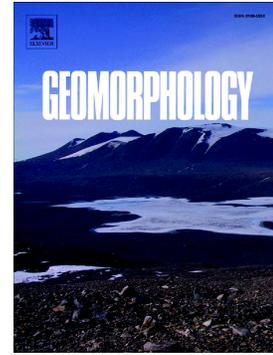


Accepted Manuscript

New method for landslide susceptibility mapping supported by spatial logistic regression and GeoDetector: A case study of Duwen Highway Basin, Sichuan Province, China

Jintao Yang, Chao Song, Yang Yang, Chengdong Xu, Fei Guo, Lei Xie



PII: S0169-555X(18)30384-2
DOI: doi:[10.1016/j.geomorph.2018.09.019](https://doi.org/10.1016/j.geomorph.2018.09.019)
Reference: GEOMOR 6528
To appear in: *Geomorphology*
Received date: 5 March 2018
Revised date: 13 September 2018
Accepted date: 21 September 2018

Please cite this article as: Jintao Yang, Chao Song, Yang Yang, Chengdong Xu, Fei Guo, Lei Xie , New method for landslide susceptibility mapping supported by spatial logistic regression and GeoDetector: A case study of Duwen Highway Basin, Sichuan Province, China. *Geomor* (2018), doi:[10.1016/j.geomorph.2018.09.019](https://doi.org/10.1016/j.geomorph.2018.09.019)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**New method for landslide susceptibility mapping supported by
spatial logistic regression and GeoDetector: A Case Study of Duwen
Highway Basin, Sichuan Province, China**

Jintao Yang^{ab}, Chao Song^{ab}, Yang Yang^{ab*}, Chengdong Xu^c, Fei Guo^d, Lei Xie^e

a. School of Geosciences and Technology, Southwest Petroleum University, Chengdu 610500, China

b. State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu 610500, China

c. State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

d. National Field Observation and Research Station of Landslides in the Three Gorges Reservoir Area of the Yangtze River, China Three Gorges University, Yichang 443002, China

e. Sixth Topographic Surveying Brigade of the National Administration of Surveying, Mapping and Geoinformation, Chengdu 610500, China

+ Jintao Yang and Chao Song contributed equally to this paper

***Corresponding author:**

E-mail: yacoco1981@126.com

Telephone: +8618681279063

Address: School of Geosciences and Technology, Southwest Petroleum University, Chengdu 610500,
China

Abstract Landslides are destructive not only to property and infrastructure but also to people living in landslide-prone regions. Landslide susceptibility mapping (LSM) is critical for preventing and mitigating the negative impacts of landslides. However, many previously proposed LSM modeling techniques included only the attribute information of spatial objects and ignored the spatial structural information of spatial objects, which led to suboptimal LSM. In addition, the selection of condition factors was not objective to such an extent that it may have reduced the reliability of LSM. To address these problems, a new method based on GeoDetector and a spatial logistic regression (SLR) model is proposed. GeoDetector is used to select condition factors based on the spatial distribution of landslides. The SLR model is used to make full use of the structural and attribute information of spatial objects simultaneously in LSM. The GeoDetector-SLR model is validated using data from the Duwen Highway Basin, which includes the epicenter of the May 12 2008 Wenchuan earthquake in southwestern China. Prediction accuracy of the GeoDetector-SLR model is found to be 86.1%, which is an 11.9% improvement over the traditional logistic regression model, indicating an improved and reliable solution for evaluating landslide susceptibility.

Keywords Landslide susceptibility mapping; GeoDetector; spatial logistic regression; spatial autocorrelation

1. Introduction

Landslides are the most common and dangerous type of natural disaster in mountainous regions, and can cause severe damage to people's lives and property (Petley, 2012). In 2014 alone, a total of 8,128 landslides occurred in China, and resulted in nearly 350 deaths and direct economic losses of an

estimated US\$827 million (<http://data.mlr.gov.cn>). Landslide susceptibility mapping (LSM) is a critical tool in disaster prevention and mitigation as it can show potential areas prone to landslides (Dai et al. 2002). However, the accuracy of LSM results is often limited by adverse effects that can increase the uncertainty of LSM and may hinder regional planning decisions. Therefore, it is necessary to improve the accuracy of landslide susceptibility assessments.

In most cases, spatial structure information exists between adjacent mapping units. To perform LSM, there are two main basic mapping units: slope-based and grid-based units (Eeckhaut et al. 2009). Compared to slope-based units, grid-based units are widely applied because they can be easily obtained. For this reason, many researchers have applied grid-based units to assess landslide susceptibility (Bai et al. 2010; Ilija and Tsangaratos 2016; Li et al. 2016). In this case, the properties of adjacent grids are similar because one spatial object (landslide surface) is usually divided into several identical grids. As a result, spatial autocorrelation usually exists in the adjacent grids (Erener and Düzgün 2012). Spatial autocorrelation provides important spatial structure information of geographic phenomena (Wang et al. 2014). Taking full advantage of this information will help reduce the uncertainty of LSM. Although a large number of methods have been applied to LSM (Akgun 2012; Bai et al. 2010; Lee and Choi 2004; Li et al. 2016; Luo and Liu 2018; Pham et al. 2015; Wang and Sassa 2005), a few of these methods have considered this important information (Erener and Düzgün 2010, 2012).

The models of LSM can be broadly divided into three categories: heuristic methods, statistical methods, and physical methods (Luo and Liu 2018). Physical methods require more exact mechanism and process information of landslides (Luo and Liu 2018) and are only suitable for a single landslide

surface or a small study area (Westen et al. 2008). Therefore, statistical methods are more commonly used for quantitative evaluation of landslide susceptibility over a large area. An assumption of independence exists when statistical models are applied; however, the data of LSM always involve spatial autocorrelation rather than being independent (Erener and Düzgün 2012). Under these circumstances, it is difficult to obtain optimal results as some statistical methods are unable to make use of significant spatial autocorrelation information (Erener and Düzgün 2010).

The inability to make full use of spatial structure information restricts the application of spatial data to some extent. Nevertheless, with the development of spatial statistics, there are now many methods with which the effects of spatial autocorrelation can be eliminated, such as data transformation and the spatial autoregression (SAR) model (Wang et al. 2010b). Data transformation reduces dependence by down-sampling to make the data follow a random distribution. However, data transformation results in larger confidence intervals and higher variance (Wang et al. 2010b). SAR makes the residual error tend toward white noise by absorbing the spatial structure information (Wang et al. 2010b), and it has been widely accepted and applied to other fields (Blangiardo and Cameletti 2015; Erener and Düzgün 2010; Lichstein et al. 2002).

In addition to underuse of spatial structure information, lack of an effective method to select factors is another cause of LSM accuracy reduction. Redundant condition factors increase the instability of the model and decrease the predictive accuracy (Jebur et al. 2014). For this reason, many studies have screened for factors that contribute significantly to landslides. For example, the factor analysis method, certainty factor method, and optimization technique were used to exclude redundant factors (Dou et al. 2015, Jebur et al. 2014, Lee and Talib 2005). These methods improve the reliability

of the LSM to a certain extent, but these methods do not account for the spatial pattern characteristics of spatial data, which reduces their accuracy. GeoDetector is a spatial statistics tool that can assesses the relative importance of different factors controlling or contributing to a geographic phenomenon (Luo and Liu 2018). And it has been widely used in other fields for it could account for the spatial patterns of spatial data (Wang et al. 2010a).

