



OPEN

Comparison of earthquake-induced shallow landslide susceptibility assessment based on two-category LR and KDE-MLR

Xinyue Fan^{1,3}, Bin Liu^{1,2}✉, Jie Luo^{1,2}, Ke Pan², Suyue Han^{1,3} & Zhongli Zhou^{1,2,3}

Geological hazards caused by strong earthquakes have caused continuous social and economic losses and destruction of the ecological environment in the hazard area, and are mostly manifested in the areas with frequent occurrence of geological hazards or the clustering of geological hazards. Considering the long-term nature of earthquakes and geological disasters in this region, this paper takes ten earthquake-stricken areas in Wenchuan earthquake zone as examples to collect shallow landslide data in 2010, combined with the spatial location of landslides and other factors. Kernel density estimation (KDE) method is used to analyze the spatial characteristics of shallow landslide. Taking the space of shallow landslide as the characteristic variable and fully considering the regulating factors of earthquake-induced landslide: terrain complexity, distance to river, distance to fault, distance to road, lithology, normalized vegetation difference index (NDVI) and ground peak acceleration (PGA) as independent variables, based on KDE and polynomial logistic regression (MLR), A quantitative model of shallow landslide in the earthquake area is constructed. The results show that: (1) PGA has the greatest impact on landslide in the study area. (2) Compared with the two-category logistic regression (two-category LR) model, the susceptibility map of landslide prediction results based on the KDE-MLR landslide susceptibility prediction model is more consistent with the actual situation. (3) The prediction accuracy of the model validation set is 70.7%, indicating that the landslide susceptibility prediction model based on KDE-MLR can effectively highlight the spatial characteristics of shallow landslides in 10 extreme disaster areas. The research results can provide decision-making basis for shallow landslide warning and post-disaster reconstruction in earthquake-stricken areas.

Earthquakes play a relatively prominent role in various natural hazards and are extremely destructive. There are many earthquakes every year around the world, and China's seismic fault zone is greatly affected by earthquake hazards. According to incomplete statistics, the cumulative death toll in China resulting from earthquakes that occurred between 1993 and 2016 accounted for 50% of global earthquake-related deaths. Moreover, more than two-thirds of China's provinces were affected by earthquakes, resulting in millions of casualties and hundreds of millions of victims^{1,2}. After a strong earthquake, the geological structure of the affected area becomes unstable and susceptible to geological hazards such as landslides, debris flows, dammed lakes, and avalanches that occur frequently in extremely earthquake-stricken areas³. Among these hazards, landslides are one of the most frequent and destructive earthquake-induced geological hazards in the world^{4–6}. According to a large amount of data, losses caused by landslides have far exceeded losses from earthquakes themselves. This had a huge impact on the sustainable development of society and the economy⁷. The Wenchuan earthquake ($M_w=7.9$) that occurred on the Longmenshan fault zone in 2008 was the most destructive earthquake in the past 100 years. The energy released was approximately three times that of the Tangshan earthquake. In addition, a post-earthquake emergency investigation of geological hazards showed that there was a total of 8060 hidden geological hazard points across 39 earthquake-stricken counties in Sichuan Province, fully demonstrating that the temporal and spatial effects of geological hazards after the Wenchuan earthquake were significant⁸.

Previous research on the evaluation of landslides, has mainly been conducted on three aspects: landslide sensitivity^{9–12}, landslide risk^{13–15} and landslide susceptibility¹⁶. Among them, landslide susceptibility can be defined

¹Geomathematics Key Laboratory of Sichuan Province, Chengdu 610059, China. ²College of Mathematics and Physics, Chengdu University of Technology, Chengdu 610059, China. ³College of Management Science, Chengdu University of Technology, Chengdu 610059, China. ✉email: liubincim@163.com

as the spatial probability of landslide occurrence based on a set of geological and environmental conditions. A landslide susceptibility map reflects the spatial distribution of the landslide probability in an area¹⁶. This analysis can be traced back to the Japanese scholar Saito, who, in the 1960s, used the results of creep tests to predict the locations of shallow landslides. After completing a series of related theoretical systems, many scholars have conducted exploratory research on other quantitative spatial prediction models of regional geological hazards, including empirical models (such as fuzzy logic and generalized additive models)^{17,18}, statistical analysis models (such as the certainty coefficient method, weights of evidence and entropy models)^{19–21} and pattern recognition models (such as artificial neural networks, support vector machines (SVMs) and adaptive, neuro-fuzzy inference systems (ANFISs))^{22–24}. Two-category LR is a generalized linear regression analysis model widely used in exploring landslide susceptibility due to the simple calculations and physical clear meaning characteristic. However, this model is relatively simple and cannot handle complex issues in actual situations, some scholars have proposed MLR. The MLR is useful in situations with several dependent variables²⁵. Previous findings indicate that the two-category LR is widely used in shallow landslide susceptibility evaluations^{26,27}. The dependent variables of most two-category LR in susceptibility analysis of shallow landslides are binary logic occurrence variables (denoted as 1) and nonoccurrence variables (represented as 0). Notably, the intensity of regional shallow landslides (spatial effects) is not considered in this model resulting in significant forecast biases.

Tobler's first law of geography states that everything is related to other things but near things are more related than distant things; that is, there is a potential dependence between the observed data of certain variables in the same or different distribution areas²⁸. Landslides are spatially manifested as regions where geological hazards or groups of hazards frequently occur. In the past, the spatial quantitative modelling of landslides primarily comprised use of statistics to explore the characteristics and development of hazards (whether a hazard occurred or not). Although the spatial distribution of geological hazards appears random, it exhibits inherent regularity, thus spatial autocorrelation is widely used in the study of geological hazards^{29,30}. Zou et al.²⁹ established spatial regression models using the relevant geological hazard data, evaluated the impact of human activities on geological hazards in the Shennongjia Mountain area, and proposed effective prevention and control strategies. Liu et al.³¹ explored the spatial–temporal distribution and conditioning factors of geological hazards. Gonzalez et al.³² investigated the spatial distribution characteristics of seismic hazards in cities and the impact of river vaults on them using the conventional spatial autocorrelation method (SPAC) and Calicatas SPAC method. Previous findings indicate that evaluating the spatial characteristics of landslides improves the accuracy of spatial prediction of landslides.

The purpose of this paper is to build a model to predict shallow landslide susceptibility. The disaster intensity of regional shallow landslides was quantitatively characterized by the model, and an objective understanding of landslide spatial characteristics was used to accurately predict the influence of the landslide disasters. By identifying the shallow landslide susceptibility area, research on shallow landslide spatial characteristics can effectively reduce casualties and property losses. In this paper, ten earthquake-stricken areas affected by the Wenchuan earthquake are studied. The evaluation factors affecting the surface sensibility of shallow landslides are examined from the aspects of topography, geological environment and inducing conditions. KDE is used to describe the multiclassification spatial characteristics of shallow landslides, and the degree of aggregation and dispersion of shallow landslides in space is fully analyzed. A landslide disaster sensitivity prediction model based on KDE-MLR is constructed using MLR. The remainder of the paper is organized as follows: in "Introduction" Section, summarize the research progress of spatial quantitative modeling of landslides. In "Overview of the study area" Section, introduce the 10 severely stricken areas of the Wenchuan earthquake, and the data sources. In "Method" Section, introduce the research methods used in this paper, including: KDE, LR algorithm (two-category LR, MLR), Model performance evaluation index. In "Results" Section, construct the landslide susceptibility prediction model based on KDE-MLR, analyze the influence degree of landslide conditioning factors, evaluate the model and discuss the validity of the model. In "Discussion" Section, conclusions are presented. In "Conclusions" Section, discussion is presented.

Overview of the study area

This paper conducts a spatial prediction of landslides in the 10 extremely earthquake-stricken areas of the Wenchuan earthquake and improves the two types of dependent variables over those used in traditional landslide evaluations (1 for landslide occurrences and 0 for landslide non-occurrences). First, the spatial characteristics of landslides are analyzed by the KDE method. Second, by taking the spatial characteristics of landslides as the dependent variable, using the advantages of the MLR method, and expanding it into three categories, then construct a landslide susceptibility prediction model based on KDE-MLR. Finally, a spatial landslide susceptibility map is generated. The research framework, which is shown in Fig. 1 is divided into five steps:

Step 1 Collect landslide geological hazard data in the study area, including the spatial location of landslide points, the number of landslides, and related data on seven conditioning factors.

Step 2 Account for factors such as the spatial locations of the landslides, KDE is used to describe the spatial characteristics of the landslide in the study area.

