qinling-Daba mountainous area, area along the coast of the Three Gorges Reservoir, southwest Guizhou mountainous area, Loess Plateau, northwest mountain, basin and plateau area, and Qinghai-Tibet Plateau. Since the 20th century, large-scale landslides in China have been disproportionally distributed by region: 70% on the first slope-descending zone of the continental landform along the eastern margin of the Qinghai-Tibet Plateau; 16% in central Hunan province and middle mountainous area of Guangxi province; and 12% in the mountainous areas of Zhejiang and Fujian provinces (Huang, 2007). However, these maps mainly contain contents on the magnitudes of the landslides, and lack losses caused by the landslides. Such losses, especially casualties, are more important than landslide frequency for disaster risk prevention.

Landslide occurrence is affected by natural environment factors, such as topography, lithology, geology, faults, soil types, land cover, and rainfall (Atkinson and Massari, 1998; Ohlmacher and Davis, 2003; Guzzetti *et al.*, 2007; García-Rodríguez *et al.*, 2008; Eeckhaut *et al.*, 2011; Ramani *et al.*, 2011; Althuwaynee *et al.*, 2016; Tsangaratos and Ilia, 2016; Samia *et al.*, 2017). Moreover, there is a clear close relationship between landslide occurrence and the impact of human activities on the natural environment. The primary reason for large-scale catastrophic landslides in China since the 1980s was the increase in anthropogenic activities. Statistical analyses have shown that 50% of large-scale landslides that occurred on the Chinese Mainland are directly or indirectly related to human activities (Huang, 2007). Many anthropogenic activities, such as road construction (Brenning *et al.*, 2015) and land-use change (Cerovski-Darriau and Roering, 2016; Tarolli and Sofia, 2016; Bartelletti *et al.*, 2017), influence the occurrence of landslides (Gill and Malamud, 2017).

Brenning *et al.* (2015) investigated landslide initiation frequency along highways in the tropical Andes of southern Ecuador. They found that landslide susceptibility was more than an order of magnitude higher in close proximity to paved interurban highways. Ayalew and Yamagishi (2005) assessed landslide susceptibility around Kakuda-Yahiko Mountain in central Japan using logistic regression analysis and considering factors, such as lithology, bed rock-slope relationship, slope gradient, aspect, elevation, and road network. Their findings showed that the road network was the chief determinant of landslide occurrence and distribution. Jadda *et al.* (2011) using a frequency ratio model and GIS technology, com-

bined with geology, geomorphology, soil, topography, land use, distance from roads, drainages, and faults to map landslide susceptibility at Alborz Mountain in Iran. They found that roads were the second most important factor after geology, and that landslide susceptibility increased as the distance from roads decreased. Taking various factors into account, such as slope gradient, lithology, land use, soil humidity, Eeckhaut *et al.* (2011) analysed landslide susceptibility in Europe based on six land-use types: cropland, forest, grassland, bare land, urban area, and regularly flooded land. In general, forests are more prone to landslides than cropland, and cropland has a higher hazard than urban areas. Bartelletti *et al.* (2017) analysed the influence of land use on landslide occurrence from a geological-morphological perspective. Their results indicated that terraced agricultural areas are more prone to landslides than woodland. Such studies have primarily focused on the impact of roads and land use types on landslide occurrence. Human activities, such as mining and real estate development, may also trigger landslides. However, there have been few studies that have analysed the general relationship between human activity and landslide occurrence. This paper is intended to bridge this gap, exploring the influence of human activities on landslides that have caused casualties in China.

Based on data for landslide-generated casualty events in China from 2000 to 2012, this study evaluates the spatial pattern of the casualty events on a national scale. It should be noted that a casualty event in this paper refers to a single landslide causing one casualty or multiple casualties. Then a logistic regression model is applied to quantitatively evaluate the relative contribution of natural environment and human activity factors to landslide casualty events. This type of analysis is very important for understanding the relationships between landslides and casualties in China, identify hotspots, and establish corresponding disaster prevention and mitigation measures.

2 Material and methods

2.1 Study area and data

China lies between 18°–54°N and 73°–135°E, covering 9.6 million km². The geological and geographic environment is complex and results in spatial and temporal differences in climate. The terrain of China is low in elevation in the east but high in the west, and mountainous areas account for nearly two-thirds of the land area. Correspondingly, nearly half of the towns are in the mountainous areas, with a population of accounting for 50% of the country's total. Landslides occur frequently in the mountainous areas resulting in serious loss of lives every year in China (Li *et al.*, 2004; Cees van Westen *et al.*, 2010; Li *et al.*, 2013).

Statistics and reports on landslide disasters in China have gradually been standardized and perfected since 2011. Real-time reports of landslide events are published by the Ministry of Land and Resources of China in the form of *Reports on Geological Disaster Situation*. Annual losses from landslide disasters are published in the form of the *China Geological Hazard Bulletin* by the China Institute for Geo-Environment Monitoring. However, data reporting landslide events prior to 2011 was not uniform. Landslide events that caused casualties were documented from 2000 to 2012 by the National Disaster Reduction Center of China (NDRCC, 2013), Ministry of Land and Resources of China (MLRC, 2013), China Institute

for Geo-Environment Monitoring (CIGEM, 2013), and news websites. The Landslides Casualty Inventory of China was established, characterizing the type, time, location, casualties, and damage caused by each landslide disaster, incorporating 576 landslide disaster events in total. Events duplicated from multiple sources were removed after considering the time and location of the disaster. Table 1 presents the data sources for the Landslides Casualty Inventory of China. There are some cases of flash flood and debris flow concurrence, such as disastrous flash floods and debris flows in Zhouqu in 2010 and Shannan Town, Ning'an City, Heilongjiang Province in June 2005. Such concurrent disasters often cause enormous casualties and property losses, but their formation mechanism is quite different from a single debris flow event, therefore in this study, landslide inventory does not include this type of disasters.