There are two common issues with using LSM. One is that spatial structure information is not fully used; the other is that the selection of condition factors is usually insufficiently effective. To resolve these issues, a new model based on GeoDetector and spatial logistic regression (SLR) is proposed in this paper and is referred to as GeoDetector-SLR. GeoDetector-SLR uses Geodetector to select condition factors and apply SLR to modeling (The SLR model is one of the SAR models, which can simultaneously utilize structural information and attribute information of spatial data). The new model is applied to the Duwen Highway Basin in the Longmen Mountain fault zone, China, and the new model is compared with the logistic regression (LR) model, which is unable to incorporate spatial structure information.

2. Study area and data

2.1. Study area

The Duwen Highway Basin is located in the Longmenshan Mountain Range, a section of the Minjiang River watershed, Wenchuan County, Sichuan Province, China (Fig. 1). Its geographic coverage is approximately 30°54'–31°36'N latitude and 103°14'–103°45'E longitude, with an area of 925 km². On May 12, 2008, the MS 8.0 May 12 Wenchuan earthquake occurred in the study area (Yin

et al. 2009), which caused secondary geological disasters such as landslides and debris flows along the Duwen Highway; the area entered an active phase expected to last for 10–20 years (Cui et al. 2008).

The study area is located between the Longmen Mountain System and the Lushan System, in the transition zone from the Qinghai–Tibet Plateau to the Sichuan Basin and the marginal mountainous terrain. This area is a typical alpine valley area, with elevation of 734–5,304 m. The overall topography of the area is slanting, high in the northwest and low in the southeast. Hillsides in this area are usually steep and with an average slope angle of 36.4° . The study area has a continental monsoon climate with an average annual rainfall of 529–1,332 mm. The Minjiang River is the main river in this area.

Duwen Highway has three sections from north to south: the Duwen–Mianjiu, Mianjiu–Yingxiu and Yingxiu–Dujiangyan sections. Phyllite and quartz sandstones of the Devonian Hanlizhai Group and the Proterozoic Huangshuihe Group are located in the Duwen–Mianjiu section, intrusive rocks such as biotite granite and plagiogranite of the Proterozoic Jinning–Jinjiang period characterize the Mianjiu–Yingxiu section, and interbedded layers of cuttings, quartz sandstone, and mudstone in the Triassic Xujiahe Formation are found in the Yingxiu–Dujiangyan section. The bedrock exposed along the Duwen Highway mainly consists of granite, diorite, limestone, phyllite, sandstone, and granite rocks (Gui-Sheng et al. 2016; Zhang et al. 2015).

Fault zones in this area are relatively well developed, making it vulnerable to geological disasters (Zhuang et al. 2010). There are two large thrust faults along the Longmenshan thrust belt at the eastern margin of the Tibetan Plateau; one is a 240-km-long surface rupture along the Beichuan fault, and the other is a 72-km-long surface rupture along the Pengguan fault. These faults were ruptured by the May

12 Wenchuan earthquake (Xu et al. 2012). Additionally, there is an NW-striking left-lateral reverse rupture about 7 km in length between the Beichuan and Pengguan faults (Xu et al. 2009).

2.2. Data

Based on remote sensing images interpretation and field geological hazards survey, 4,841 landslides were obtained (Fig. 1). These images were observed by the Pleiades satellite on December 7, 2014, including multi-spectral images at 2m resolution and panchromatic images at 0.5m resolution. A total of 15 explanatory variables were used in this paper, and all of them are the monitoring results of the national geographic conditions survey. They were provided by the Sichuan Province Bureau of Surveying, Mapping and Geoinformation, China. These monitoring results all passed stringent quality checks by the geographical experts and geologists appointed by Sichuan Bureau of Surveying, Mapping and Geoinformation, China. (http://www.mlr.gov.cn/xwdt/chxw/201510/t20151019_1384486.htm). Therefore, the quality of these data is reliable, and has been verified in several published papers and an unpublished government report (Meng et al. 2016; Wang et al. 2017).

The landslide factors were divided into geological/topographic (fault zone, seismic intensity, rock mass, geologic time, elevation, roughness, slope, aspect, and river), ecological (vegetation, precipitation, land stress, and soil erosion), and socio-economic (roads, residences, and hydropower stations). Details of the data, such as type and structure, are shown in Table 1. The specific properties of these factors can be found in Meng et al. 2016; Wang et al. 2017. Seven explanatory factors (Fig. 2) were selected as potential influencing factor by GeoDetector (factor selection is described in Section

3). Two discrete factors were included in the selected factors, and the classification of the two discrete factors was done by the data provider.

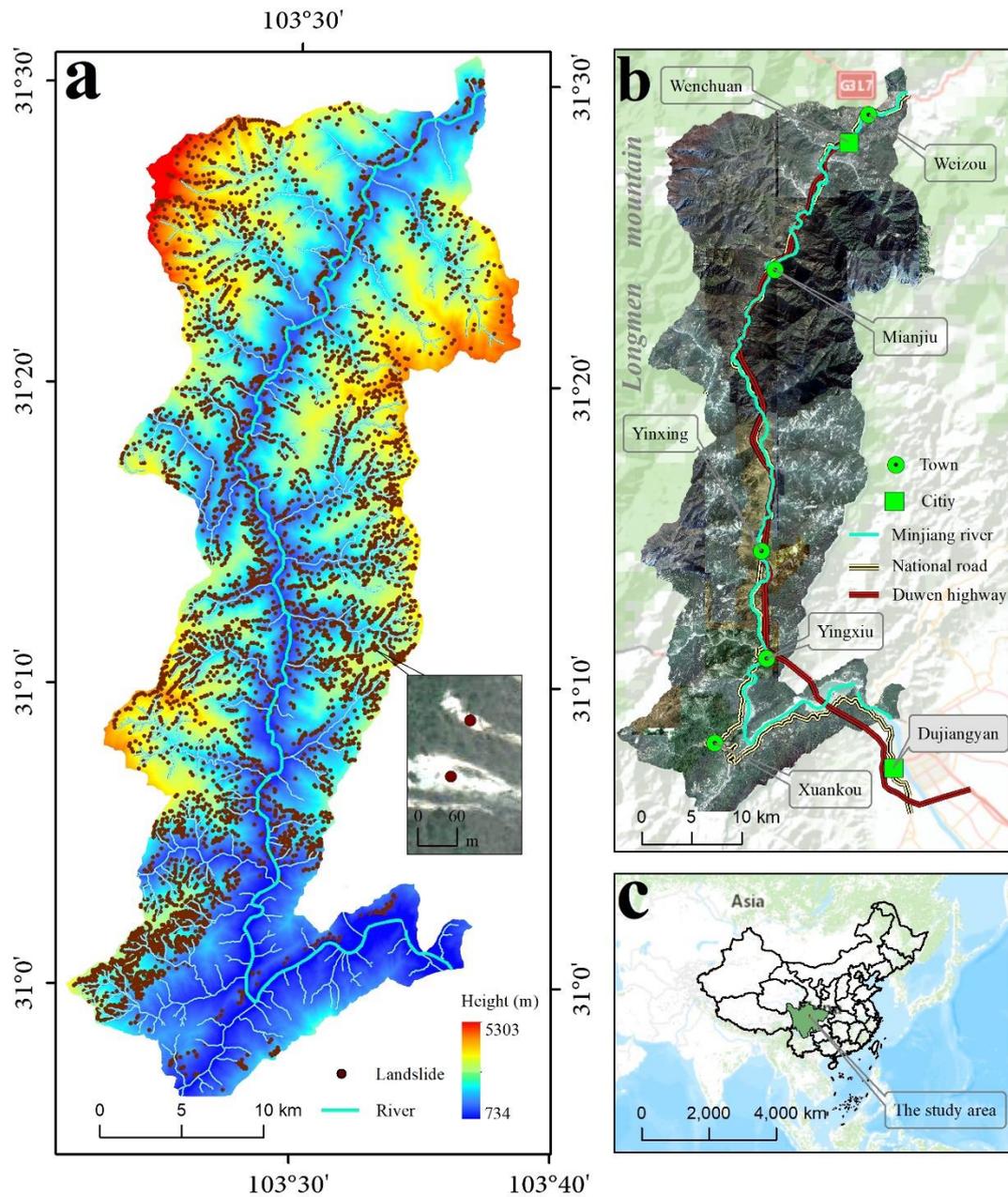


Fig. 1 Landslide inventory map of the Duwen Highway Basin. (a) landslide inventory, (b) remote sensing image map, (c) the location of study area.