Step 3 Use the MLR and take the spatial characteristics of landslides as dependent variables and the conditioning factors as independent variables, then construct a landslide susceptibility prediction model based on KDE-MLR.

Step 4 Randomly select 80% of the spatial characteristics of landslides in the study area as training samples and 20% as verification samples, then continuously adjust the model parameters and evaluate the effectiveness and performance of the model.

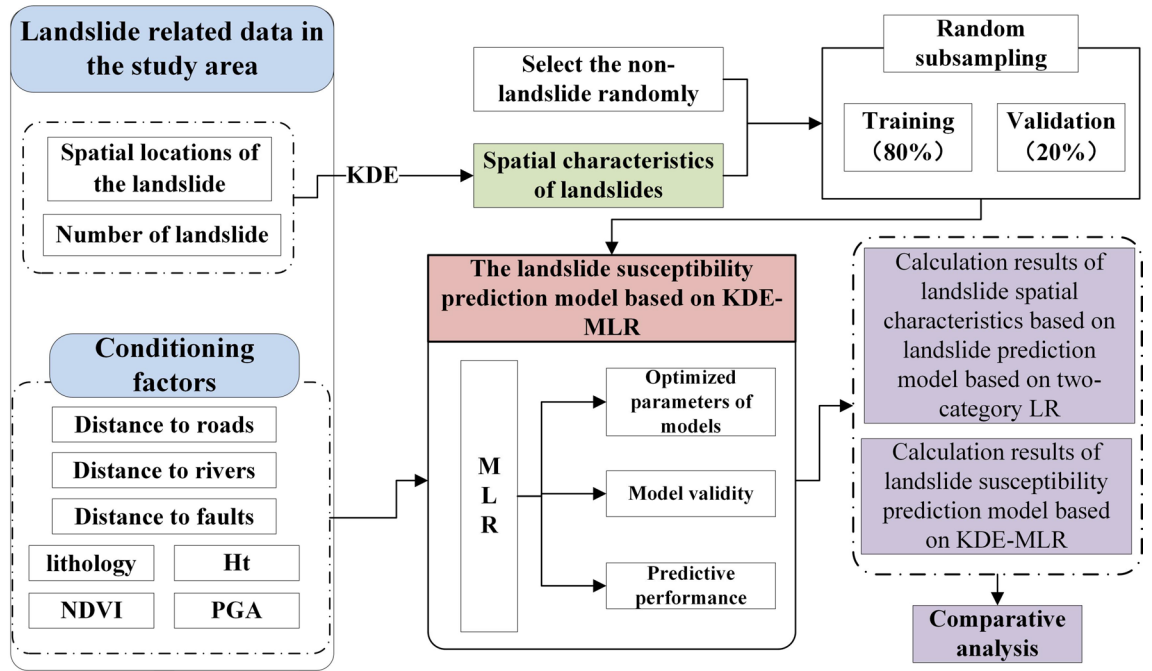


Figure 1. Methodological flowchart of the study.

Step 5 Compare and analyze the susceptibility maps generated by the Landslide prediction model based on two-category LR and the landslide susceptibility prediction model based on KDE-MLR.

Study area. In 2008, the 5.12 Wenchuan earthquake shook Sichuan, Gansu and Shaanxi provinces. This paper studies the 10 extremely earthquake-stricken areas of the Wenchuan earthquake that were announced by the Ministry of Land and Resources of China. These 10 extremely earthquake-stricken areas suffered the most damage during the earthquake and are all located in Sichuan Province: Wenchuan, Beichuan, Qingchuan, Maoxian, Anxian, Mianzhu, Shifang, Dujiangyan, Pingwu and Pengzhou (Fig. 2). The study area covers approximately 26,000 km² and has a total population of over 3.5 million. The topography of the region is relatively complex, with elevations ranging between 490 and 5600 m, and the average annual precipitation is between 490 and 1400 mm³³. In addition, the terrain in this area is complex, located in the Longmenshan fault zone (one of the fault zones with the largest topographic gradient in the world), and there have been many strong earthquakes. By investigating the occurrence of geological hazards (e.g., landslides) in the extremely earthquake-stricken areas

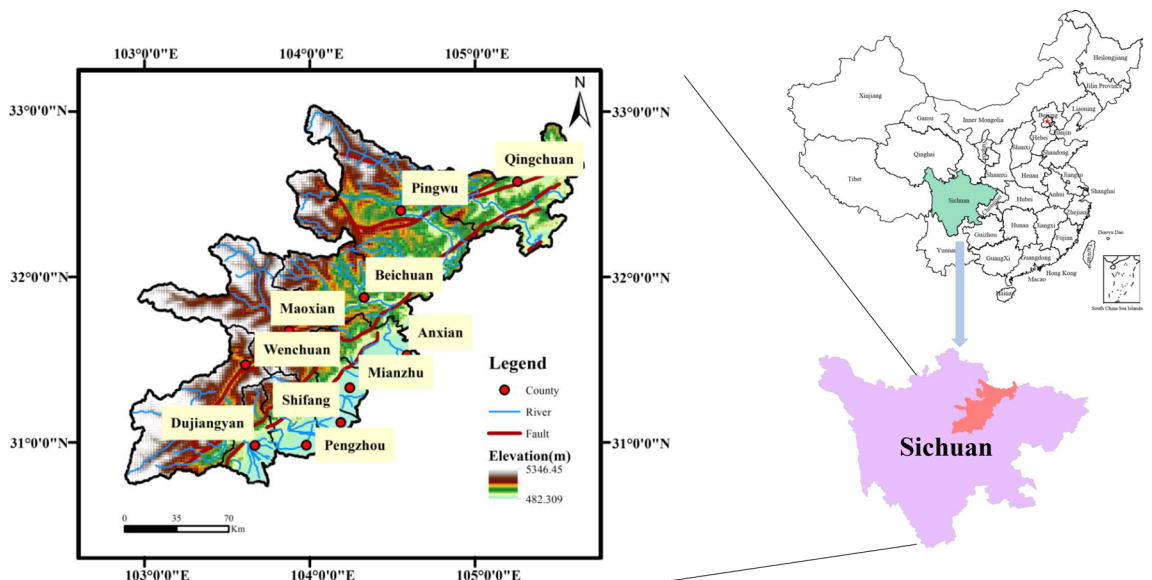


Figure 2. Map of the study area (Names of software: ArcGIS 10.5 and Office 2019).

of the Wenchuan earthquake, landslides can be found to have increased sharply 2–3 years after the earthquake³⁴. Therefore, this paper takes 2010 as an example to analyze the spatial distribution characteristics of landslides in the 10 extremely earthquake-stricken areas and establishes a quantitative spatial landslide model to provide a reference for hazard monitoring and decision-making to formulate emergency plans.

Data sources. *Data preparation.* In this study, the landslide that occurred in 2010 (2 years after the Wenchuan earthquake) was selected as the research object due to the long-term nature of the earthquake and geological hazards. The spatial location and number of shallow landslides in the study area of this paper were obtained by Landsat-8 30 m resolution remote sensing data interpretation, and grid processing was carried out with 60 m*60 m pixels in ArcGIS 10.5 software. According to the results of remote sensing analysis and statistics, there were 885 landslide points in the study area in 2010. However, in real life, the occurrence of shallow landslides often changes the whole area, which may make the model ignore the change in nonlandslide points in the calculation process. This reduces the accuracy of the landslide susceptibility evaluation map due to inconsistency with the actual situation. A total of 885 non-landslide points were randomly selected in a 1:1 ratio outside the buffer zone, 10 km away from the landslide point to reduce errors caused by this situation. The spatial distribution of the 885 non-landslide points is shown in Fig. 3.

Conditioning factors of landslides. The selection of conditioning factors is one of the most important tasks in spatial predictive modelling³⁵. Although there is currently no clear guideline for selecting conditioning factors when constructing prediction models, the selection of conditioning factors usually depends on the characteristics of the study area and the scale of the hazard^{36,37}. For this paper, we selected seven landslide conditioning factors or independent variables based on the quality and availability of data: terrain information entropy (Ht) (Fig. 4a), PGA (Fig. 4b), distance to roads (Fig. 4c), distance to rivers (Fig. 4d), distance to faults (Fig. 4e), NDVI (Fig. 4f) and lithology (Fig. 4g). Among them, Ht is a topographic and geomorphic factor, distance to rivers, distance to faults, lithology, and NDVI are geological environment factors, and distance to roads and PGA are inducing factors.