Table 1 Data sources for the Landslides Casualty Inventory of China

Source	Data compiler	Cases	Channel
China Geological Hazard Bulletin	China Institute for Geo-Environment Monitoring	81	CIGEM
Reports on Geological Disaster Situation	Ministry of Land and Resources of China	95	MLRC
Yesterday Disaster Report	NDRCC	228	NDRCC
Web news search	News websites	172	Network collection and arrangement

Based on identifications of counties where landslide disasters occurred, a distribution map of landslide casualty events in China for the period 2000–2012 was generated. As shown in Figure 1, landslide casualty events occurred in a total of 334 counties. These counties are mainly distributed in the Sichuan Basin and Yunnan-Guizhou Plateau (39%) and southeast hilly area (34%) in southern China. About 12% of the counties are distributed in the loess hilly area and about 10% in Tianshan and Qilian mountains. There are a few counties distributed in the North and Northeast China Plains and Tibetan Plateau.

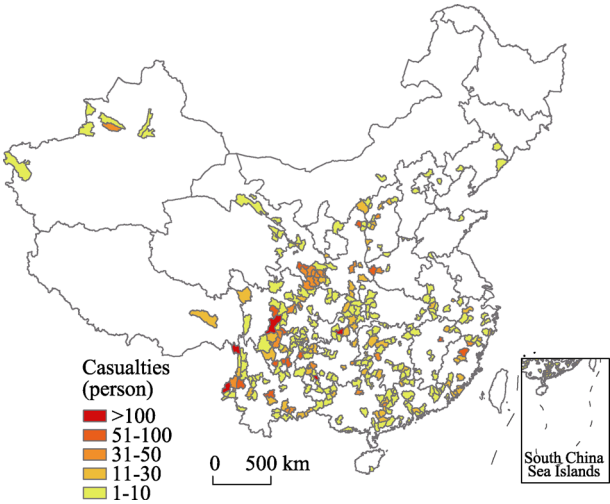


Figure 1 Distribution of counties in China with landslide casualties (2000–2012)

Based on analysing the characteristics of China’s natural environment, six factors including topography, faults, precipitation, lithology, vegetation, and soil types were selected as natural environment influencing factors. These factors are illustrated in Figure 2: relative topographic relief (Figure 2a), Quaternary active faults (Figure 2b), mean annual precipitation (Figure 2c), vegetation coverage (Figure 2d), lithology (Figure 2e), and soil type (Figure 2f). These maps for each county were generated using the ArcGIS spatial analysis mod-

ules. The datasets were obtained from the Computer Network Information Center of CAS (2015), Institute of Geographic Sciences and Natural Resources Research of CAS (2015), and Hartmann and Moosdorf (2012). Comparing Figures 1 and 2, it is clear that these natural environmental factors have a spatial correlation with landslide casualty events in China. The counties with high relative relief, fault zones, more precipitation, and high vegetation coverage are more prone to landslide events that have caused casualties.

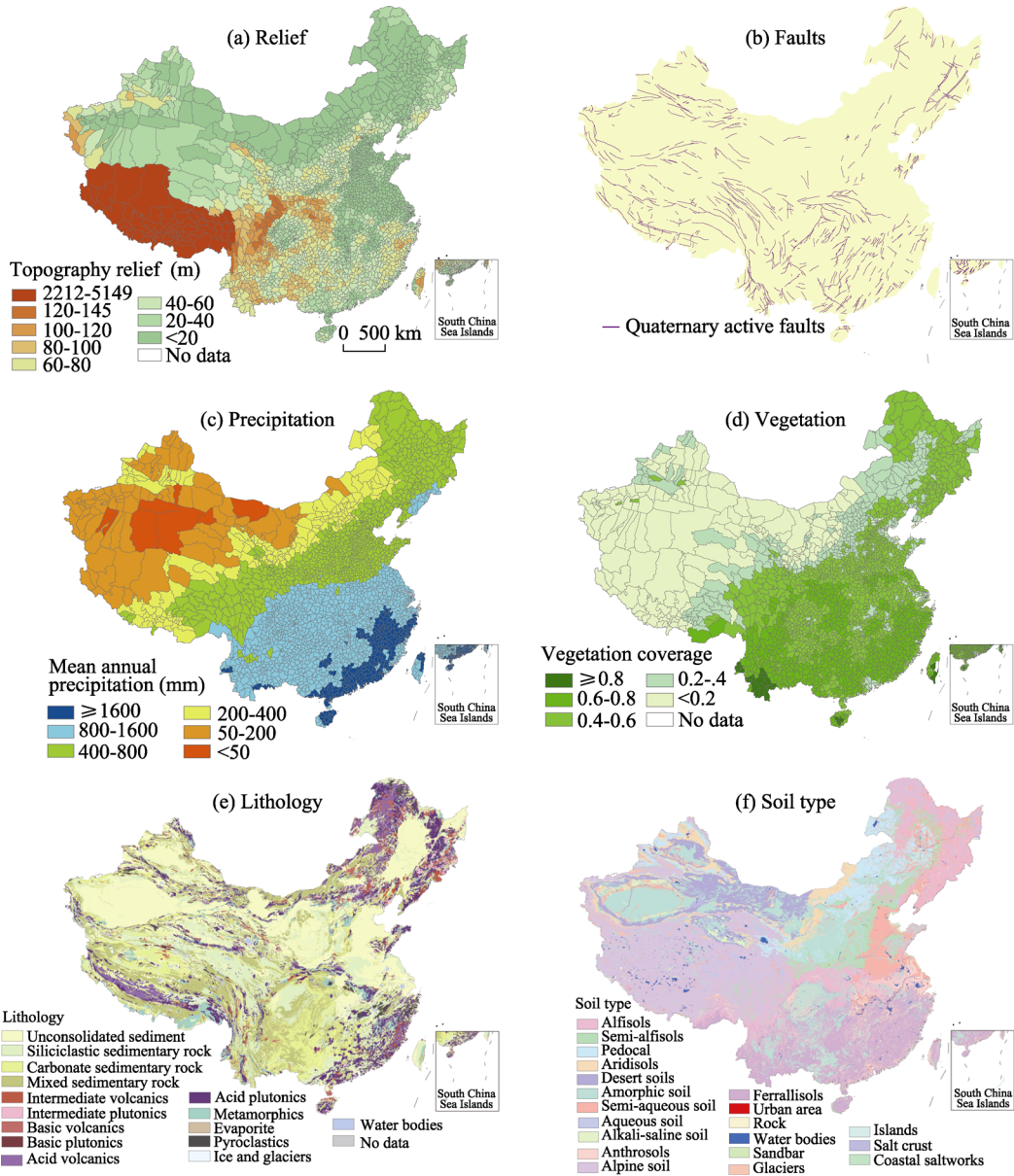


Figure 2 Distribution of natural environment factors in landslide casualty events across China

There are some counties with these described natural environmental characteristics but no casualty events have occurred. For instance, some counties in Yunnan and Fujian provinces have similar environmental characteristics with their surrounding casualty counties and yet

no landslide casualties have occurred since 2000. The difference appears to be driven by human activity factors, such as population density (Figure 3a) and human-economic activity intensity, which have provided the physical characteristics leading to casualties.