Principal conditional factors used in this study are shown in Fig. 2 (a-g). In it, the factor ‘rock mass’ (Fig. 2a) has five levels: the harder rock group, hard rock group, soft rock group, softer rock group, and loose rock group, denoted respectively as E, D, C, B, and A (Meng et al. 2016; Wang et al. 2017). The classification criteria are based on the "Engineering Rock Mass Classification Standard", which is one of China's national standards (GB50218-2014). (http://www.mohurd.gov.cn/wjfb/201508/t20150829_224347.html). Fig. 2b is the slope angle map. After the 12 May 2008 Wenchuan earthquake, the ecological environment has undergone major changes (Zhang et al. 2015); therefore, seismic intensity is regarded as the condition factor in this study. The seismic intensity of the 2008 Wenchuan earthquake is shown in Fig. 2d, the four seismic zones (VIII-XI) are the seismic intensities distribution of the 2008 Wenchuan earthquake and is expressed as M_s 8, M_s 9, M_s 10, and M_s 11, respectively. ‘seismic intensity’ was produced by the China Earthquake Administration (CEA 2008). Fig. 2c is the residential area buffer map. Roughness of terrain is shown in Fig. 2e, where larger values denote rougher surfaces. Fig. 2f is the road buffer map. Finally, Fig. 2g represents elevations of the area.

Table 1 The names, structures, types, descriptions, and classification of environmental variables.

NDVI is the abbreviation of " Normalized Difference Vegetation Index ". The data is provided by the Sichuan Province Bureau of Surveying, Mapping and Geoinformation, China.

Variables	Name	Data structure	Variable type	Data description	Class
Y	landslide point	point	binary	landslide occurred or not	landslide
X1	fault zone	line	continuous	distance to line	geological
X2	seismic intensity	polygon	discrete	2008 Wenchuan earthquake	geological
X3	rock mass	polygon	discrete	the hardness of the rock and soil	geological

X4	geologic time	polygon	discrete	different times of Rocks and stratum	geological
X5	elevation	raster	continuous	height above sea level	topographic
X6	roughness	raster	continuous	the roughness index of terrain	topographic
X7	slope	raster	continuous	extracted from DEM	topographic
X8	aspect	raster	discrete	extracted from DEM	topographic
X9	river	line	continuous	distance to river	topographic
X10	precipitation	polygon	discrete	precipitation classification	ecology
X11	soil erosion	raster	discrete	soil erosion degree	ecology
X12	NDVI	raster	discrete	the vegetation of surface	ecology
X13	road	line	continuous	distance to road	ecology
X14	hydropower station	point	continuous	distance to point	ecology
X15	settlement	point	continuous	distance to point	ecology

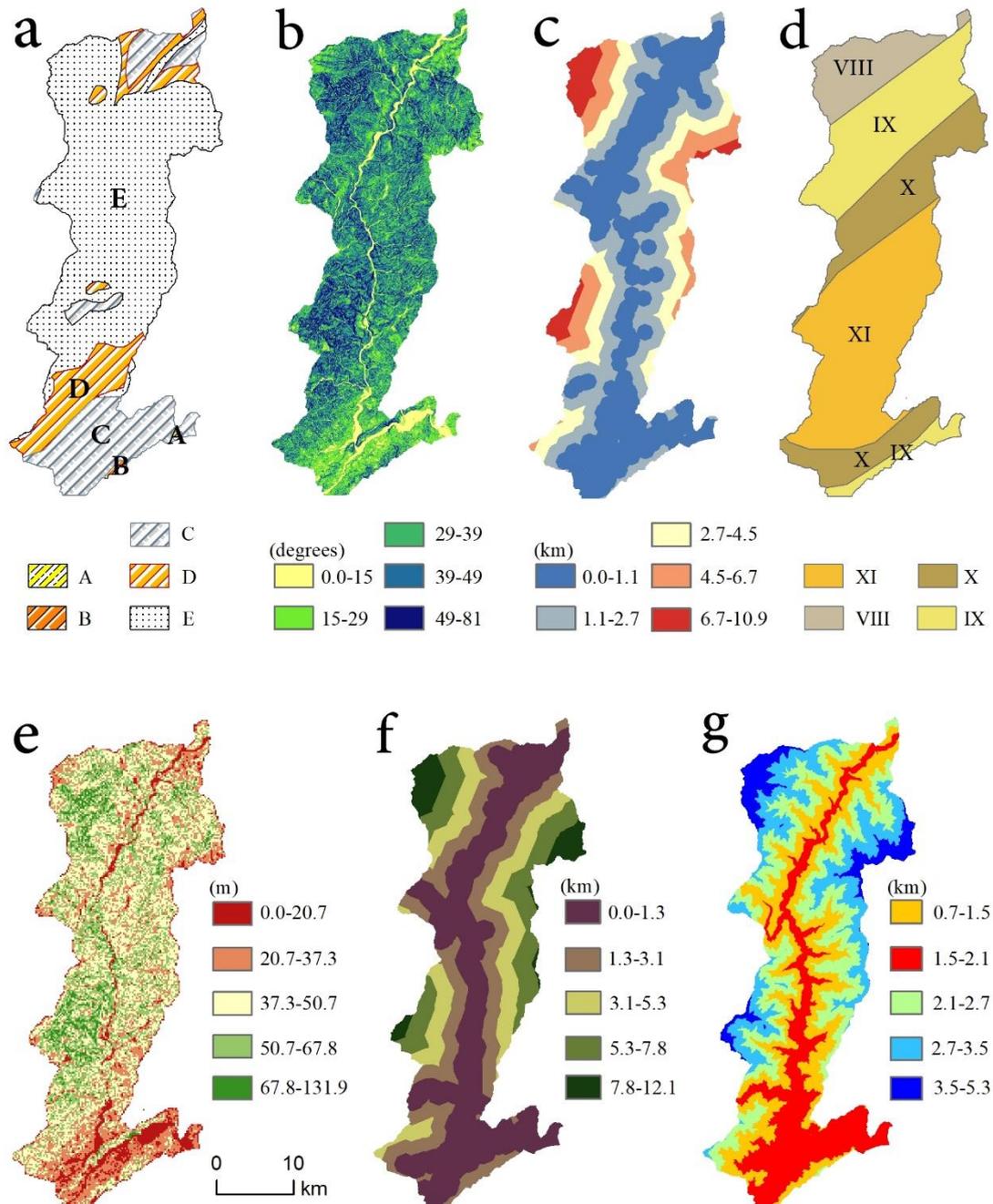


Fig. 2 Thematic maps of condition factors. (a) Rock mass. The harder rock group, hard rock group, soft rock group, softer rock group, and loose rock group of rock mass, denoted respectively as E, D, C, B, and A. (b) slope, (c) distance from settlements, (d) seismic intensity. The four seismic zones (VIII-XI) are the seismic intensities distribution of the 2008 Wenchuan earthquake and is expressed as M_s 8, M_s

9, M_s 10, and M_s 11, respectively. (e) terrain roughness, (f) distance to highway, (g) elevation. The data is provided by the Sichuan Province Bureau of Surveying, Mapping and Geoinformation, China.