All conditioning factors were generated using ArcGIS software. The topographic information entropy was calculated using 1:35 0000 DEM data³⁰. The distance to the fault was derived from the 1:500,000 regional geological map³⁸. The distance to the river was obtained from 1:3 million National Hydrogeological Atlas (<https://www.osgeo.cn/map/m04dd>). Statistical yearbook and historical data were obtained through remote sensing interpretation. Lithology data were retrieved from 1:500,000 regional geological map^{38,39}. NDVI was obtained from Landsat-7 30-m resolution remote sensing interpretation. PGA data were retrieved from the United States

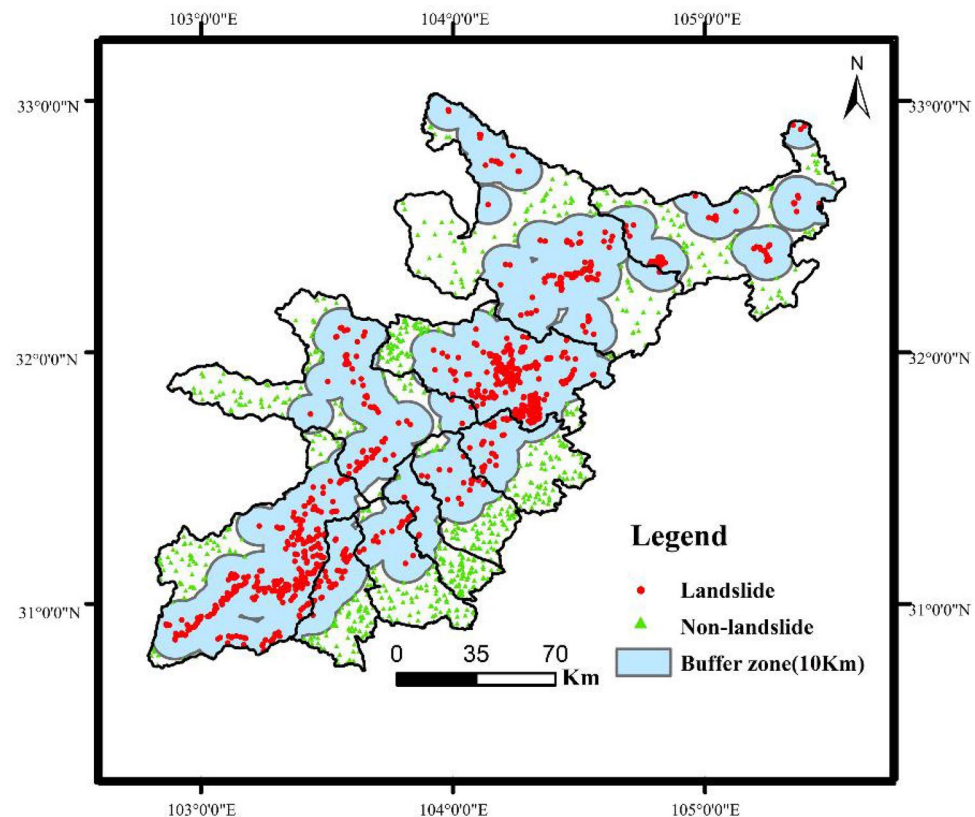


Figure 3. Spatial distribution of landslides (Name of software: ArcGIS 10.5).

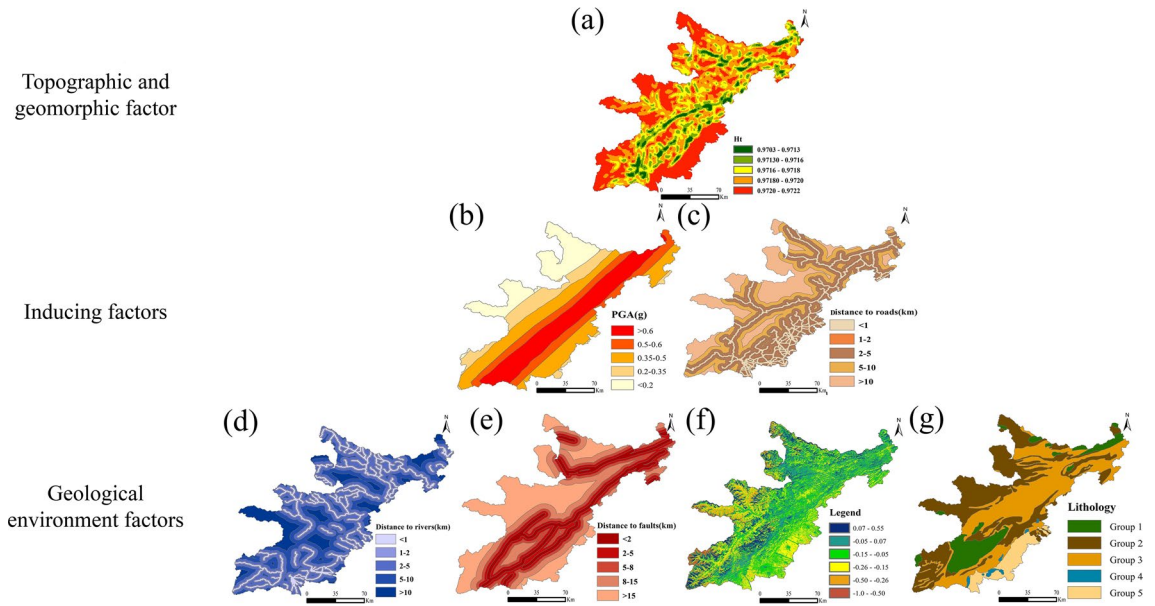


Figure 4. Conditioning factors of landslides: (a) Ht, (b) PGA, (c) distance to roads; (d) distance to rivers, (e) distance to faults, (f) NDVI, and (g) lithology (Names of software: ArcGIS 10.5 and Office 2019).

Geological Survey (USGS) report. The 7 conditioning factor layers were gridded with 60×60 m pixels in ArcGIS 10.5 software, generating a total of 1,002,7131 grids, which are used as input data for generating the landslide susceptibility map, as shown in Fig. 4.

Method

Kernel density estimation (KDE). In order to deeply reveal the spatial distribution characteristics of earthquake and landslide geological hazards, this paper selects the KDE to quantitatively calculate the spatial distribution characteristics of landslides, and analyzes the spatial continuous trend of landslides in the hardest hit area of the Wenchuan earthquake. KDE is mainly used to estimate the density of point or line features around each output raster cell⁴⁰. Through the two-dimensional greyscale expression or three-dimensional surface expression of the earthquake-induced landslide core density calculation results, the distribution characteristics of the cluster or dispersion of the landslide point groups can be simply and intuitively obtained. The two-dimensional kernel density is an extension of the one-dimensional kernel density, and the density value at a certain point is estimated by calculating the kernel estimator:

$$f(s) = \frac{1}{nh^2} \sum_{i=1}^n K_0\left(\frac{|dis|}{h}\right) \tag{1}$$

In formula (1), s is the spatial location of an estimated point, n is the number of landslide points within the bandwidth, d_{is} is the distance from the i_{th} landslide point to s , and h is the bandwidth. The bandwidth formula is as follows:

$$h = 0.9 \times \min\left(SD, \sqrt{\frac{1}{\ln(2)} \times D_m}\right) \times N^{-0.2} \tag{2}$$

$$SD = \sqrt{\frac{\sum_{i=1}^M (x_i - \bar{X})^2}{M} + \frac{\sum_{i=1}^M (y_i - \bar{Y})^2}{M}} \tag{3}$$

In formulas (2) and (3), D_m is the median distance from each landslide point to the average centre, N is the sum of the landslide points, SD is the standard distance, M is the total number of landslide points, x_i and y_i are the coordinates of the landslide points i and \bar{X} and \bar{Y} are the average center of the landslide points. K_0 is the kernel function, and its formula is shown:

$$K_0\left(\frac{|d_{is}|}{h}\right) = \begin{cases} \frac{3}{\pi} \left(1 - \frac{|d_{is}|^2}{h^2}\right)^2 & 0 < d_{is} \leq h \\ 0 & d_{is} > h \end{cases} \tag{4}$$

LR algorithm. *Two-category LR.* The two-category LR method is a nonlinear, multivariate statistical method that is widely used to evaluate landslide susceptibility^{41–43}. The two-category LR model can perform different types of independent variable analysis, including analyses of continuous variables and discrete variables. It comprehensively evaluates various conditioning factors based on actual landslide point samples and non-landslide point samples and can better solve the problem of factor interdependence^{44,45}. The two-category LR method does not require the independent variables to conform to the normal distribution and it has no restriction on the distribution of the identified variables. It can be used to predict the probability of dependent variables with binomial characteristics. When the dependent variable in the two-category LR model is of two types, namely, occurrence of a landslide (denoted as 1) and nonoccurrence of a landslide (denoted as 0), the relationship between the probability of landslide occurrence and the conditioning factor is shown in the following formula⁴⁶:

$$Z = a_0 + a_1x_1 + a_2x_2 + \dots + a_ix_i \quad (5)$$

$$P = \frac{\exp(z)}{1 + \exp(z)} \quad (6)$$

In formulas (5) and (6), is the two-category LR prediction value representing the probability of landslide occurrence and having a value in the range [0, 1], Z represents the sum of the linear weight values after the variables are superimposed, is the constant term of the logistic regression, x_i represents each conditioning factor and is the two-category LR coefficient of the i th conditioning factor.

