

Maximum Entropy-Based Model of High-Threat Landslide Disaster Distribution in Zhaoqing, China

Shaoxiong Yuan^{1,2,3}, Guangqing Huang^{1,2,3,*}, Haixian Xiong^{1,2,3}, Qinghua Gong², Jun Wang², Jun Chen²

¹ Guangzhou Institute of Geochemistry, Graduate School of the Chinese Academy of Sciences,
510640 Guangdong, China;
gdetrc@163.com

² Guangdong Open Laboratory of Geospatial Information Technology and Application,
Guangdong Geological Hazard Emergency Technology Research Center,
Guangzhou Institute of Geography,
100 Xianlie Middle Rd, 510070 Guangdong, China
hgq@gdas.ac.cn

³ University of Chinese Academy of Science, 100049 Beijing, China

Received 15 July 2017

Accepted 5 August 2017

Abstract

Landslide disaster that threatened over 100 people in Zhaoqing, China, were taken as samples. Sixteen environmental factors were selected, including altitude, slope degree, slope aspect, lithology, soil texture, normalized differential vegetation index (NDVI), average annual rainfall, distance to developed land, and distance to roads. The Maximum Entropy model was employed for simulation analysis of landslides. The results suggest that: NDVI, lithology, distance to rivers, distance to roads, rainfall variance, and altitude are the leading environmental factors that affect landslide disasters. Of the factors taken into consideration, distance to developed land contributes as much as 43.6% of the AUC (area under the curve) value of the landslide distribution model. In fact, this factor became the absolute leading variable over even the NDVI, indicating that high-threat landslide disasters in the study area are highly correlated with human activities. The closer the landslide location was to developed land, rivers, and roads, the more likely a landslide was to occur. Using the MaxEnt model, the high-threat landslide in Zhaoqing can be favourably simulated. The AUC of the model's prediction precision reached 0.769 without distance to developed land; whereas, the AUC of the model's precision reached 0.845 with distance to developed land taken into account.

Keywords: Maximum Entropy; Zhaoqing, China; Landslide Disaster; Distribution Model; Landslide Susceptibility

1. Introduction

The term landslide refers to the phenomenon of an area of land detaching from a slope. Due to the fact that landslides clearly have an impact on the economy as well as human life and property, governments and institutions have made an effort to evaluate landslide

susceptibility, disasters, and risk level of regions, and they have marked the spatial distribution of landslides on maps in order to improve their ability to prevent and respond to such disasters¹⁻². Quantitative studies of landslide disaster mapping based on GIS spatial data and set prediction models have been carried out since the 1990s³⁻⁵. Chung and Fabbri⁶ proposed a joint

* Corresponding author at: Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangdong Geological Hazard Emergency Technology Research Center, Guangzhou Institute of Geography, 100 Xianlie Middle Rd, 510070 Yuexiu, Guangzhou, Guangdong, China. E-mail address: gdetrc@163.com (G.Q. Huang).

conditional probability model and constructed a landslide disaster prediction map using GIS multi-layer spatial data in 1999. Subsequently, models [e.g., the Artificial Neural Networks⁷, Neuro-fuzzy⁸, and the Weights of evidence Model⁹] were successively used in studies of landslide disaster prediction mapping. However, it is difficult to establish zoning for landslide disasters and risks due to the complexity of the landslide-related environmental factors¹⁰.