3. Methods

The process of LSM is divided into three main phases. The first phase including the selection of condition factors. The second phase is model incorporation, in this phase the condition factors selected by GeoDetector and the landslide inventory layer was divided into basic mapping units, and their attribute tables were extracted to form test data sets and training data sets. Modeling in the third phase, which include model training, mode verification and model comparison. Fig. 3 illustrates the research flow of this article.

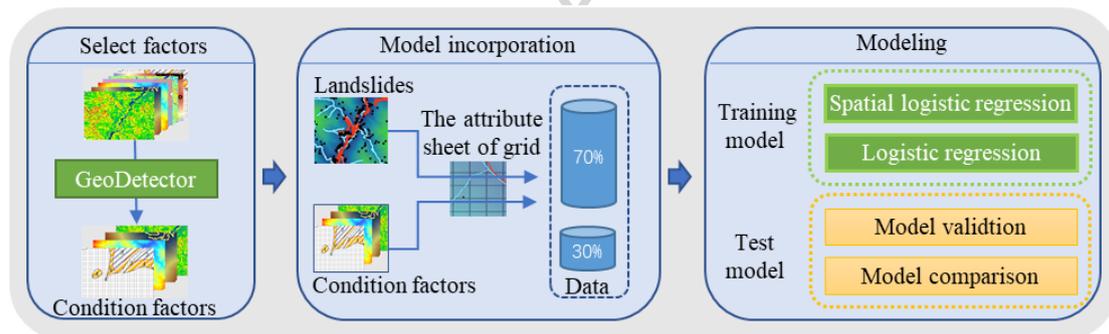


Fig. 3 Research flow. Research flow includes three major phases: Factor selection in the first phase, Model incorporation in the second phase, and modeling in the third phase.

3.1. GeoDetector

An evaluation index system of condition factors needs to be established when undertaking LSM. However, a universal framework for selection of condition factors is lacking (Luo and Liu 2018). The GeoDetector is utilized for quantitatively evaluating one potential factor relative to a spatial phenomenon and can be used for factor selection (Wang et al. 2010a). GeoDetector was first applied

for detecting potential disease factors (Wang et al. 2010b). Because GeoDetector makes very few assumptions about the data, the application of GeoDetector has been extended to remote sensing and geoscience (Luo et al. 2016). The GeoDetector software is freely available from <http://www.GeoDetector.org/>.

The core hypothesis of GeoDetector is that if an independent variable has an important influence on a dependent variable, the spatial distributions of the independent variable and the dependent variable should be similar (Wang and Hu 2012). That similarity can be measured based on the ratio of local variance to global variance (Wang et al. 2016). As such, the study area was divided into basic mapping units for LSM, and the landslide rate R (Fig. 4a) of each unit was taken as the y -variable by GeoDetector. Based on the type of landslide data, the value of R can be calculated in two ways. If the historical landslide representation is polygon-based, R is the ratio of the landslide area to the area of the basic mapping unit. If the historical landslide representation is point-based (center of the landslide body), R is expressed by the count of landslide points in the mapping unit. To calculate the local variance of each region of R in the x -layer (Fig. 4b), GeoDetector specifies that the x -variable entered must be partitioned data (Wang et al. 2016). Therefore, it is necessary to reclassify continuous variables such as the elevation and slope. The principle of GeoDetector is as follows:

$$q = 1 - \frac{1}{N\sigma^2} \sum_{i=1}^m N_i \sigma_i^2 \quad q \in [0,1] \quad (1)$$

where m is the stratum count in the x -layer, N is the number of mapping units in the study area, σ_i^2 is the variance of R in the i^{th} stratum, and σ^2 is the variance of R in the entire area. Large values of the index q indicate a large contribution of the x -layer to landslide occurrence.

Assuming the x -layer (Fig. 4b) is a potential influencing factor of y -layer (Fig. 4a), such as lithology or rock mass, it can be divided into three strata. This step is shown in Fig. 4b, where the x -layer is divided into regions labeled I, II, and III, which show its attributes. The local and global variances were calculated by overlaying the analysis of the y -layer and the x -layer (Fig. 4c) to detect whether the x -layer contributes to the spatial distribution of landslides. If x is related to R , there will be a similar spatial pattern between R and x . In other words, the value of R in each stratum of x will be more homogeneous and have lower variance. Ideally, x would completely explain the spatial pattern of R , where the variance of R is close to 0 and q is close to 1. In a case, where the partition of x is completely unrelated to the spatial pattern of R , the variance of R in each partition of the x -layer is the same, and the value of q is zero.

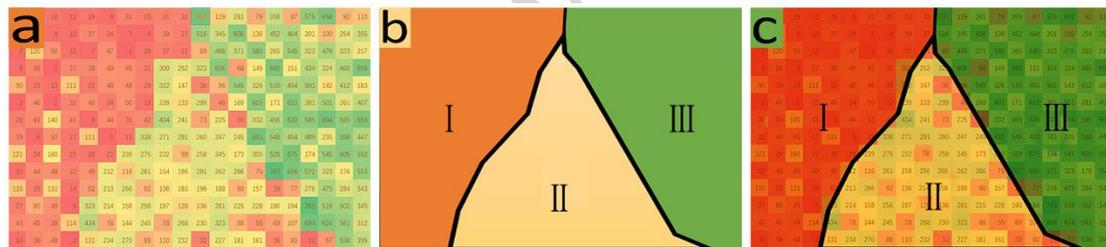


Fig. 4 Principles of GeoDetector. Panel a is the y -layer showing the landslide rate (R) for each unit, b is the x -layer as a potential condition factor, and c shows the spatial distribution similarity between the y -layer and x -layer obtained by overlapping these layers.

3.2. Implementation of GeoDetector

As mentioned above, GeoDetector requires the potential influence factor x to be partition data. Therefore, continuous variables were reclassified into five classes using the natural break method (Fig. 2.). The study area was divided into regular grids to calculate the variable R (Fig. 4a), and the

frequency of landslides in the grid was taken as the R value. The R -layer and x -layer were then overlapped, and the attributes of each layer were joined. Then, the attribute table of each mapping unit was exported as the input data to GeoDetector. The relative contributions of all potential factors calculated by GeoDetector are shown in Fig. 5. The first seven variables with q -statistic index greater than 0.05 were selected as the independent variables in the SLR model.

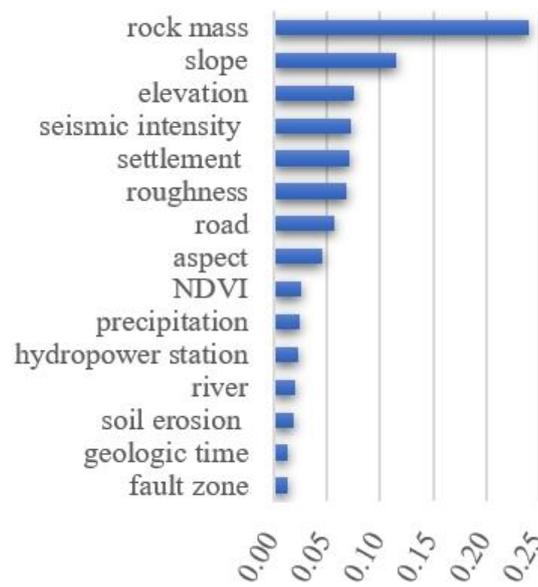


Fig. 5 The q -statistic indices calculated by GeoDetector. Graphical representation of the relative contributions of potential factors to landslide formation. NDVI is the abbreviation of Normalized Difference Vegetation Index.