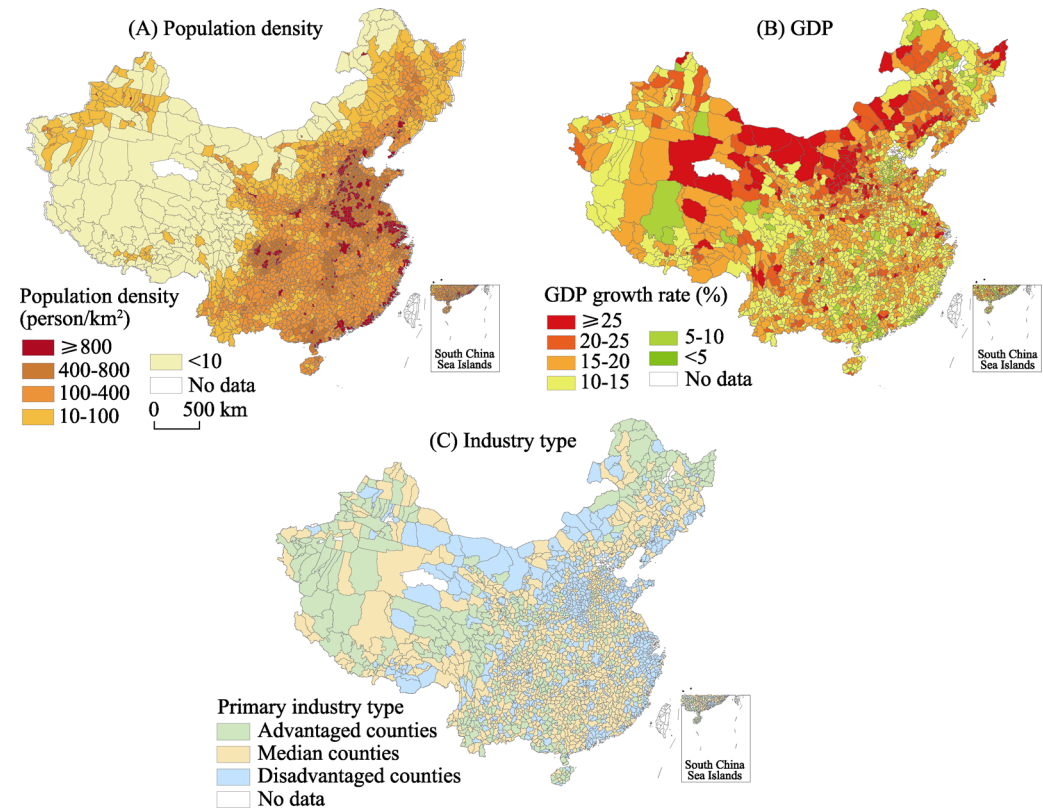


Figure 3 Distribution of human-economic activity factors in landslide casualty events across China

Economy is a most comprehensive indicator reflecting human activity intensity. Since 2000, China's economy has maintained rapid growth. From 2000 to 2012, China's GDP had an average annual growth rate of 10.14% (in comparable price); however this rapid economic growth has been achieved through over-exploiting resources at the expense of the natural environment. The World Bank's empirical data show that a society's golden development period is also a period of high incidence of conflicts and high-risk emergence (Zhang and Tong, 2009). Therefore, GDP growth rate (Figure 3b) and industry type (Figure 3c) were selected to mirror the intensity of human activity. In this study, GDP growth rate in each county was calculated as the average annual GDP growth rate from 2000 to 2012 at current prices. Industry type was calculated as the proportion of the primary industry in GDP in 2010. The natural break classification method used in Hu *et al.* (2011) and Luo *et al.* (2016) was applied to divide the counties into three categories: the primary industry advantaged counties, median counties, and disadvantaged counties.

Comparing Figures 3 and 1, there is not a clear spatial correlation between landslide casualty events and human activity. In the population density map, counties with high population density are mainly concentrated in eastern China, while those with low population

density are mainly in western China. In comparison, counties with high GDP growth rate are scattered across various provinces, and industrial types also showed a mixed county distribution. Therefore, to accurately evaluate the correlation between potential factors and landslide casualties, a logistic regression model was used to analyse the data.

2.2 Modelling

2.2.1 Logistic regression model

Logistic regression analysis is a type of probabilistic nonlinear regression model. The advantage of logistic regression for statistical analyses is that independent variables can be either continuous or discrete (with dummy variables). In recent years, logistic regression models have been widely applied to landslide hazard and susceptibility assessments of factors influencing the occurrence of geological disasters (Chau and Chan, 2005; Lee and Pradhan, 2006; García-Rodríguez *et al.*, 2008; Falaschi *et al.*, 2009; Bui *et al.*, 2011; Eeckhaut *et al.*, 2011; Shirzadi *et al.*, 2012; Devkota *et al.*, 2013; Pourghasemi *et al.*, 2013; Xu *et al.*, 2013; Nourani *et al.*, 2014). The binary logistic regression equation is often used to study the relationship between classified observation results and their influencing factors, which is provided as follows:

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_n)} \quad (1)$$

where P is the dependent variable in the interval $[0, 1]$ that reflects the occurrence probability of a casualty event caused by landslides, x_i is an independent variable related to influencing factors ($i = 1, 2, \dots, k$); β_0 is a constant; and β_i is the regression coefficient for the influencing factors x_i , reflecting the influence of independent variable x_i on P .