MLR. In practical problems, the dependent variable may have multiple values. When there are more than two dependent variable values, MLR analyses need to be used. MLR analysis is a variant of two-category LR and its basic concept is essentially the same as that of two-category LR^{47,48}. MLR analysis assumes that dependent variable categories are completely independent. It also takes one of the categories as a reference group and calculates a reference group regression coefficient⁴⁹.

The MLR model can be considered a $J-1$ two-category LR model of J dependent variables and each independent dependent variable is compared with the reference group to calculate the advantage of the reference group⁵⁰. Assuming that $Y = J$ is the reference group, the LR model of each category is as follows:

$$\lambda_j = \log \left(\frac{P_j}{P_J} \right) = \alpha_j + \beta_{1j}x_{1j} + \dots + \beta_{ij}x_{ij} \quad (7)$$

In formula (7), $j = 1, \dots, J - 1$, α_j is the intercept, x_i represents each conditioning factor, and β_{ij} is the MLR coefficient of the No. i th conditioning factor. The MLR model is expressed in the form of probability as follows:

$$P_j = \frac{\exp(\lambda_j)}{1 + \sum_{i=1}^{J-1} \exp(\lambda_j)} \quad (8)$$

$$P_J = \frac{1}{1 + \sum_{i=1}^{J-1} \exp(\lambda_j)} \quad (9)$$

In formulas (8) and (9), P_j represents the probability that the dependent variable is a certain category and P_J represents the probability of the reference group. The values of the dependent variable Y in this study are 0, 1 and 2. Since most of the locations in the 10 extremely earthquake-stricken areas are non-landslide points, the dependent variable $Y = 0$ is set as the reference group.

Model performance evaluation index. To evaluate the performance of both the two-category LR and KDE-MLR landslide susceptibility prediction models, this paper uses an error matrix to evaluate the accuracy of each model⁵¹, as shown in Table 1. The overall accuracy is principally used to determine the performance of the models and is obtained by comparing the predicted model category with the actual category.

In MLR analysis, training sets are used to train the model. When predicting the remaining test datasets, the probability of each category is compared by calculating the predicted dependent variable probabilities for

Actual category	Forecast category		
	0	1	2
0	N1	N2	N3
1	N4	N5	N6
2	N7	N8	N9

Table 1. Error matrix used for the accuracy assessment. 0, 1, 2 are dependent variables respectively; N1 means the actual category is 0, the forecast category is 0, and so on.

different categories and the reference group. The highest category is the result of the final classification⁵⁰. The effectiveness of the model is determined by its overall accuracy rate, which expresses the accuracy of the model. The accuracy rate is the ratio of the number of samples that predict the correct classification of the model to the total number of samples. The specific formula is as follows:

$$\text{Overall accuracy} = \frac{N1 + N5 + N9}{\sum_{i=1}^9 Ni} \quad (10)$$

Results

Calculation results of the spatial characteristics of landslides. To further understand the spatial distribution characteristics of landslides, this paper comprehensively considers the spatial location and number of landslides. This paper also uses the KDE formula in Sect. [Kernel density estimation \(KDE\)](#) to calculate the spatial characteristics of landslides of the 10 extremely earthquake-stricken areas, as shown in Fig. 5.

Based on the value of spatial distribution characteristics that calculated by KDE, this paper uses the average value (3.3) as the demarcation point to divide the spatial characteristics of landslides into two levels (High, Medium). Thus, the spatial characteristics of landslides of all landslide points is divided into 2 levels, High and Medium. In Fig. 5, approximately 62% of the landslide points have high spatial aggregation of landslides. These points mainly lie in Wenchuan and Beichuan. The KDE value of 337 landslide disaster points in the entire study area is lower than 3.3, the landslide spatial characteristics is moderate, and the distribution in other counties (cities) is relatively uniform.

Analysis of the influence degree of factors on landslides. *Analysis of landslide conditioning factors based on LR.* In the two-category LR landslide prediction model, this paper randomly selects 80% of the 885 landslide and non-landslide points (619 landslide points and 619 non-landslide points) as the training set and 20% (266 landslide points and 266 non-landslide points) as the validation set. Two-category LR is performed based on the extracted landslide conditioning factors and training samples, and the regression coefficient of each conditioning factor is obtained. As shown in Table 2, the greater the absolute value of the coefficient, the greater the impact it has on landslide susceptibility.

Table 2 shows that the absolute value of the PGA regression coefficient is the highest (0.418), which indicates that this factor has the largest influence on earthquake-induced landslides. The lithology, distance to rivers and Ht also have relatively large impacts on landslides in the study area, with absolute values of their regression coefficients between 0.2 and 0.4. The regression coefficients of the distance to roads, NDVI and distance to faults are all less than 0.1, indicating that these three conditioning factors have relatively little correlation with landslides.

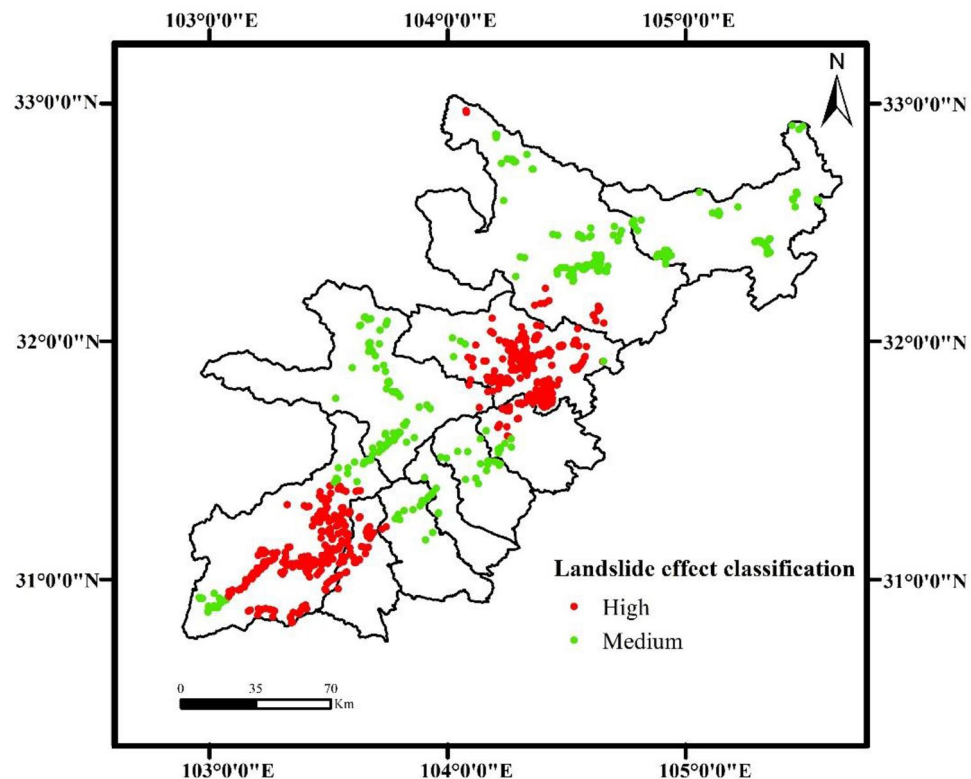


Figure 5. Calculation results of the spatial characteristics of landslides (Name of software: ArcGIS 10.5).

Serial number	1	2	3	4	5	6	7
Conditioning factor	Distance to roads	PGA	Distance to rivers	Distance to faults	NDVI	Ht	Lithology
Absolute value of the standardized regression coefficient	0.0575	0.418	0.3663	0.008	0.052	0.2611	0.381

Table 2. Two-category LR coefficient results.