Maximum Entropy Modeling has been successfully applied to computer vision, space physics, natural language processes, and many other fields¹¹. In terms of natural science, the MaxEnt model is often used in the geographic distribution of species^{12,13}, studies of wildlife¹⁴, and analysis and simulation of habitat suitability¹⁵ with a high prediction precision that has been widely recognized. There are several applications of this model in the landslide disaster distribution prediction since 2013, but the model's variables and prediction accuracy is still under discussion. Compared to those complex models that require precise sampling, the MaxEnt model may be used more easily for landslide prediction since it is able to analyze the relevance and importance of the driving factors of landslides without preset main trigger factors. As a result, an environmental variable test is crucial for landslide prediction^{16,17}. Felicísimo et al.¹ conducted a comparative analysis of landslide susceptibility mapping by means of logistic regression, multivariate adaptive regression spline function, categorical regression tree, and MaxEnt. They argued that with lithology taken into consideration, the categorical regression tree model and the MaxEnt model were the optimal methods, but the former lacked stability, the prediction accuracy of MaxEnt model is 77.80%. M. Convertino et al.¹⁶ used MaxEnt model to detect fingerprints of landslide drivers, they show that potential landslide hazard is strongly correlated with variation in the 12 and 48h rainfall with a return time of 10 years, and the prediction accuracy is 87%. By simulating and predicting landslide disasters in Pyeongchang, Korea using the MaxEnt model, Kim et al.¹⁸ concluded that the number of landslide disasters would increase during 2050–2090, during which rainfall would be the leading factor, with prediction precision reaching 92%. M. S. O'Banionl & M. J. Olsen¹⁶ generated seismically induced landslide hazard maps for western Oregon using MaxEnt Model, the prediction accuracy is lower

than 65%. Spatial modeling was conducted by Hong et al.¹⁹ who concluded that the MaxEnt model had favorable prediction precision in landslide susceptibility mapping using a data-driven evidential belief function (EBF), frequency ratio (FR), maximum entropy (MaxEnt), and logistic regression (LR), the prediction accuracy of MaxEnt model is 79.03%. It is evident that the MaxEnt model has received increasing recognition as a viable method of landslide disaster distribution simulation, but on the selection of variables and predictive accuracy of the model is still room for further study.

Guangdong Province, China has a complex natural environment and frequent geological disasters that result in a large loss of human life and property²⁰. In this study, the study area is set in Zhaoqing, Guangdong Province, China, with 16 environmental factors selected, including altitude, slope degree, slope aspect, lithology, soil texture, soil type, NDVI, annual average rainfall, annual maximum rainfall, rainfall during the wettest month, rainfall during the wettest season, rainfall variable coefficient, distance to roads, distance to developed land, distance to rivers, and distance to faults. The MaxEnt model is used to model major landslide disasters in Zhaoqing and predict the spatial distribution of landslide disasters in an effort to produce information that can be used in various fields, including geological disaster prevention and urban planning.

2. Methods

2.1. Study Area

Located in the Midwestern part of Guangdong Province and the middle reaches of the Xijiang River, Zhaoqing City (Fig. 1) is located at 22°47' – 24°24' N and 111°21' – 112°52' E where a subtropical monsoon climate prevails (Fig. 1). Zhaoqing City has an annual average temperature of 21.2°C and an annual average rainfall of 1650 mm, which is mainly concentrated in April and September. The main factors that trigger geological disasters in the study area are heavy rainfall, building construction by cutting slopes in mountainous areas, resource exploitation, and engineering construction. Geological disaster types are mainly residual deluvial weathering, slope collapse, and landslides in mountainous areas and surface collapse in the limestone region.