In the logistic regression model, independent variables can be continuous or categorical. Continuous variables can be added directly to Equation (1). After regression analysis, we obtained the partial regression coefficient β_i , which refers to the degree of influence on P in the case of continuous change in the independent variable. Categorical variables were first classified by values, assigned to corresponding numerical values, and then added to Equation (1). The coefficient β_i enabled us to evaluate the strength of the relationship between a variable and landslides occurrence, compared with a reference category. Stronger relationships identified the categorical variables having an impact on the dependent variables.

The dependent variable in this study was the presence (1) or absence (0) of a casualty event caused by landslides in a single county. Of the 2215 Chinese counties, those that appeared in the Landslides Casualty Inventory of China were classified to 1, otherwise they were classified as 0. P was set to 1 for counties with casualty event, and P was 0 for counties with no casualty event. Due to the lack of data in the municipal districts of Beijing, Tianjin, and Shanghai, and the special terrain of Tibet, seven counties with casualty events in these areas were not included. The 327 counties with landslide casualty events were applied to establishing the model. Bringing human-economic activity and natural environment data from these counties into Equation (1) resulted in quantitative relationships between influencing factors and landslide casualty events in China according to coefficient β_i .

In total, 327 “1” counties and 1888 “0” counties were identified. Using information from all “0” counties in the logistic regression model analysis would have resulted in a bias to-

wards “0” counties. Currently, no conclusions have been drawn about the number of 1 and 0 samples that should be used in the logistic regression. According to King and Zeng (2001), too few 1 samples “grossly undervalued” the probability of prediction, and they suggested keeping the rate of 0 samples to 1 samples between 0.2 and 1. Therefore, to better predict “1” casualty counties, and considering the information of “0” casualty counties, weighting of the data was performed using all the “0” non-casualty counties as samples and four times the number of “1” casualty counties as samples in the logistic model analysis. Therefore, the number of “1” casualty county samples was 327×4 (i.e., 1308) and the number of “0” non-casualty county samples was 1888. The ratio of 1 to 0 is 0.69. This is called the over-sampling method, which is currently widely used in machine learning to achieve uniformity of training sample data (Yen *et al.*, 2006).

For the independent variables, topographic relief, mean annual precipitation, vegetation coverage, GDP growth rate, and population density were continuous variables. As there is a big difference in the numerical values of these factors, the first step in the analysis was to conduct natural logarithm standardization processing. The processed data were then used as input in the logistic regression modelling.

Faults, lithology, soil type, and industry type were categorical variables in this study. Fault variables were assigned values based on the occurrence of a Quaternary active fault in the county; mapped occurrence was assigned a value of 1, otherwise they were 0. Lithology variables were classified into 15 categories, including unconsolidated sediment, siliciclastic sedimentary rock, pyroclastics, mixed sedimentary rock, carbonate sedimentary rock, evaporite, acid volcanic, intermediate volcanics, basic volcanics, acid plutonics, intermediate plutonics, basic plutonics, metamorphics, water bodies, ice, and glaciers based on the classification of the world lithology map (Hartmann *et al.*, 2012). Lithology variables were converted into 14 binary dummy variables with unconsolidated sediment as the reference category. Lithology was assigned based on the largest coverage. For instance, for carbonate sedimentary rock accounting for the largest area in one county compared to other lithology types, the county was assigned a value of 1 in the carbonate sedimentary rock variable and 0 in other lithology variables. Soil type variables were classified to 12 categories, including alfisols, semi-alfisols, pedocal, aridisols, desert soils, amorphous soil, semi-aqueous soil, aqueous soil, alkali-saline soil, anthrosols, alpine soil, and ferrallisols, based on the "Soil Classification" system (GB). The soil type variables were converted into 11 binary dummy variables with alfisols as the reference category. For industry type, the natural break method was applied to divide the industry type variables into three categories based on the proportion of the primary industry in GDP in 2010. Primary industry accounting for the proportion of GDP $\geq 31.21\%$ resulted in categorization as a primary industry advantaged county. The proportion of primary industry to GDP of 14.70%–31.21% resulted in categorization as a primary industry median county. The remaining counties were categorized as primary industry disadvantaged. The industry type variables are converted into two binary dummy variables with primary industry advantaged county as the reference category.

2.2.2 Evaluation methods

The performances of the model were firstly evaluated based on the confusion matrix and receiver operating characteristic (ROC) curve. The confusion matrix can be used to deter-

mine the percentages of correctly classified observations and number of false positive (FP), false negative (FN), true positive (TP), and true negative (TN) observations. A high percentage of correctly classified observations and low number of false positive and false negative observations indicate a better model fitting. The ROC curve measures the goodness of fit of the model prediction by plotting the true positive rates and false positive rates at different susceptibility threshold values. The area under the ROC curve (AUC) ranges from 0.5 to 1, with higher values indicating a better predictive capability of the model. AUC values less than 0.7 indicate poor predictive ability, values between 0.7 and 0.8 indicate moderate ability, values between 0.8 and 0.9 indicate good ability, and values > 0.9 indicate excellent predictive ability (Swets, 1988).

To further evaluate the robustness of the model, a simple random sampling was conducted ten times to divide the counties into 70% training, i.e., 229 casualty counties, 1322 non-casualty counties, and 30% validation groups (Chung and Fabbri, 2003; Poiraud, 2014). Then, a 70% training group was used to create ten models and the remaining 30% was used to validate the ten models. Category variables, lithology, soil type and industry type, were not included in the process of random sampling cross validation due to the excessive number of classes in the category variables.

3 Results

Table 2 shows that nine variables, including topographic relief, mean annual precipitation, vegetation coverage, GDP growth rate, population density, faults, lithology, soil type, and industry type were selected using the likelihood ratio variable selection method based on the p -value < 0.05 , which is indicated with the Sig. column in Table 2. The results indicated that these variables make a difference to casualty event caused by landslides.