Analysis of landslide conditioning factors under KDE-MLR. In the landslide susceptibility prediction model based on KDE-MLR, the dependent variable (the spatial characteristics of the landslides) is a categorical variable and the independent variable (conditioning factor) is a categorical, or continuous, variable in the MLR analysis⁵². MLR uses different criteria in determination of the MLR coefficients of discrete variables and continuous variables. Therefore, continuous variables were classified in this study and treated as separate variables (Table 3).

The results of the landslide susceptibility prediction model based on KDE-MLR were calculated by SPSS Modeler software. The same sample ratio used in the two-category LR landslide prediction model was used to divide the dataset. In this study, 80% of the dataset was used as the training set (0: 708 individuals in this set, 1: 269 individuals in this set, and 2: 439 individuals in this set) and 20% was used as the validation set (0: 177 individuals in this set, 1: 67 individuals in this set, and 2: 110 individuals in this set) (0 represents no landslide point). The dataset comprised 3 categories, with $Y = 0$ as the reference group. Two groups of regression coefficients with different spatial characteristics of landslides levels were obtained. The results are shown in Fig. 6:

- (1) When the dependent variable is the moderate spatial characteristics of landslides ($Y = 1$), lithology and NDVI have the highest effect on the region, and the two variables were positively correlated. The result indicated that the probability of $Y = 1$ was higher in areas with softer lithology and lower NDVI values compared with the reference group. This implies that areas with less vegetation cover and softer strata were prone to moderate spatial aggregation of landslides. The regression coefficients of each level of the distance to the road, the distance to the fault zone, the distance to the river, PGA, and Ht were all negative values. This finding implies that the intermediate level of spatial characteristics of landslides was negatively correlated with these factors. These results indicate that shorter distance to the road, shorter the distance to the fault zone, and shorter distance to the river are associated with moderated spatial characteristics of landslides are more moderate. In addition, the regression coefficients of PGA and Ht fluctuate according to the classification.
- (2) When the dependent variable represents the high spatial characteristics of landslides ($Y = 2$), the regression coefficients of NDVI and lithology are basically consistent with the case of $Y = 1$. In this case, the high spatial characteristics of landslides exhibited a high dependence on lithology and NDVI, and the spatial characteristics of landslides were positively correlated with lithology and NDVI. The regression coefficient value gradually decreased with increase in the distance to faults. However, all coefficients had positive values, indicating that landslide susceptibility was higher in areas closer to the faults and the influence of this factor on $Y = 2$ was higher than its influence on $Y = 0$. The regression coefficients of Ht had negative values, implying a larger value of Ht was associated with a higher correlation of spatial characteristics of landslides. The regression coefficient value of PGA was small, but the change was relatively stable. The regression coefficients of the distance to the road and the distance to the river exhibited distinct correlations according to the classification. The absolute value of the regression coefficient was largest when the distance to the road was 1–2 km and the distance to the river was less than 1 km. The impact on the landslide disaster was largest under these conditions.

The importance of each conditioning factor is obtained by combining the two sets of regression coefficients. From Table 4, it can be concluded that the influence on landslide geological hazards from high to low is PGA, NDVI, lithology, distance to roads, distance to rivers, distance to faults, and Ht.

Serial number	Conditioning factor	Grading situation	Classification based on
1	Distance to roads	Level 1: ≤ 1 km; Level 2: 1 ~ 2 km; Level 3: 2 ~ 5 km; Level 4: 5 ~ 10 km; Level 5: ≥ 10 km	53
2	Distance to faults	Level 1: ≤ 2 km; Level 2: 2 ~ 5 km; Level 3: 5 ~ 8 km; Level 4: 8 ~ 15 km; Level 5: ≥ 15 km	54
3	Distance to rivers	Level 1: ≤ 1 km; Level 2: 1 ~ 2 km; Level 3: 2 ~ 5 km; Level 4: 5 ~ 10 km; Level 5: ≥ 10 km	55
4	PGA	Level 1: < 0.2 g Level 2: 0.2 ~ 0.35 g Level 3: 0.35 ~ 0.5 g Level 4: 0.5 ~ 0.6 g Level 5: > 0.6 g	56
5	NDVI	Level 1: 0.0425 ~ 0.5456; Level 2: $-0.0847 \sim 0.0425$; Level 3: $-0.2181 \sim -0.0847$; Level 4: $-0.4666 \sim -0.2181$; Level 5: $-1 \sim -0.4666$	Natural fracture method
6	Ht	Level 1: > 0.9720 ; Level 2: 0.9718 ~ 0.9720; Level 3: 0.9716 ~ 0.9718; Level 4: 0.9712 ~ 0.9716; Level 5: < 0.9712	Natural fracture method
7	Lithology	Level 1: hardest (Rb > 60 MPa); Level 2: harder (Rb = 40 ~ 60 MPa); Level 3: softer (Rb = 20 ~ 40 MPa); Level 4: weak (Rb < 20 MP a); Level 5: loose (Unstable rock formation)	

Table 3. Classification of Influential Factors.

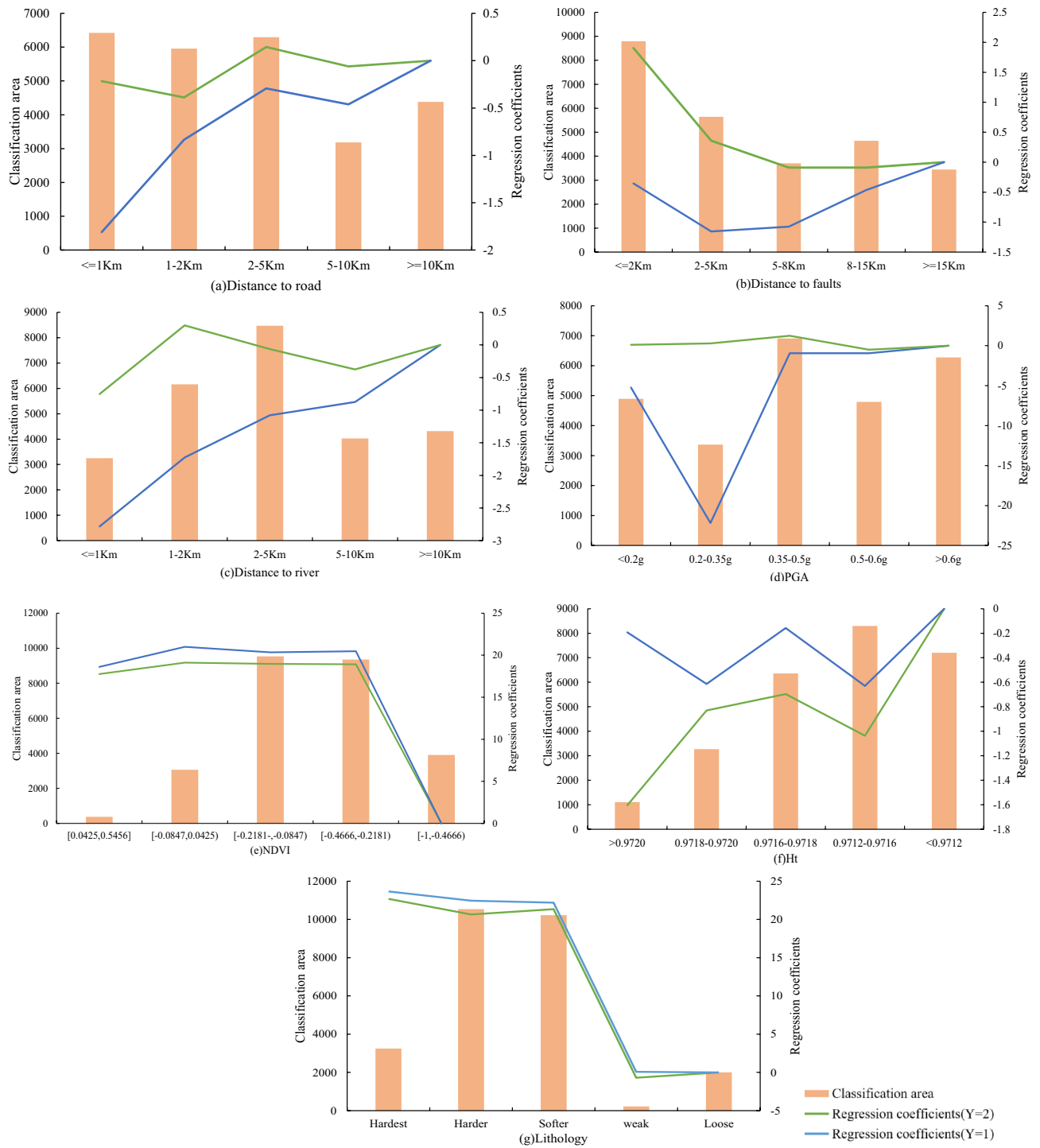


Figure 6. MLR coefficients under the grading of various conditioning factors (Name of software: Office 2019).