3.3. Spatial logistic regression

The main objective of the LR model is to determine the probability of an event occurring by constructing a regression relationship between a binary variable and multiple independent variables (Bai et al. 2010). LR models are ideally suited to the types of data used in landslide susceptibility

assessments because they can contain both discrete and continuous variables (Akgun 2012; Lee and Sambath 2006). The principle equations governing the LR model are as follows:

$$Y = \ln \frac{P(x)}{1-P(x)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \varepsilon \quad (i = 1, 2, \dots, n) \quad (2)$$

$$P = \frac{e^Y}{1+e^Y} \quad (3)$$

where X_1, X_2, \dots, X_i are independent variables and $\beta_0, \beta_1, \dots, \beta_i$ are the regression coefficients to be determined. The function $P(x)$ represents the probability of a landslide occurring (x is a condition factor). In the application of spatial data, the Y variable may be not only related to the X_i variables but also related to itself (the surrounding landslides). Thus, there is spatial autocorrelation in the data (Erener and Düzgün 2012). In that case, it is necessary to include this important spatial structure information in the LR model (Wang et al. 2010b), and Eq. (1) can be modified as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \rho WY + \varepsilon. \quad (4)$$

Eq. (4) is the spatial auto-regression (SAR) model (Anselin 1988; Erener and Düzgün 2012; Lichstein et al. 2002), which is robust with respect to spatial autocorrelation data because both spatial attribute information and spatial structure information can be accounted for simultaneously. In Eq. (4), ρ is the spatial autocorrelation parameter, ε is the error term obeying a Gaussian distribution, and W is the spatial weight matrix of $n \times n$ dimensions where n is the total number of samples. This matrix defines the adjacency relationship between landslide mapping units. The factor ρWY is the spatial structure effect caused by spatial autocorrelation (Wang et al. 2010b).

If there is no spatial autocorrelation in y -variables, then $\rho = 0$, and Eq. (4) becomes the same general regression model as Eq. (1). Unlike the non-spatial model, the model ρWY contains the spatial structure information of the spatial object and is absorbed as a latent variable in the SAR model. This

step prevents the non-spatial model from incorporating it into the residual to derive a biased estimate.

Integration of Eq. (4) with Eq. (2) results in the SLR model, and W of the spatial structure effect ρWY

as follows:

$$W = \begin{bmatrix} 0 & f(d_{12}) & \cdots & f(d_{1j}) \\ f(d_{21}) & 0 & \cdots & f(d_{2j}) \\ \vdots & \vdots & \ddots & \vdots \\ f(d_{i1}) & f(d_{i2}) & \cdots & 0 \end{bmatrix}, \quad (5)$$

$$f(d_{ij}) = \frac{d_{ij}}{\sum_1^j d_{ij}}, \quad (6)$$

$$L = y \ln \frac{\exp(\alpha + X\beta + \rho W y)}{1 + \exp(\alpha + X\beta + \rho W y)} - (1 - y) \ln(1 + \exp(\alpha + X\beta + \rho W y)). \quad (7)$$

Eq. (5) is the weight matrix, and $f(d_{ij})$ is an inverse distance weighting function whose expression is Eq.

(6). d_{ij} represent the distance between the i -th and j -th mapping units.

Estimation of parameters such as traditional least-squares may result in deviations and inconsistencies, issues which are usually resolved with the maximum likelihood function shown in Eq.

(7). To reduce overhead time during calculations, an integrated nested Laplacian approximation was used to solve the model (Blangiardo and Cameletti 2015).

3.4. Implementation of SLR model

The study area of this work is 925 km² in extent. After a compromise between computing efficiency and spatial resolution was calculated, a 200 × 200 m grid was chosen as the basic mapping unit. The center of the landslide scar was taken as the mark of a landslide in this study. And assign an approximation average elevation of the landslide body to the central point. If there is a landslide point in the grid, the value of Y is 1; if the converse is true, then Y is 0. The eight x -layers and the y -layers were overlapped, and the attribute table was exported for use as input in the SLR model. Prior to

entering the data, the data were divided into 30% and 70% as testing data and training data, respectively.

To compare the advantages and disadvantages of SLR and LR, both were developed for the model using the freely available statistical software package R (not to be confused with the landslide rate, R). The INLA package (<http://www.r-inla.org/>) in R was used to infer and solve the model. The expression of the SLR model is as follows:

$$Y = (-0.046 \times \text{settlement}) + (0.393 \times \text{slope}) + (-0.012 \times \text{ELEVATION}) + (0.101 \times \text{road}) + (0.418 \times \text{roughness}) + (-1.299 \times \text{seismic intensity VIII}) + (-0.878 \times \text{seismic intensity IV}) + (-0.857 \times \text{seismic intensity V}) + (-0.771 \times \text{seismic intensity VI}) + (0.533 \times \text{lithology E}) + (-0.599 \times \text{lithology D}) + (-1.764 \times \text{lithology C}) + (-1.488 \times \text{lithology B}) + (-1.404 \times \text{lithology A}). \quad (8)$$

4. Result

4.1. Verification and comparison

Verification and comparison of the model included the following three aspects: goodness of fit, complexity of the model, and predictive accuracy. The deviance information criterion (DIC) is a comprehensive index used to quantify goodness of fit and evaluate the complexity of a model (Spiegelhalter et al. 2002). Lower DIC values indicate higher reliability of the model. The receiver operator characteristics (ROC) curve (Akgün and Bulut 2007; Yesilnacar and Topal 2005) was chosen to express the predictive capabilities of the models. However, the ROC curve (Fig. 6) cannot fully represent the details of the predictive ability of the model. Therefore, a confusion matrix was used to

quantitatively evaluate the accuracy of 0-value predictions and 1-value predictions, and the overall predictive accuracy. In this study, the DIC and the confusion matrix were used for comparing the model degree of fit and model prediction accuracy, respectively (Table 2). The area under the ROC curve (AUC) (Chung and Fabbri 2003) and the confidence of the AUC values are shown in Table 2. Compared with the prediction accuracy of traditional spatial LR, that of SLR is significantly increased by 11.9% (Table 3). The reasons for this are detailed in the discussion section.

Table 2 Comparison of logistic regression (LR) and spatial logistic regression (SLR) in terms of model verification results.

Method	Deviance information criterion	Standard Error	Area Under the ROC Curve	Confidence interval (95%)
LR	19841.46	0.08	0.79	(0.78,0.81)
SLR	17244.78	0.04	0.93	(0.92,0.94)

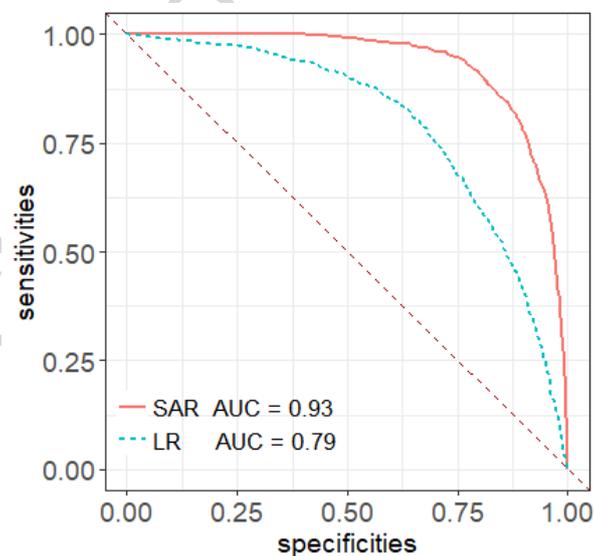


Fig. 6 The receiver operator characteristics (ROC) curve of logistic regression model (LR) and spatial logistic regression model (SLR). AUC is the acronym of area under the ROC curve.