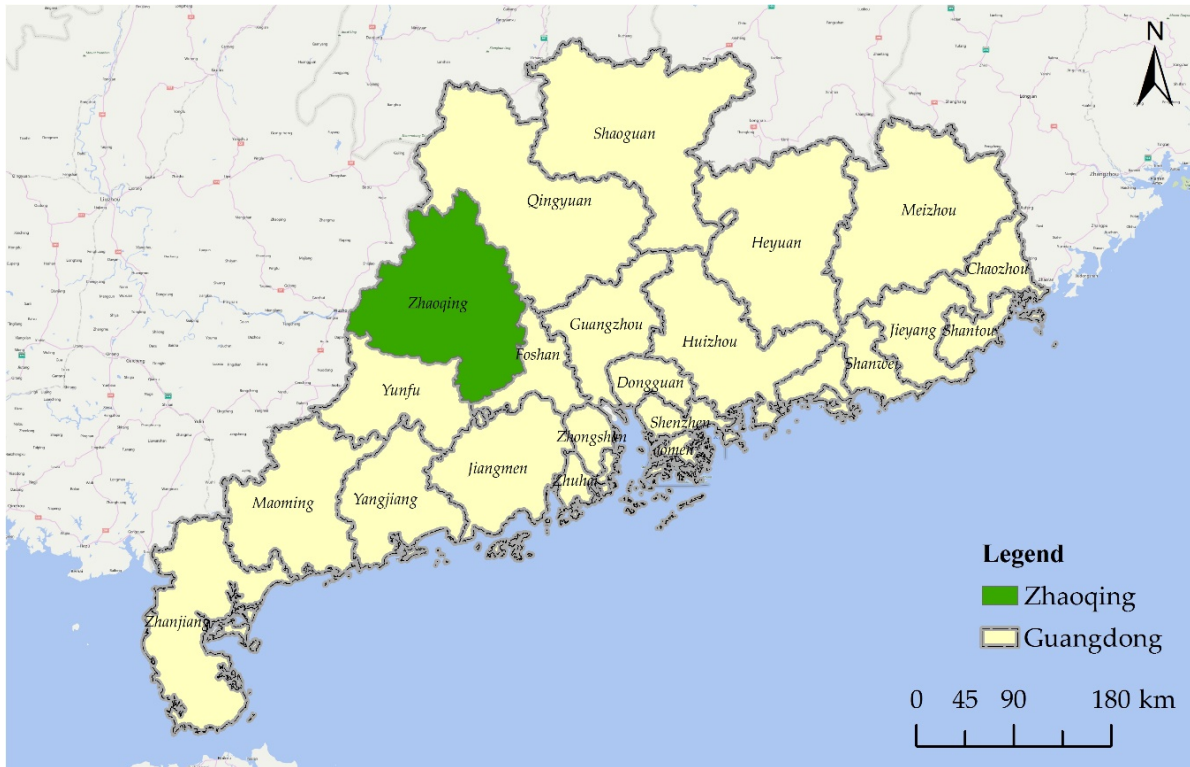


Fig. 1. Study area: Zhaoqing, Guangdong Province.

2.2. Model Description

In the MaxEnt model, all of the pixels in the study area are taken as possible distribution space that constitutes the maximum entropy. Pixels in the known object distribution spots are used as sampling points. Constraint conditions can be worked out according to the pixels of environmental variables, such as climatic variables, altitude, soil type, and vegetation forms in order to explore the possible distribution of the maximum entropy within such condition constraints (i.e., identifying the pixels with the same environmental variables as object distribution spots) and can thereby, predict the distribution of the object in the study site¹². In this model, the Jackknife test is employed to analyze the importance of the environmental variables; the area under the curve (AUC) ROC curve (receiver operating characteristic curve) is used to evaluate the precision of the MaxEnt model. A higher AUC means a greater relevance between environmental variables and the predicted target geographical distribution model, and thus, it is more likely to differentiate distribution from lack of distribution of certain objects and produce a

better prediction. In general, an AUC that is higher than 0.75 indicates that the model's prediction is effective²¹. The MaxEnt²² software for species habitat modeling (version: 3.3.3k) was used in this study. See Table 1 for the model's key parameters.

Table 1. Maxent model parameter settings.

Parameter	Setting
Random seed	Yes
Random test percentage	25%
Replicates	15
Replicated run type	Subsample
Maximum iterations	5000

2.3. Landslide Data

Landslide data for the position, type, longitude, and latitude of potential locations of geo-disasters is from the Guangdong Province Geological Disaster Prevention Scheme (2014–2016), which was compiled by the Division of Geological Environment, Guangdong Office of Land and Resources. This scheme is reported

as having high precision because potential locations of landslides that threaten over 100 people were determined to be high-threat landslide spots, which coincide with the information from disaster monitors. A landslide distribution map of landslide locations in Zhaoqing was created (Fig. 2).