Serial number	1	2	3	4	5	6	7
Conditioning factor	Distance to roads	PGA	Distance to rivers	Distance to faults	NDVI	Ht	Lithology
Predictor importance	0.10	0.42	0.10	0.09	0.12	0.05	0.12

Table 4. The importance of conditioning factors.

Performance evaluation of landslide prediction model. The Error matrix in Table 1 is used to calculate the overall accuracy of the KDE-MLR landslide susceptibility prediction model. The results in Table 5 show that the overall accuracy of the training datasets used to build the model is 71%, the overall accuracy of the validation datasets used to verify the model is 70.7%, and the overall accuracy is more than 70%. This demonstrates that the landslide susceptibility prediction model based on KDE-MLR credibly predicts the probability distribution of the spatial aggregation of landslides in the study area⁵⁷. Similarly, the two-category LR landslide prediction model was verified. At the same time, the two-category LR landslide prediction model is verified, and its overall accuracy rate is about 72%, indicating that the two-category LR landslide disaster prediction model is also feasible, because the two models have similar accuracy rates. Therefore, the problem of error caused by model accuracy was excluded in the follow-up study.

Comparative analysis of landslide susceptibility. This paper assigns a corresponding weight value to each conditioning factor based on the regression coefficient. Then, by superimposing each conditioning factor layer, distribution maps of landslide predictions for both the two-category LR-based landslide prediction model (Fig. 7a) and the KDE-MLR model (Fig. 7b) are obtained.

In Table 6 and Fig. 7:

- (1) In the predicted results of the two models above, more than 50% of landslides occur in areas with high landslide susceptibility. In the KDE-MLR model, the area with a landslide susceptibility level of 3 is approximately 5441 km². In the three-level landslide susceptibility area, divided by the two-category LR model according to the natural segment point method, the area of the high-level susceptibility is approximately 7104 km², and the landslide density is significantly lower than the predicted result in the KDE-MLR model. Overall, the landslide susceptibility prediction model based on KDE-MLR constructed in this paper can effectively highlight the spatial characteristics of landslides in 10 extremely earthquake-stricken study areas.
- (2) The landslide susceptibility prediction model based on KDE-MLR is used to predict the hazard level areas with high spatial characteristics of landslides. According to Fig. 7, it can be seen that most of them are distributed in Beichuan and Wenchuan areas. On the whole, with the increase of susceptibility level, the regional landslide density has also gradually increased. Generally, as the susceptibility level increases, the

	Observed	Predicted			Overall accuracy (%)
		0	1	2	
Training step	0	566	42	100	71.0
	1	116	106	47	
	2	80	26	333	
Validation step	0	132	14	31	70.7
	1	31	26	10	
	2	12	5	93	

Table 5. Accuracy results of the susceptibility prediction model of landslides based on KDE-MLR.

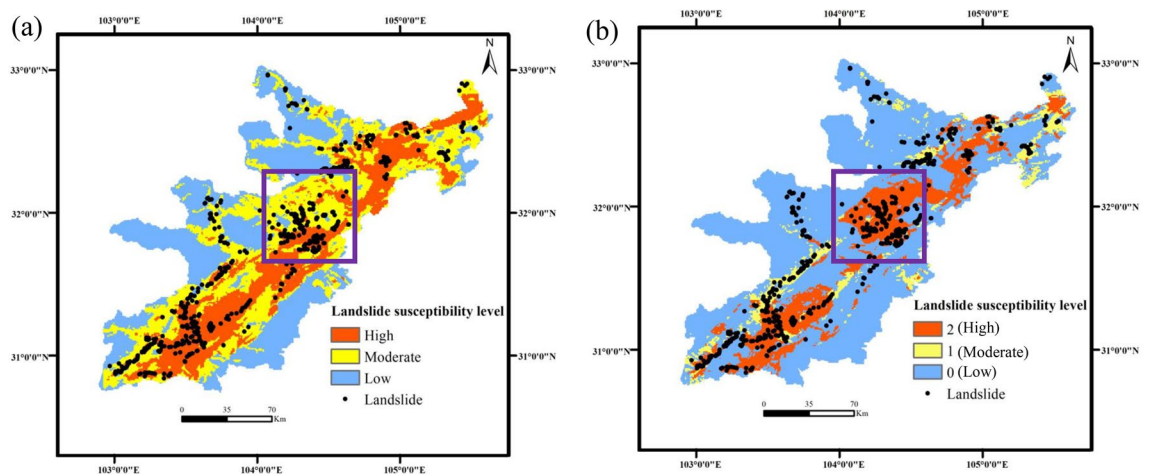


Figure 7. Spatial distribution of the prediction results of landslides in extremely earthquake-stricken areas: (a) Landslide prediction distribution map based on the two-category LR-based landslide prediction model; (b) Landslide prediction distribution map based on the KDE-MLR model (Names of software: ArcGIS 10.5 and Office 2019).

Model	Forecast result	Number of grids	Area (Km ²)	Area percentage (%)	Percentage of landslides (%)	Landslide density (10 ⁻³)
Landslide prediction model based on two-category LR	Low	3,659,903	9640	36.5	9	0.022
	Moderate	3,669,930	9666	36.6	33.2	0.08
	High	2,697,298	7104	26.9	57.8	0.19
The susceptibility prediction model of landslides based on KDE-MLR	0	7,079,154	18,645	70.6	28.8	0.036
	1	882,388	2324	8.8	17.6	0.17
	2	2,065,589	5441	20.6	53.6	0.229

Table 6. Forecast result statistics chart.