Table 3 The confusion matrix of the logistic regression model (LR) and spatial logistic regression model (SLR).

Method	Landslide occurred	Prediction		Percent	Accuracy
		Yes	No		
LR	Yes	486	235	67.4%	74.2%
	No	1553	4662	75.0%	
SLR	Yes	609	849	84.5%	86.1%
	No	112	5366	86.3%	

The spatial structure effect was taken as a latent variable in the SLR model as Equation (4) shows, and it was extracted by the SLR model (Fig. 7). A larger spatial structure effect indicates a stronger spatial effect, and thus greater deviation of the LR model results.

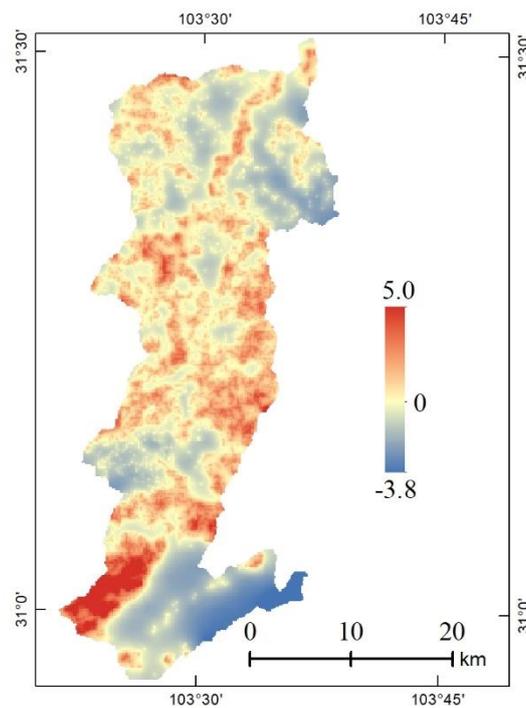


Fig. 7 Map of the study area illustrating the range of the spatial structure effect. Note the high-value areas of spatial autocorrelation in the south.

4.2. Landslide susceptibility mapping

The probability of a landslide occurring was classified using the natural break method to perform LSM (Fig. 8). The susceptibility map output from the LR model is shown in Fig. 8a, and the output of the SLR model is shown in Fig. 8b. The actual landslide point density map was used as a reference to represent real landslides (Fig. 8c). Compared with the actual landslide distribution, the SLR results are highly consistent. However, the most landslide prone areas of the LR's output shows large deviation from the actual landslide distribution in areas of high spatial effect values. The outputs were then combined with the spatial structure effect and the output of the two models for further analysis.

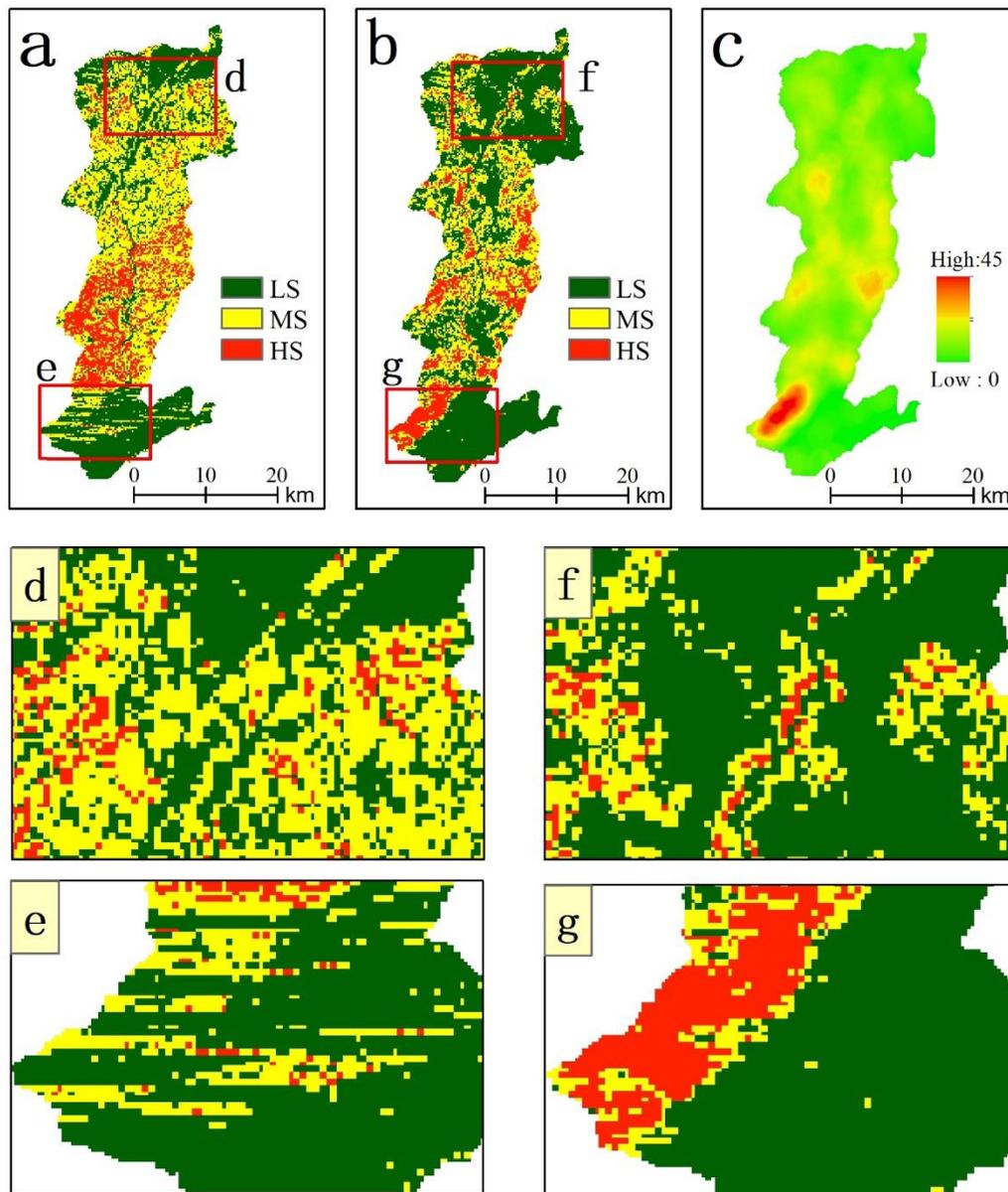


Fig. 8 Landslide susceptibility map. Panel (a) is the landslide susceptibility map produced by the LR model; (b) is the landslide susceptibility map produced by SLR; (c) is the density map of actual landslide spatial points. LS, MS, and HS respectively indicate three landslide susceptibility levels of low, moderate, and high.

5. Discussion

A new model for LSM based on SLR and the use of Geodetector has been proposed in this paper. The predictive accuracy of LSM is improved by making full use of spatial data and objective selection of condition factors. The new model was applied to the Duwen Highway Basin, China, and the results of the LSM show significant improvement compared with the LR model, especially in areas with strong spatial effects. In theory, the SLR model accounted for attribute information and spatial structural information at the same time, which avoids the defect of insufficient information usage. In practice, a more reliable map of landslide susceptibility was acquired, and could be used to provide more accurate decision-making information for landslide risk management.