The degree of influence of each variable depends on the size of $\text{Exp}(\beta)$. $\text{Exp}(\beta)$ is an odds ratio (OR for short), indicating the ratio of casualty event occurrence probability due to every variable change. $\text{Exp}(\beta) > 1$ indicates a positive influence; $\text{Exp}(\beta) < 1$ shows a negative influence. Greater differences between $\text{Exp}(\beta)$ and 1, indicate a greater degree of influence of the variable. $\text{Exp}(\beta)$ of the influencing variables in Table 2 indicate that topographic relief, mean annual precipitation, GDP growth rate, population density, and faults had a positive effect on the occurrence of landslide casualty events. While vegetation coverage had a negative influence on the occurrence of landslide casualty events. The negative effect of vegetation coverage indicates that although counties with high numbers of landslide casualty events were mainly distributed in the southern region with high vegetation coverage on the national scale, an increase in vegetation coverage can reduce occurrences of landslide casualty events when the other natural environment factors are similar. Regarding soil type, aridisols, desert soils, amorphic soils, and anthrosols increased the possibility of the landslide casualty events in comparison with alfisols. Regarding lithology, carbonate sedimentary rock and acid volcanics decreased the possibility landslide casualty events compared with unconsolidated sediment. Regarding industry type, primary industry median county and primary industry disadvantaged county increased the possibility landslide casualty events compared with the reference, primary industry advantaged county.

Table 2 Variables applied in the logistic regression model

Variables	β	S.E.	Sig.	Exp(β)
ln (Topography relief)	1.922	0.117	0.000	6.834
ln(GDP growth rate)	0.996	0.209	0.000	2.706
ln(Mean annual precipitation)	0.535	0.178	0.003	1.707
ln(Vegetation coverage)	-0.333	0.159	0.037	0.717
Faults	0.374	0.093	0.000	1.453
ln(Population density)	0.317	0.063	0.000	1.373
Lithology ^a			0.000	
siliciclastic sedimentary rock	-0.649	0.171	0.000	0.523
pyroclastics	-0.448	0.330	0.174	0.639
mixed sedimentary rock	-0.173	0.168	0.302	0.841
carbonate sedimentary rock	-0.572	0.178	0.001	0.564
acid volcanic	-1.574	0.408	0.000	0.207
intermediate volcanics	-0.866	0.475	0.069	0.421
basic volcanics	-19.936	9093	0.998	0.000
acid plutonics	-0.204	0.184	0.268	0.815
intermediate plutonics	-20.172	12480	0.999	0.000
basic plutonics	-19.612	40190	1	0.000
metamorphics	0.238	0.309	0.442	1.268
water bodies	-17.337	27290	0.999	0.000
Soil type ^a			0.000	
semi-alfisols	-0.396	0.241	0.1	0.673
pedocal	0.567	0.31	0.068	1.762
aridisols	1.183	0.477	0.013	3.263
desert soils	2.083	0.532	0.000	8.026
amorphic soil	0.447	0.169	0.008	1.564
semi-aqueous soil	0.857	0.468	0.067	2.357
alkali-saline soil	-17.918	12030	0.999	0
anthrosols	0.893	0.284	0.002	2.443
alpine soil	0.467	0.3	0.119	1.595
ferrallisols	0.524	0.157	0.001	1.69
Industry type ^a			0.000	
primary industry median county	0.536	0.126	0.000	1.71
primary industry disadvantaged county	0.352	0.145	0.015	1.422
Constant	-11.913	1.344	0.000	0.000

a. The reference categories of lithology, soil type and industry type are unconsolidated sediment, alfisols and primary industry advantaged county respectively

Substituting the parameters in Table 2 into Equation (1) resulted in Equations (2) and (3), as follows:

$$\begin{aligned}
 Z = & -11.913 + 1.922X_1 + 0.996X_2 + 0.535X_3 - 0.333X_4 + 0.374X_5 + 0.317X_6 - \\
 & 0.649X_{71} - 0.448X_{72} - 0.173X_{73} - 0.572X_{74} - 1.574X_{75} - 0.866X_{76} - 19.936X_{77} - \\
 & 0.204X_{78} - 20.172X_{79} - 19.612X_{710} + 0.238X_{711} - 17.337X_{712} - 0.396X_{81} + \\
 & 0.567X_{82} + 1.183X_{83} + 2.083X_{84} + 0.447X_{85} + 0.857X_{86} - 17.918X_{87} + 0.893X_{88} + \\
 & 0.467X_{89} + 0.524X_{810} + 0.536X_{91} + 0.352X_{92}
 \end{aligned} \quad (2)$$

$$P = \frac{\exp(Z)}{1 + \exp(Z)}$$

(3)

where P is the occurrence probability of casualty event caused by landslides; X_1 is $\ln(\text{topographic relief})$; X_2 is $\ln(\text{GDP growth rate})$; X_3 is $\ln(\text{mean annual precipitation})$; X_4 is $\ln(\text{vegetation coverage})$; X_5 is fault; X_6 is $\ln(\text{population density})$; X_{71} is siliciclastic sedimentary rock; X_{72} is pyroclastics; X_{73} is mixed sedimentary rock; X_{74} is carbonate sedimentary rock; X_{75} is acid volcanic; X_{76} is intermediate volcanics; X_{77} is basic volcanics; X_{78} is acid plutonics; X_{79} is intermediate plutonics; X_{710} is basic plutonics; X_{711} is metamorphics; X_{712} is water bodies; X_{81} is semi-alfisols; X_{82} is pedocal; X_{83} is aridisols; X_{84} is desert soils; X_{85} is amorphic soil; X_{86} is semi-aqueous soil; X_{87} is alkali-saline soil; X_{88} is anthrosols; X_{89} is alpine soil; X_{810} is ferrallisols; X_{91} is primary industry median county; and X_{92} is primary industry disadvantaged county. Equations (2) and (3) form the Probability Model for Landslide Casualty Events in China (LCEC model).

Table 3 shows the correct percentages in confusion matrix for the LCEC model. The accuracy in predicting the occurrence of casualty event was 74.9%, and the accuracy in predicting no occurrence of casualty event was 75.0%, giving a total accuracy of 75.0%. Figure 4 shows the ROC curve for the LCEC model. The area under the ROC curve (AUC) is 0.826 with a standard error of 0.007. Therefore, the LCEC model was of relatively high predictive value.