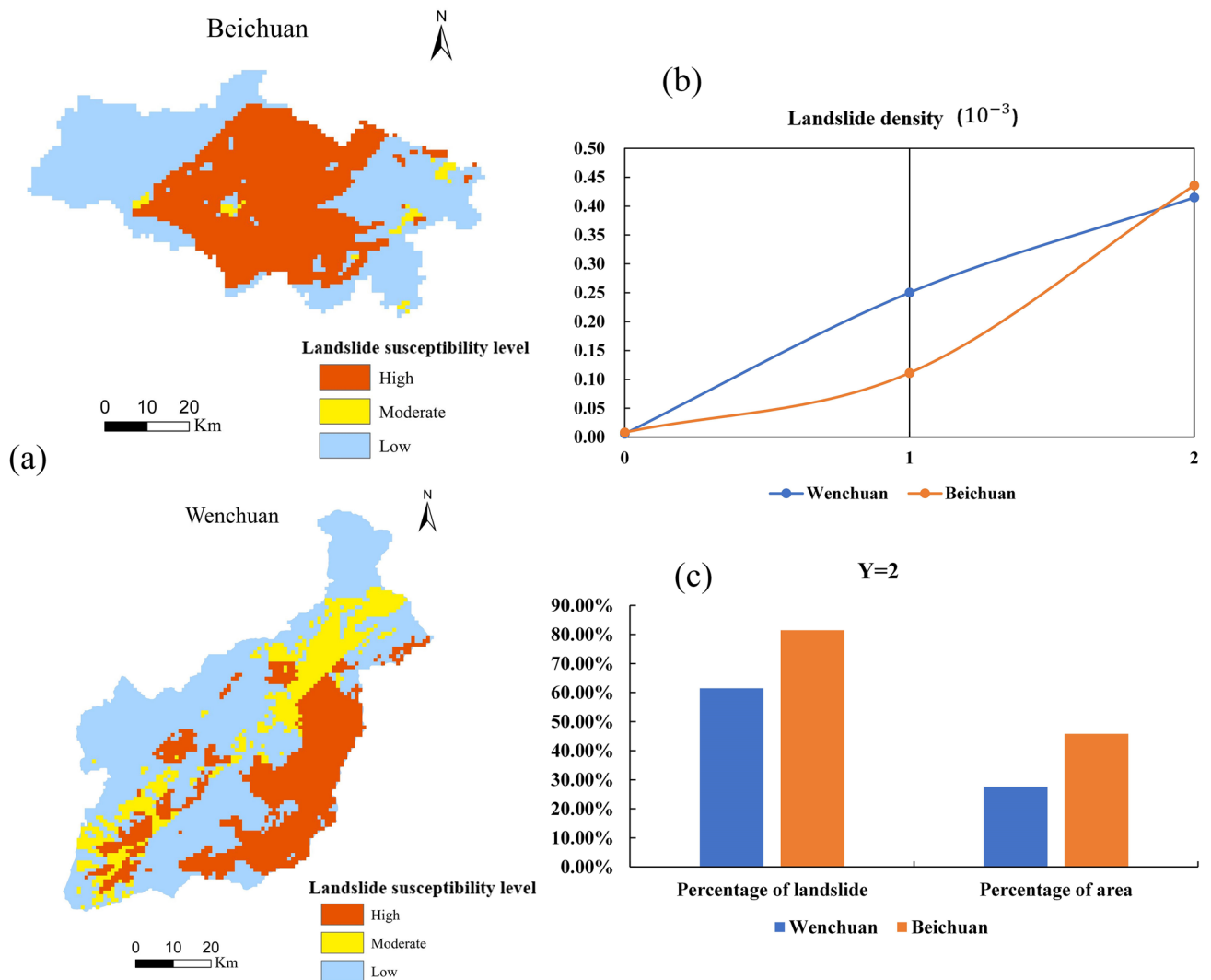


Figure 8. Map of Wenchuan and Beichuan with regional statistical results (Names of software: ArcGIS 10.5 and Office 2019).

density of regional landslides also increases. From the statistical analysis shown in Fig. 8, there are a total of 552 landslides in Wenchuan and Beichuan, and approximately 71% of those landslides are in areas with a predicted landslide susceptibility levels of 2. In Beichuan, approximately 80% of the landslides are located in the Y = 2 area, which is more consistent with the spatial distribution of the actual landslide points.

Discussion

This paper takes 10 extremely earthquake-stricken areas of Wenchuan as examples, constructs a spatial quantitative modelling of landslide of earthquake-stricken area which based on the KDE-MLR. This provides a decision-making basis for the early warning of landslide in earthquake-stricken areas. But it can also be improved from the following aspects:

- (1) The nonlandslide points in this paper are selected at a certain distance away from the disaster points. According to Zuo and Wang⁵⁸, that the positive training set usually uses known data. However, not all the negative points can be treated as true negative samples, which means that selecting of random negative training points creates uncertainty. For the spatial prediction model of landslides that is driven by supervisory data, identifying negative samples will be the focus of research in the future.
- (2) When predicting landslides, this paper uses the KDE algorithm to describe the spatial characteristics of landslides. Kristan et al.⁵⁹ proposed a new online estimation method of the probability density function based on kernel density estimation (KDE). This method maintains and updates the nonparametric model of the observed data, and has great advantages over the selection of bandwidth and kernel function in this paper. Additionally, the advantages of machine learning methods in the spatial prediction of landslides are constantly being shown. Therefore, in future research, this paper will consider to using machine learning methods for prediction.

Conclusions

Taking as examples the landslides in the 10 areas extremely earthquake-stricken areas of the Wenchuan earthquake in 2010, this paper comprehensively considers the spatial location of landslides and other factors after the earthquake and uses the KDE method to quantitatively characterize the spatial characteristics of landslides and improve the intensity description of traditional landslides (the "0–1" occurrence assignment method). Next, the spatial distribution characteristics of landslides after the earthquake are calculated. Then, taking the spatial characteristic of landslides as the dependent variable and each conditioning factor as the independent variable, an MLR model is used to establish a quantitative spatial landslide model. After comparing and analyzing the traditional two-category LR landslides prediction model, the specific conclusions are as follows:

- (1) In the landslide susceptibility prediction model based on KDE-MLR, the absolute value of the logistic regression coefficient of PGA is the largest ($PGA = 0.418$), indicating that this conditioning factor has the greatest impact on the landslides in the study area. The value of $PGA = 0.42$ is basically consistent with these results. NDVI is also has a high value in the KDE-MLR model, however, in the two-category LR model, this conditioning factor has the smallest impact on the prediction results. The predictive variable terrain information entropy is the least important in the KDE-MLR model. The four conditioning factors of lithology, distance to roads, distance to rivers and distance to faults are basically consistent between the two prediction models.
- (2) The accuracy of the KDE-MLR model when predicting the spatial distribution of landslides in the study area is approximately 70.7%, which is similar to the prediction accuracy of the traditional two-category LR landslide model. However, the landslide prediction susceptibility maps show that the KDE-MLR model susceptibility predictions are more consistent with the spatial distribution of actual landslides. Compared to the landslide prediction results obtained from the two-category LR model, the KDE-MLR model highlights the spatial characteristics of landslides in the 10 extremely earthquake-stricken areas and more accurately expresses the spatial distribution characteristics of landslides. This shows that on the basis of combining GIS technology, the KDE-based, MLR method has great advantages for studying the spatial prediction of earthquake-induced landslides over the traditional, two-category LR method. Furthermore, it has great prospects for promotion and application.

Data availability

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Received: 22 March 2022; Accepted: 12 January 2023

Published online: 16 January 2023

References

1. Li, X. *et al.* Spatiotemporal characteristics of earthquake disaster losses in China from 1993 to 2016. *Nat. Hazards* **94**(2), 843–865 (2018).
2. Han, P., Tian, S., Fan, X. & Sheng, X. Statistical analysis and forecasting of the secondary disasters induced by lushan earthquake. *J. Nat. Disasters* **27**(1), 120–126 (2018).
3. Ramirez, M. R. & Peek-Asa, C. L. Epidemiology of traumatic injuries from earthquakes. *Epidemiol. Rev.* **27**(1), 47–55 (2005).
4. Dagdelenler, G., Nefeslioglu, H. A. & Gokceoglu, C. Modification of seed cell sampling strategy for landslide susceptibility mapping: An application from the Eastern part of the Gallipoli Peninsula (Canakkale, Turkey). *Bull. Eng. Geol. Environ.* **75**(2), 575–590 (2016).
5. Tsangaratos, P. & Ilia, I. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Prefecture, Greece. *Landslides* **13**(2), 305–320 (2016).
6. Chen, W., Ding, X., Zhao, R. & Shi, S. Application of frequency ratio and weights of evidence models in landslide susceptibility mapping for the Shangzhou District of Shangluo City, China. *Environ. Earth Sci.* **75**(1), 64 (2016).
7. Stahl, T. *et al.* Earthquake science in resilient societies. *Tectonics* **36**(4), 749–753 (2017).
8. Xu, C. & Xu, X. Comment on "Spatial distribution analysis of landslides triggered by 2008.5.12 Wenchuan Earthquake, China" by Shengwen Qi, Qiang Xu, Hengxing Lan, Bing Zhang, Jianyou Liu [Engineering Geology 116 (2010) 95–108]. *Eng. Geol.* **133**, 40–42 (2012).
9. Junsong, L. & Menglan, W. Sensor-based mountain landslide sensitivity and logistics supply chain management optimization. *Arab. J. Geosci.* **14**, 1612.
10. Pathak, D. Knowledge based landslide susceptibility mapping in the Himalayas. *Geoenviron. Disasters* **3**(1), 8 (2016).