Spatial structure information had different effects on the results in different regions. In the southwestern part of the study area (Fig. 8g), landslides occurred 45 times per square kilometer. This was the region most prone to landslides and was also the area with the strongest spatial autocorrelation effect (red part of Fig. 8g and c). This strong spatial autocorrelation effect means that a deviation that cannot be neglected in this region occurred in the traditional LR model. Because of the side effects of spatial autocorrelation, the area of high landslide occurrence in Fig. 8e was incorrectly predicted to be an area with low landslide occurrence, which would have a severe impact on decision-making concerning landslide risk management. In the northern part of the study area, the spatial distribution of landslides was relatively weak (Fig. 8d), which corresponds to the lower spatial effect value in Fig. 7. The LR result in this area was consistent with the actual landslide distribution, and there was no major deviation like that in Fig. 8e. Thus, the LR model has improved the accuracy in Fig. 8f compared with Fig. 8e, and the traditional LR model is therefore more reliable when there is a weak spatial effect.

The SLR model is robust with spatial autocorrelation effects. The map produced by the SLR model shows good consistency with the spatial pattern of actual landslide points in the region with the strongest spatial effects (Fig. 8g). This finding substantiates the robustness of the SLR model when using SLM data with a strong spatial autocorrelation effect. In addition, the spatial details of Fig. 8f were accurately identified by the SLR model with the relatively weak spatial effect region, which shows that the SLR model is equally applicable in regions with both high and low spatial effects; therefore, it can serve as a general solution for most LSM scenarios. Although the SLR model was more accurate than the traditional LR model in this study (Table 3), the improvement in prediction accuracy does not mean that the results of traditional LR are entirely unreliable. This result only reflects the strong spatial effect of landslides in the study area, and the results of the LR model, which do not account for spatial structure information, have a larger deviation.

Why is the prediction accuracy of the SLR model greatly improved compared with the LR model? According to the first law of geography (Tobler 1970), adjacent landslide areas have similar geographical environments, with the presence of landslides themselves indicating greater potential for landslides in the surrounding area. In this paper, the spatial autocorrelation effect of high-value areas (red region in Fig. 7) and high-value parts of landslide-intensive areas (red region of Fig. 8c) are spatially consistent, which indicates that spatial autocorrelation can be an important indicator in landslide-prone regions. However, traditional LR only considers the attribute information of the spatial data and ignores the influence of the spatial patterns of landslides themselves, which caused traditional LR models to incorporate autocorrelation information into residuals. This issue leads to large deviations in traditional LR model regions with strong spatial effects.

Removing redundant information and making full use of potential information are both important for LSM to improve accuracy. LSM models based on statistical theory will have unstable results if they contain redundant variables or do not contain significant variables (Jebur et al. 2014). Many studies therefore excluded insignificant condition factors in LSM, whereas very few studies have considered how to include the potential information implicated in the condition factors. This may be related to data unavailability, but if the data have important potential information that has not been included in the model, it will result in information waste. The advantage of the methods described in this paper is that redundant variables are eliminated by using GeoDetector, and the potential information is further extracted using the SLR model.

Compared with other models, spatial models are rarely used in landslide risk assessment. Perhaps a reason behind this is that during evaluation of landslide susceptibility, the Y-variable is binary, which means its spatial effect is not easy to extract, and increases the complexity of the model. Another reason may be that the effect is not strong enough in the study area such that some LSM methods have not fully accounted for the impact of spatial autocorrelation. Some works have compared spatial and traditional models (Erener and Düzgün 2010), the results of these studies are consistent with those of this article. However, if variable screening models are integrated, the advantages will be more pronounced.

Spatial autocorrelation is always accompanied by spatial heterogeneity. Consequently, the sensitivity of the results to the stratification of factors varies from region to region. Spatial heterogeneity affects the weight of condition factors, which caused the dominant factor in landslides vary from stratification to stratification (Chalkias et al. 2014; Wang et al. 2016). Thus, to obtain

optimal results, the study area should be divided into several zones based on a certain condition factor and modeling in each zone. Relative to a regional scale, the study area for this article was small; thus, the sample size may be insufficient if the study area is divided into several zones. Therefore, stratification heterogeneity was not considered for this area. Nevertheless, as supplement, we included three partitioning condition factors as dummy variables in the LSM model. That means in addition to the stratification heterogeneity of continuous factors, the stratification heterogeneity of the categorical factors has been taken into account. Thus, the effect of stratification heterogeneity was reduced.

Although spatial modeling has demonstrated improvements in goodness-of-fit and prediction accuracy, some limitations still exist in this study. First, the complexity of the SLR model means that its computational efficiency is much lower than that of the LR model (the time cost of the SLR model is about 1,200 fold more than that of the LR model), which results in the need to find a compromise between resolution and computational efficiency for large study areas such as the one presented here (a 200 × 200 meter grid is relatively coarse resolution). Secondly, the year 2014 was six years after the 2008 Wenchuan earthquake. Although the destructive effect of the event has weakened year by year (Zhang et al. 2015), the impact of it on the distribution of landslides in the study area still cannot be neglected. Nevertheless, the seismic intensity of the earthquake had weaker control than the factors of rock mass, slope, and elevation, which indicates that the earthquake six years prior was less influential to landslide occurrences than the topographic and ecological condition factors (Fig. 5). Regardless, if more seismic impact data were included, the results would have been more convincing.

Both GeoDetector and SLR can include both the spatial structural information and attribute information of spatial data. Therefore, for future landslide susceptibility modeling, a combination of

GeoDetector and other LSM models may be considered. Besides, the spatial effect may be considered as a potential variable for other LSM models so that more models can make full use of the structural information of spatial objects.

6. Conclusion

This study aimed to improve the reliability of LSM, and a new model based on GeoDetector and SLR was built. The new method solves two common problems in LSM: the selection of condition factors being insufficiently objective, and the information of spatial objects being underused. The new model provides a general solution that makes full use of spatial structure information and accurately selects condition factors. Improvements made with this new model is expected to significantly enhance the reliability of landslide susceptibility maps.

Acknowledgments

We would like to thank the two anonymous reviewers for their constructive and professional reviews of earlier versions of the paper. Moreover, we are grateful to Prof. JinFeng Wang, who provided the freeware GeoDetector, and the contributors to the open source R package INLA. This work was supported by the Open Fund of the State Key Laboratory of Geohazard Prevention and Geoenvironment Protection (No. KLGSI2016-03); the Technology Project of the Sichuan Bureau of Surveying, Mapping and Geoinformation (No. J2017ZC05); the Science and Technology Strategy School Cooperation Projects of the Nanchong City Science and Technology Bureau (No. NC17SY4016); and the National Natural Science Foundation of China (No. 41701448).

References

- Akgun Aykut (2012) A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. *Landslides* 9:93-106.
<https://doi.org/10.1007/s10346-011-0283-7>
- Akgün A, Bulut F (2007) GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region. *Environmental Geology* 51:1377-1387.
<https://doi.org/10.1007/s00254-006-0435-6>
- Bai SB, Wang J, Lü GN, Zhou PG, Hou SS, Xu SN (2010) GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology* 115:23-31.
<https://doi.org/10.1016/j.geomorph.2009.09.025>
- Blangiardo M, Cameletti M (2015) Spatial and spatio-temporal Bayesian models with R-INLA. 4:105-118. Wiley. UK <https://doi.org/10.1002/9781118950203>
- Chalkias C, Kalogirou S, Ferentinou M (2014) Landslide susceptibility, Peloponnese Peninsula in South Greece. *Journal of Maps* 10:211-222
<https://doi.org/10.1080/17445647.2014.884022>
- Cui P, Wei F, He S (2008) Mountain disasters induced by the earthquake of May 12 in Wenchuan and the disasters mitigation. *Journal of Mountain Science* 26:280-282. (In Chinese)
- Dai FC, Lee CF, Ngai YY (2002) Landslide risk assessment and management: an

overview. *Engineering Geology* 64:65-87.