Table 3 Confusion matrix for the LCEC model

Observed	Predicted		Percentage correct (%)
	Casualty		
	0	1	
0	1416	472	75.0
1	328	980	74.9
Overall percentage			75.0

The AUC values from ten training models ranged from 0.811 to 0.831. The validation results of these ten models ranged from 0.781 to 0.821 (Figure 4). These results indicate that the LCEC model performed accurately.

To summarize, topographic relief, mean annual precipitation, vegetation coverage, GDP growth rate, population density, faults, lithology, soil type, and industry type all contributed to the occurrence of landslide casualty events in China. In Table 2, the odds ratio (OR) of GDP growth rate is 2.706. This indicates that for every 2.72% increase in the GDP growth rate, the occurrence probability of a landslide casualty

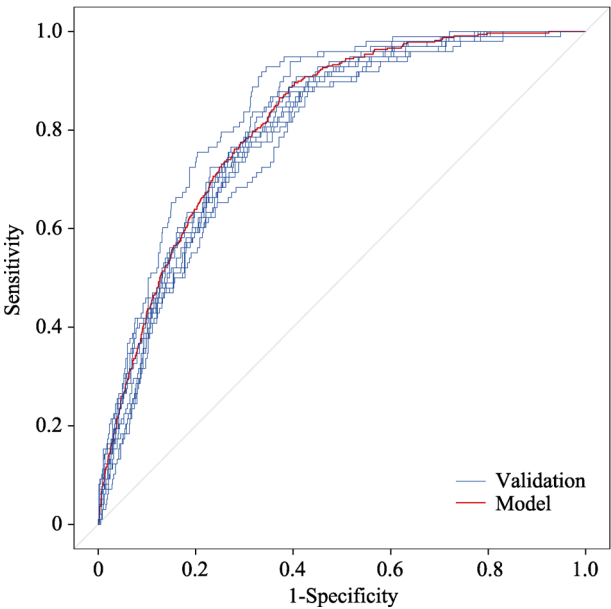


Figure 4 ROC curve for LCEC model and validation of models produced from 10 samples of 70% training data

event increased by 2.706 times. Thus, as shown, rapid growth in GDP has an important impact on casualties caused by landslides. This has been confirmed on a county-by-county basis, where rapid growth of GDP has occurred at the cost of environmental damage, which in turn has resulted in numerous casualty events.

4 Discussion

After placing the values of independent variables from the 2215 counties in China into Equations (2) and (3), the occurrence probability of landslide casualty events was determined for each county (Figure 5). At present, no conclusions exist to classify landslide susceptibility and hazard maps in the literature. In this study, we divide the occurrence probability into five categories: very low, low, medium, high, and very high based on the classification method in Hu *et al.* (2011) and Luo *et al.* (2016). Figure 5 shows that the very high and high occurrence probability areas are similar to Figure 1, mainly distributed in the Sichuan Basin and Yunnan-Guizhou Plateau, southeast hilly area in southern China, loess hilly area, and Tianshan and Qilian mountains.

However, there were 80 counties with occurrence probability > 0.78 . Among them, 42 counties have suffered casualty events. County GDP growth exceeded average provincial GDP growth in 27 out of the remaining 38 counties where no casualty events occurred from 2000 to 2012. Based on probability analysis, these 27 counties are locations where it is necessary for China to spare efforts in their landslides risk prevention in the future. The natural environmental conditions of these counties, such as topographic relief, mean annual precipitation, and vegetation coverage, are prone to landslides. Moreover, these counties have witnessed a fast-growing economy for nearly a decade. According to the distinctive characteristics of economic development in each county, they were divided into three categories: poverty-stricken, mineral-rich, and real-estate overexploited counties (Table 4).

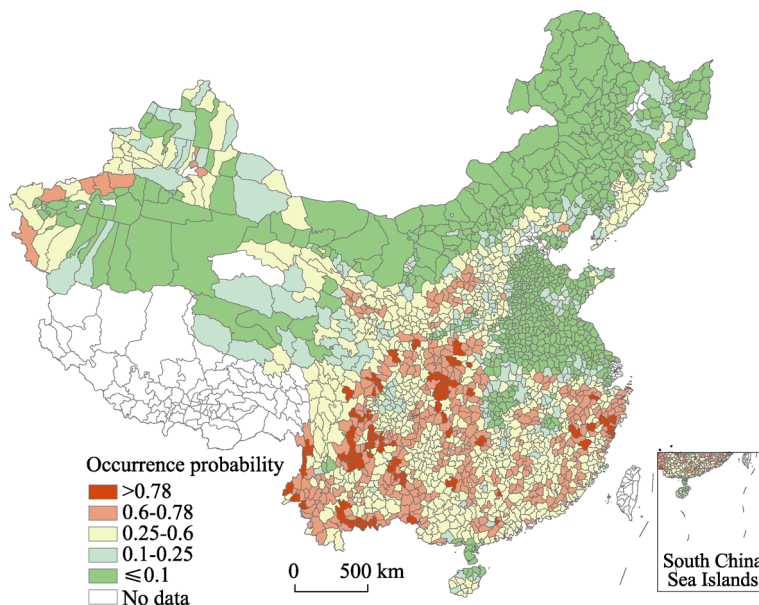


Figure 5 Distribution of landslide casualty event probabilities across China

Table 4 Counties, stressing geological disaster risk prevention

Category	County
Poverty-stricken counties	Jinyang County, Yongshan County, Qianjiang District, Luding County, Mao County, Mianning County, Zhenkang County, Shizhu County, Xichang City, Huili County, Ludian County, Mojiang County, Pingbian County, Hekou County, Yuanyang County
Mineral-rich counties	Feng County, Jingxi County, Miyi County, Emeishan city, Hongya County, Hongya County, Huidong County, Shui Fu County, Shangluo municipal district
Real-estate overexploited counties	Lishui municipal district, Zhoushan municipal district, Qingtian County, Wuyishan City

Poverty-stricken counties have been the focus of state poverty relief work. Despite the support of state aid-the-poor programmes, and the relatively rapid development of the region's economy in recent years, they are still quite economically disadvantaged and disaster-prevention work has been minimal. Therefore, measures must be taken to intensify landslide disaster prevention work, which in the process may facilitate a decrease in poverty.