11. Eslaminezhad, S. A., Omarzadeh, D., Eftekhari, M. & Akbari, M. Development of a data-driven model to predict landslide sensitive areas. *Geogr. Tech.* **16**(1), 97–112 (2021).
12. Jennifer, J. J. & Saravanan, S. Artificial neural network and sensitivity analysis in the landslide susceptibility mapping of Idukki district, India. *Geocarto Int.* <https://doi.org/10.1080/10106049.2021.1923831> (2021).
13. Dikshit, A., Sarkar, R., Pradhan, B., Acharya, S. & Alamri, A. M. Spatial landslide risk assessment at phuentsholing, Bhutan. *Geosciences* **10**(4), 131 (2020).
14. Schuster, R. L. & Fleming, R. W. Economic losses and fatalities due to landslides. *Environ. Eng. Geosci.* **23**(1), 11–28 (1986).
15. Pollock, W., Grant, A., Wartman, J. & Abou-Jaoude, G. Multimodal method for landslide risk analysis. *MethodsX* **6**, 827–836 (2019).
16. Guzzetti, F., Reichenbach, P., Ardizzone, E., Cardinali, M. & Galli, M. Estimating the quality of landslide susceptibility models. *Geomorphology* **81**(1), 166–184 (2006).
17. Vakhshoori, V. & Zare, M. Landslide susceptibility mapping by comparing weight of evidence, fuzzy logic, and frequency ratio methods. *Geomat. Nat. Hazards Risk* **7**(5), 1731–1752 (2016).
18. Kritikos, T., Robinson, T. R. & Davies, T. Regional coseismic landslide assessment without historical landslide inventories: A new approach. *J. Geophys. Res. Earth Surf.* **120**(4), 711–729 (2015).
19. Devkota, K. C. *et al.* Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **65**(1), 135–165 (2013).
20. Regmi, A. D. *et al.* Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab. J. Geosci.* **7**(2), 725–742 (2014).
21. Youssef, A. M., Al-Kathery, M. & Pradhan, B. Landslide susceptibility mapping at Al-Hasher area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models. *Geosci. J.* **19**(1), 113–134 (2015).
22. Chen, W., Panahi, M. & Pourghasemi, H. R. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *CATENA* **157**, 310–324 (2017).
23. Pradhan, B. & Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* **25**(6), 747–759 (2010).
24. Xu, C., Dai, F., Xu, X. & Lee, Y. H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **145**, 70–80 (2012).
25. Teshale, A. B. Factors associated with unmet need for family planning in sub-Saharan Africa: A multilevel multinomial logistic regression analysis. *PLoS ONE* **17**, e0263885 (2022).
26. Bai, S. B. *et al.* GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology* **115**(1), 23–31 (2010).
27. Bui, D. T., Lofman, O., Revhaug, I. & Dick, O. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat. Hazards* **59**(3), 1413–1444 (2011).
28. Tobler, W. An alternative formulation for spatial-interaction modeling. *Environ. Plan. A* **15**(5), 693–703 (1983).
29. Zou, F., Zhan, Q. & Zhang, W. Quantifying the impact of human activities on geological hazards in mountainous areas: Evidence from shennongjia, China. *Nat. Hazards* **90**(1), 137–155 (2018).
30. Liu, B., Chen, X., Zhou, Z., Tang, M. & Li, S. Research on disaster resilience of earthquake-stricken areas in Longmenshan fault zone based on GIS. *Environ. Hazards* **19**(1), 50–69 (2020).
31. Liu, Y., Yuan, X., Liang, G., Huang, Y. & Zhang, X. Driving force analysis of the temporal and spatial distribution of flash floods in Sichuan province. *Sustainability* **9**(9), 1527 (2017).
32. Gonzalez, H. R., Mora, C. J. C., Aguirre, G. J. & Novelo, C. D. A. The velocity structure and its relationship to seismic hazard in Tuxtla Gutierrez, Chiapas. *Rev. Mex. Cienc. Geol.* **30**(1), 121–134 (2013).
33. Wu, C. S. Study on the optimization of ecological restoration monitoring in the severely affected areas of Wenchuan Earthquake. *Human Univ. Sci. Technol.* (2013) (in Chinese).
34. Shen, P., Zhang, L. M., Fan, R. L., Zhu, H. & Zhang, S. Declining geohazard activity with vegetation recovery during first ten years after the 2008 Wenchuan earthquake. *Geomorphology* **352**, 106989 (2020).
35. Crozier, M. J. *Landslides: Causes, Consequences and Environment* 252 (Croom Helm Australia Private Limited, 1986).
36. Ayalew, L., Yamagishi, H., Marui, H. & Kanno, T. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Eng. Geol.* **81**(4), 432–445 (2005).
37. Anderson, M., Glade, T. & Crozier, M. *Landslide and Risk* (John Wiley, 2005).
38. Li, Y. Q. Geological Architecture and Formation Mechanism of the transitional zone between Longmenshan Mountains and Sichuan Basin. Doctoral dissertation, China University of Geosciences (Beijing). <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=1018029944.nh&dbname=CDFDTEMP> (2018) (in Chinese).
39. Xu, Q., Liu, C. L., Zhang, B., Liang, N. & Tong, L. Q. Slope instabilities in the severest disaster areas of 5.12 Wenchuan earthquake. *J. Eng. Geol.* **17**(01), 39–49 (2009) (in Chinese).
40. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**(3), 1065–1076 (1962).
41. Chong, X. U. *et al.* Application of logistic regression model on the Wenchuan earthquake triggered landslide hazard mapping and its validation. *Hydrogeol. Eng. Geol.* **40**(3), 98–104 (2013).
42. Dai, F. C., Lee, C. F., Li, J. & Xu, Z. W. Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environ. Earth Sci.* **40**(3), 381–391 (2001).
43. Ayalew, L. & Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **65**(1), 15–31 (2005).
44. Mathew, J., Jha, V. K. & Rawat, G. S. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides* **6**(1), 17–26 (2009).
45. Conoscenti, C. *et al.* Gully erosion susceptibility assessment by means of GIS-based logistic regression: A case of Sicily (Italy). *Geomorphology* **204**, 399–411 (2014).
46. Achu, A. L., Aju, C. D. & Reghunath, R. Spatial modelling of shallow landslide susceptibility: A study from the southern Western Ghats region of Kerala, India. *Ann. GIS Geogr. Inf. Sci.* **26**(2), 113–131 (2020).
47. Luo, Y. D. & Lu, L. Network attack detection based on artificial neural network and genetic algorithm. *Comput. Eng. Des.* <https://doi.org/10.16208/j.issn.ssn1000-7024.2021.09.007> (2021) (in Chinese).
48. Gortmaker, S. L., Hosmer, D. W., & Lemeshow, S. Applied logistic regression. *Contemp. Sociol.* **23**(1), 159 (1994)
49. Long, J. S. *Regression Models for Categorical and Limited Dependent Variables* (Sage, Thousand Oaks, 1997).
50. Chan, H. C., Chang, C. C., Chen, P. A. & Lee, J. T. Using multinomial logistic regression for prediction of soil depth in an area of complex topography in Taiwan. *CATENA* **176**, 419–429 (2019).
51. Story, M. & Congalton, R. G. Accuracy assessment: A user's perspective. *Photogramm. Eng. Remote Sens.* **52**(3), 397–399 (1986).
52. Wright, R. E. Logistic regression. In *Reading and Understanding Multivariate Statistics* (eds Grimm, L. G. & Yarnold, P. R.) 217–244 (American Psychological Association, Washington, 1995).
53. Yan, Y. Q. *et al.* Landslide susceptibility evaluation of Batang fault zone in eastern Qinghai-Tibet Plateau based on weighted weight of evidence model. *Mod. Geol.* <https://doi.org/10.19657/j.geoscience.1000-8527.2020.091> (2021).

54. Liu, J., Li, S. & Chen, T. Landslide susceptibility evaluation based on optimized random forest model. *J. Wuhan Univ. (Information Science Edition)* **43**(7), 1085–1091. <https://doi.org/10.13203/j.hugis20160515> (2018).
55. Liu, F. Z., Wang, L. & Xiao, D. Application of machine learning model in landslide susceptibility evaluation. *Chin. J. Geol. Hazard Prev.* <https://doi.org/10.16031/j.cnki.issn.1003-8035.2021.06-12> (2021) (in Chinese).
56. Liu, R., Li, L. Y., Yang, X., Yang, Y. T. & Yang, M. Landslide susceptibility evaluation based on factor strong correlation analysis method. *Earth Environ.* <https://doi.org/10.14050/j.cnki.1672-9250.2020.48.108> (2021).
57. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
58. Zuo, R. & Wang, Z. Effects of random negative training samples on mineral prospectivity mapping. *Nat. Resour. Res.* **29**(6), 1–13 (2020).
59. Kristan, M., Leonardis, A. & Skocaj, D. Multivariate online kernel density estimation with Gaussian kernels. *Pattern Recogn.* **44**(10–11), 2630–2642 (2011).

Acknowledgements

This work was supported by the Sichuan Science and Technology Program (NO. 2021JDR0108), Opening Fund of Geomathematics Key Laboratory of Sichuan Province (NO. scsxdz2020yb05), Opening Fund of Sichuan Mineral Resources Research Center (NO. SCKCZY2022-YB017) and the General Program of Sichuan Center for Disaster Economy Research (NO. ZHJJ2022-YB002), Chengdu University of Technology Development Funding Program for Young and Middle-aged Key Teachers (NO. 10912-JXGG2020-06251).

Author contributions

X.F. and S.H. wrote the main manuscript text, K.P. performed the material preparation, J.L. performed the data collection, X.F., B.L. and Z.Z. performed the analysis. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023