[https://doi.org/10.1016/S0013-7952\(01\)00093-X](https://doi.org/10.1016/S0013-7952(01)00093-X)

Dou J, Tien BD, Yunus AP, Jia K, Song X, Revhaug I, Xia H and Zhu Z (2015)

Optimization of causative factors for landslide susceptibility evaluation using remote sensing and gis data in parts of niigata, japan. *Plos One* 10: e0133262.

<https://doi.org/10.1371/journal.pone.0133262>

Eeckhaut M V D, Reichenbach P, Guzzetti F, Rossi M (2009). Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium. *Natural Hazards & Earth System Sciences* 9:507-521. <https://doi.org/10.5194/nhess-9-507-2009>

Erener A, Düzgün HSB (2010) Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides* 7:55-68.

<https://doi.org/10.1007/s10346-009-0188-x>

Erener A, Düzgün HSB (2012) Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? *Environmental Earth Sciences* 66:859-877. <https://doi.org/10.1007/s12665-011-1297-0>

Gui-Sheng H (2016) Case study of the characteristics and dynamic process of catastrophic debris flows in Wenchuan County, China. *Earth Sciences Research Journal* 20(2):661-669. <http://dx.doi.org/10.15446/esrj.v20n2.48026>

Ilija I, Tsangaratos P (2016) Applying weight of evidence method and sensitivity

analysis to produce a landslide susceptibility map. *Landslides* 13:379-397.

<https://doi.org/10.1007/s10346-015-0576-3>

Jebur MN, Pradhan B, Tehrany MS (2014) Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sensing of Environment* 152:150-165.

<https://doi.org/10.1016/j.rse.2014.05.013>.

Lee S, Choi J (2004) Landslide susceptibility mapping using GIS and the weight-of-evidence model International. *Journal of Geographical Information Systems* 18:789-814. <https://doi.org/10.1080/13658810410001702003>

Lee S, Sambath T (2006) Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models.

Environmental Geology 50:847-855.

<https://doi.org/10.1007/s00254-006-0256-7>

Lee S, Talib JA (2005) Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology* 47: 982-990.

<https://doi.org/10.1007/s00254-005-1228-z>

Li L, Lan H, Guo C, Zhang Y, Li Q, Wu Y (2016) A modified frequency ratio method for landslide susceptibility assessment. *Landslides* 14:727-741.

<https://doi.org/10.1007/s10346-016-0771-x>

Lichstein JW, Simons TR, Shiner SA, Franzreb KE (2002) Spatial Autocorrelation and Autoregressive Models in Ecology. *Ecological Monographs* 72:445-463.

[https://doi.org/10.1890/0012-9615\(2002\)072\[0445:SAAAMI\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2002)072[0445:SAAAMI]2.0.CO;2)

Luo W, Liu CC (2018) Innovative landslide susceptibility mapping supported by geomorphon and geographical detector methods. *Landslides* 15:465-471.

<https://doi.org/10.1007/s10346-017-0893-9>

Luo W, Jasiewicz J, Stepinski T, Wang J, Xu C, Cang X (2016) Spatial association between dissection density and environmental factors over the entire conterminous United States. *Geophysical Research Letters* 43:692-700.

<https://doi.org/10.1002/2015GL066941>

Meng Xiangrui, Pei Xiangjun, Liu Qinghua, Zhang Xiong, Hu Yunhua (2016) GIS-Based Environmental Assessment from Three Aspects of Geology, Ecology and Society along the Road from Dujiangyan to Wenchuan. *Mountain Research* 34:110-120. (In Chinese)

Petley D (2012) Global patterns of loss of life from landslides. *Geology* 40:927-930.

<https://doi.org/10.1130/G33217.1>

Pham BT, Bui DT, Pourghasemi HR, Indra P, Dholakia MB (2015) Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theoretical & Applied Climatology* 122:1-19. <https://doi.org/10.1007/s00704-015-1702-9>

Spiegelhalter DJ, Best NG, Carlin BP, Angelika VDL (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society* 64:583-639.

<https://doi.org/10.1111/1467-9868.00353>

Tobler W R (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46:234-240. <https://doi.org/10.2307/143141>

Wang HB, Sassa K (2005) Comparative evaluation of landslide susceptibility in Minamata area, Japan. *Environmental Geology* 47:956-966.

<https://doi.org/10.1007/s00254-005-1225-2>

Wang JF, Yong GE, Lianfa LI, et al (2014) Spatiotemporal data analysis in geography. *Acta Geographica Sinica* 69:1326-1345. (In Chinese)

Wang JF, Hu Y (2012) Environmental health risk detection with GeogDetector. *Environmental Modelling & Software* 33:114-115.

<https://doi.org/10.1016/j.envsoft.2012.01.015>

Wang JF, Li X H, Christakos G, et al (2010a) 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* 24:107-127. <https://doi.org/10.1080/13658810802443457>

Wang JF, Liao YL, LX (2010b) *Spatial Data Analysis Tutorial*. Science Press, Beijing, China. (In Chinese)

Wang JF, Zhang TL, Fu BJ (2016) A measure of spatial stratified heterogeneity. *Ecological Indicators* 67:250-256.

<https://doi.org/10.1016/j.ecolind.2016.02.052>

Wang Z, Wang Y, Wang L, Zhang T and Tang Z (2017) Research on the

comprehensive evaluation system of eco-geological environmental carrying capacity based on the analytic hierarchy process. *Cluster Computing* pp:1-10.

<https://doi.org/10.1007/s10586-017-1242-4>

Westen CJV, Castellanos E, Kuriakose SL (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology* 102:112-131. <https://doi.org/10.1016/j.enggeo.2008.03.010>

Xu C, Dai F, Xu X, Yuan HL (2012) GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* s145–146:70-80.

<https://doi.org/10.1016/j.geomorph.2011.12.040>

Xu X, Wen X, Yu G, Chen G, Klinger Y, Hubbard J, Shaw J (2009) Coseismic reverse- and oblique-slip surface faulting generated by the 2008 Mw 7.9 Wenchuan earthquake, China. *Geology* 37:515-518.

<https://doi.org/10.1130/G25462A.1>

Yesilnacar E, Topal T (2005) Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology* 79:251-266.

<https://doi.org/10.1016/j.enggeo.2005.02.002>

Yin Y, Wang F, Sun P (2009) Landslide hazards triggered by the 2008 Wenchuan earthquake, Sichuan, China. *Landslides* 6:139-152.

<https://doi.org/10.1007/s10346-009-0148-5>

Zhang H, Chi T, Fan J, Hu K, Peng L (2015) Spatial Analysis of Wenchuan

Earthquake-Damaged Vegetation in the Mountainous Basins and Its

Applications. *Remote Sensing* 7:5785-5804.

<https://doi.org/10.3390/rs70505785>

Zhuang Jianqi, Cui Peng, Ge Yonggang, et al (2010) Risk assessment of collapses and

landslides caused by 5.12 Wenchuan earthquake-A case study of

Dujiangyan-Wenchuan Highway. *Chinese Journal of Rock Mechanics &*

Engineering 29:3735-3742. (In Chinese)

<http://data.mlr.gov.cn> (2018) (Web references)

http://www.mlr.gov.cn/xwdt/chxw/201510/t20151019_1384486.htm (2018). (Web references)

<http://www.GeoDetector.org/> (2018). (Web references)

<http://www.r-inla.org/> (2018). (Web references)

Highlights

1. Spatial autocorrelation is important information for geographic data. Taking full advantage of this information can improve the accuracy of the landslide susceptibility map.
2. Making full use of spatial data attribute information and spatial structure information is conducive to reduce the uncertainty of landslide susceptibility mapping.
3. Eliminating redundant information and mining potential information can improve the predictability of the landslide susceptibility mapping.
4. Model integration can simultaneously meet the need to reduce multiple adverse impact on landslide landslide susceptibility mapping.