Mineral-rich counties have greatly contributed to the high-speed growth of GDP because of their mineral wealth. For example, Fengxian County in Shaanxi Province is one of the four major lead-zinc ore bases in the country, with 3.5 million tons of ore reserves accounting for 80% of the whole province. For nearly a decade, Fengxian has prioritized its mineral development to such an extent that its average GDP growth is twice that of the province. However, mineral development has caused serious damage to the region's vegetation and ecology, which increases the probability of future landslide casualty events. For this reason, development of mineral resources should be combined with appropriate measures to protect the ecological environment and conduct landslide disaster prevention work.

Real-estate overexploited counties are municipal districts primarily in eastern China. In view of their relatively high economic significance, decreasing dependence on raw materials of industrial development, extensive processing development, and involvement of an increasingly large proportion of tertiary industries, these counties have sufficient economic strength to undertake precautions against natural disasters. However, vigorous development of the real estate industry and tourism has expanded development to the mountainous region. For instance, Zhejiang is a province with a vast mountainous area that occupies 70.4% of the land area. Lishui city in southwestern Zhejiang Province has about 88.42% of the city's total mountainous area. In recent years, real estate development has been rapid. The area under construction has risen from $5.4055 \times 10^6 \text{ m}^2$ in 2003 to $19.1346 \times 10^6 \text{ m}^2$ in 2012. Therefore, many newly constructed buildings have been erected in mountainous regions, and more residents will live in landslide disaster prone places. Therefore, landslide disaster prevention work should be strengthened in these eastern counties, in spite of their rapid economic development.

5 Conclusion and perspective

Based on county-level casualty data in the Landslides Casualty Inventory of China and a logistic regression model, this study analysed the spatial pattern of landslide casualty events and quantitatively evaluated the effect of natural environment and human activity factors on landslide casualty events. The spatial distribution of casualty events was higher in south

China than in north China. Hotspots of landslide casualty events included the western Sichuan mountainous area and Yunnan-Guizhou Plateau region, southeast hilly area, northern part of the loess hilly area, and Tianshan and Qilian mountains.

The LCEC model was established and results showed that both the natural environment and human activity were factors affecting the probability of landslide casualty events; sub-factors included topographic relief, mean annual precipitation, vegetation coverage, GDP growth rate, population density, faults, lithology, soil type, and industry type. In descending degree, topographic relief, GDP growth rate, mean annual precipitation, vegetation coverage, fault zones, and population density were most important in contributing to landslide casualty events. For every 2.72% increase in the GDP growth rate, the occurrence probability of casualty event caused by landslides increased 2.706 times. This result quantitatively proved that the fast-growing economy damaged the natural environment and triggered more landslide disasters.

An analysis of the LCEC model shows that about 71% of the counties with high occurrence probability but no casualty events have been rapidly developing economically and the GDP growth of these counties has been far greater than their host provinces. They can be divided into three categories: poverty-stricken, mineral-rich, and real-estate overexploited counties, which relate their natural and economic factors. These countries are key areas where great importance must be placed on landslide disaster risk prevention.

In Table 2, the hypothesis of a constant significance level <0.05 , i.e., constant = 0, is false, which indicates that in addition to the factors considered in this study, there are likely other factors that contribute to the occurrence of landslide casualty events. Therefore, additional and more detailed data are required to use as indicators of natural environment and human activity factors. Such data will further improve the accuracy of the LCEC model. The study did consider the GDP, per capita GDP, and per area GDP as human activity factors, but the results showed that these factors had no significant effect on landslide casualty events. The lack of correlation may have been due to the long time period that the size of a regional GDP develops, which makes it less sensitive to local developments that impact county-scale GDP. Therefore, these factors cannot effectively reflect the intensity of local human activities.

Due to lack of detailed information on risk elements in the Landslides Casualty Inventory of China, this study only adopted a binary logistic regression model to analyse the occurrence of landslide casualty events. With improvements to the database, methods, such as Geodetector (Wang and Li, 2010), can also be applied to detect the spatial association between influencing factors and landslide casualty events.

References

- Althuwaynee O F, Pradhan B, Lee S, 2016. A novel integrated model for assessing landslide susceptibility mapping using CHAID and AHP pair-wise comparison. *International Journal of Remote Sensing*, 37(5): 1190–1209.
- Atkinson P, Massari R, 1998. Generalised linear modelling of susceptibility to landsliding in the Central Apennines, Italy. *Computers & Geosciences*, 24(4): 373–385.
- 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): 15–31.
- Bartelletti C, Giannecchini R, D'Amato Avanzi G *et al.*, 2017. The influence of geological-morphological and

- land use settings on shallow landslides in the Pogliaschina T. basin (northern Apennines, Italy). *Journal of Maps*, 13(2): 142–152.
- Brenning A, Schwinn M, Ruiz-Páez A P *et al.*, 2015. Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province. *Nat. Hazards Earth Syst. Sci.*, 15: 45–57.
- Bui D T, Lofman O, Revhaug I *et al.*, 2011. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Natural Hazards*, 59(3): 1413–1444.
- Cees van Westen, Fan Xuanmei, Huang Runqiu, 2010. 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. In: Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies. The Safeland Project Reports.
- Cerovski-Darriau C, Roering J J, 2016. Influence of anthropogenic land-use change on hillslope erosion in the Waipaoa River Basin, New Zealand. *Earth Surface Processes and Landforms*, 41(15): 2167–2176.
- Chau K T, Chan J E, 2005. Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island. *Landslides*, 2(4): 280–290.
- China Institute for Geo-Environment Monitoring (CIGEM), 2007a. Rockfall and Landslide Disaster Map of China. Beijing: SinoMaps Press. (in Chinese)
- China Institute for Geo-Environment Monitoring (CIGEM), 2007b. Debris Flow Disaster Map of China. Beijing: SinoMaps Press. (in Chinese)
- China Institute for Geo-Environment Monitoring (CIGEM), 2013. China Geological Hazard Bulletin (2004–2012). China Geological Environmental Monitoring Institute Web. <http://www.cigem.gov.cn/>. Accessed on 20 June 2014.
- Chung C-J F, Fabbri A G, 2003. Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*, 30(3): 451–472. (in Chinese)
- Computer Network Information Center of CAS, 2015. Geospatial Data Cloud. <http://www.gscloud.cn/>. Accessed on August, 2015. (in Chinese)
- Devkota K C, Regmi A D, Pourghasemi H R *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. *Natural Hazards*, 65(1): 135–165.
- Eeckhaut M, Hervás J, Jaedicke C *et al.*, 2011. Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides*, 9(3): 357–369.
- Falaschi F, Giacomelli F, Federici P *et al.*, 2009. Logistic regression versus artificial neural networks: Landslide susceptibility evaluation in a sample area of the Serchio River valley, Italy. *Natural Hazards*, 50(3): 551–569.
- García-Rodríguez M J, Malpica J A, Benito B *et al.*, 2008. Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. *Geomorphology*, 95(3): 172–191.
- Gill J C, Malamud B D, 2017. Anthropogenic processes, natural hazards, and interactions in a multi-hazard framework. *Earth-Science Reviews*, 166: 246–269.
- Guzzetti F, Peruccacci S, Rossi M *et al.*, 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, 98(3): 239–267.
- Hartmann J, Moosdorf N, 2012. The new global lithological map database GLiM: A representation of rock properties at the earth surface. *Geochemistry, Geophysics, Geosystems*, 13(12): 1–37.
- Hu Y, Wang J, Li X *et al.*, 2011. Geographical detector-based risk assessment of the under-five mortality in the 2008 Wenchuan earthquake, China. *PloS One*, 6(6): e21427.
- Huang R Q, 2007. Large-scale landslides and their sliding mechanisms in China since the 20th century. *Chinese Journal of Rock Mechanics and Engineering*, 26(3): 433–454. (in Chinese)
- Institute of Geographic Sciences and Natural Resources Research of CAS, 2015. Data Sharing Network of Earth System Science. <http://www.geodata.cn/Portal/index.jsp>. Accessed on June, 2015. (in Chinese)
- Jadda M, Shafri HZM, Mansor SB, 2011. PFR model and GiT for landslide susceptibility mapping: A case study from Central Alborz, Iran. *Natural Hazards*, 57(2): 395–412.
- King G, Zeng L, 2001. Logistic regression in rare events data. *Political Analysis*, 9(2): 137–163.
- Lee S, Pradhan B, 2006. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic re-

- gression models. *Landslides*, 4(1): 33–41.
- Li Y, Meng H, Dong Y *et al.*, 2004. Main types and characteristics of geo-hazard in China: Based on the results of geo-hazard survey in 290 counties. *The Chinese Journal of Geological Hazard and Control*, 15(2): 29–34. (in Chinese)
- Li Y, Qu X, Yang X *et al.*, 2013. The spatial and temporal distribution of China geo-hazard and key prevention area. *The Chinese Journal of Geological Hazard and Control*, 24(4): 71–78. (in Chinese)
- Luo W, Jasiewicz J, Stepinski T *et al.*, 2016. Spatial association between dissection density and environmental factors over the entire conterminous United States. *Geophysical Research Letters*, 43(2): 692–700.
- Ministry of Land and Resources of China (MLRC), 2013. Report on Geological Disaster Situation. Ministry of Land and Resources of China Web. <http://www.mlr.gov.cn/dzhj/dzzh/zqxqbg/>. Accessed on 20 May 2014. (in Chinese)
- National Disaster Reduction Center of China (NDRCC), 2013. Yesterday's Disaster 2004–2012. Ministry of Civil Affairs National Disaster Reduction Center. (in Chinese)
- Nourani V, Pradhan B, Ghaffari H *et al.*, 2014. Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. *Natural Hazards*, 71(1): 523–547.
- Ohlmacher G C, Davis J C, 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology*, 69(3): 331–343.
- Poiraud A, 2014. Landslide susceptibility-certainty mapping by a multi-method approach: A case study in the Tertiary basin of Puy-en-Velay (Massif central, France). *Geomorphology*, 216: 208–224.
- Pourghasemi H, Moradi H, Aghda S F, 2013. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural Hazards*, 69(1): 749–779.
- Ramani S E, Pitchaimani K, Gnanamanickam V R, 2011. GIS based landslide susceptibility mapping of Tevankarai Ar sub-watershed, Kodaikkanal, India using binary logistic regression analysis. *Journal of Mountain Science*, 8(4): 505–517.
- Samia J, Temme A, Bregt A *et al.*, 2017. Characterization and quantification of path dependency in landslide susceptibility. *Geomorphology*, 292: 16–24.
- Sheng L, Wang W, Zhong S, 2013. China Statistical Yearbook. Beijing: China Statistics Press. (in Chinese)
- Shirzadi A, Saro L, Joo O H *et al.*, 2012. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. *Natural Hazards*, 64(2): 1639–1656.
- Swets J A, 1988. Measuring the accuracy of diagnostic systems. *Science*, 240(4857): 1285–1293.
- Tarolli P, Sofia G, 2016. Human topographic signatures and derived geomorphic processes across landscapes. *Geomorphology*, 255: 140–161.
- Tsangaratos P, Ilia I, 2016. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides*, 13(2): 305–320.
- The State Council, 2004. Geological Disaster Prevention Regulations. http://www.gov.cn/gongbao/content/2004/content_63064.htm. Accessed on 15 May 2014. (in Chinese)
- Wang J F, Li X H, Christakos G *et al.*, 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *International Journal of Geographical Information Science*, 24(1): 107–127.
- Xu C, Xu X, Dai F *et al.*, 2013. Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake of China. *Natural Hazards*, 68(2): 883–900.
- Yen S J, Lee Y S, Lin C H *et al.*, 2006. Investigating the effect of sampling methods for imbalanced data distributions. In: Systems, Man and Cybernetics, 2006. SMC'06. *IEEE International Conference*, 5: 4163–4168.
- Zhang H B, Tong X, 2009. Public policy in a high-risk society. *Journal of Nanjing Normal University (Social Science)*, (6): 23–28. (in Chinese)