There were 159 potential landslide locations that each would threaten over 100 people in Zhaoqing, Guangdong Province from 2014–2016. In this study, 75% of those potential locations were randomly selected to construct a landslide susceptibility model, while the remaining 25% used for model verification. The calculation was repeated 15 times to determine the mean value, which was used as the prediction result.

2.4. Environmental Variable Data

2.4.1. Environmental Variable Selection

To construct a landslide prediction model, it is particularly important to carefully select environmental

variables, which are usually landforms, soil texture, soil, land use, vegetation, and precipitation^{2,19,23-27}. In accordance with previous studies, altitude, slope degree, slope aspect, lithology, soil texture, soil type, NDVI, annual average rainfall, annual maximum rainfall, rainfall during the wettest month, rainfall during the wettest season, rainfall variable coefficient, distance to roads, distance to rivers, and distance to faults were chosen as environmental variables because scholars have commonly used them for landslide prediction; however, we also added distance to developed land as an environmental variable, as shown in Fig. 3–Fig. 10. The distance to developed land factor was seldom taken into consideration in previous studies, but the author of this study and some local scholars^{27,28} argue that landslide disasters in Guangdong Province, China were related to development of the area. As a result, “distance to developed land” was used as an environmental variable in this study to investigate its role in landslide disasters in Zhaoqing, Guangdong Province.

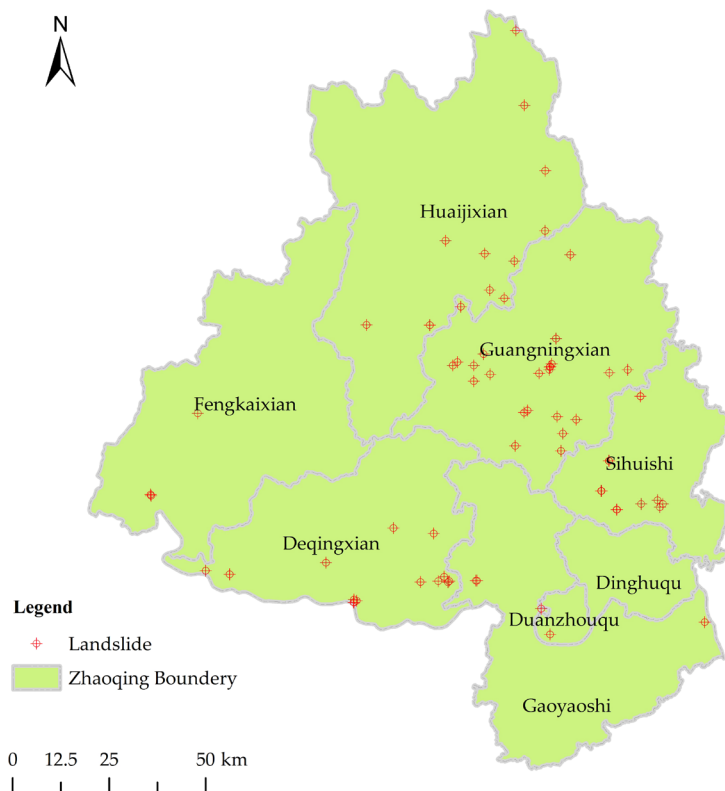


Fig. 2. Potential locations of high-threat landslides (Each of these landslides would threaten over 100 people).

The NDVI variable and the distance to developed land variable have several similarities and differences. Developed land (or urban area) can be classified using NDVI combined with normalized difference built-up index²⁹. NDVI is a non-linear variable, equivalent to the infrared/red ratio. Distance to developed land is generated using Euclidean distance tools and is linear variable. The distance to developed land variable describes each cell's spatial relationship to nearby

developed land based on the straight-line distance between them. In this study, both NDVI and developed land are deprived from Landsat8 image data analysis. The mean NDVI of developed land is 0.35 with a standard deviation of 0.19. The area encompassed by the distance to developed land variable (with a distance of less than 373.73 units) overlaps the area encompassed by the NDVI variable for values in the range of 0–0.5.

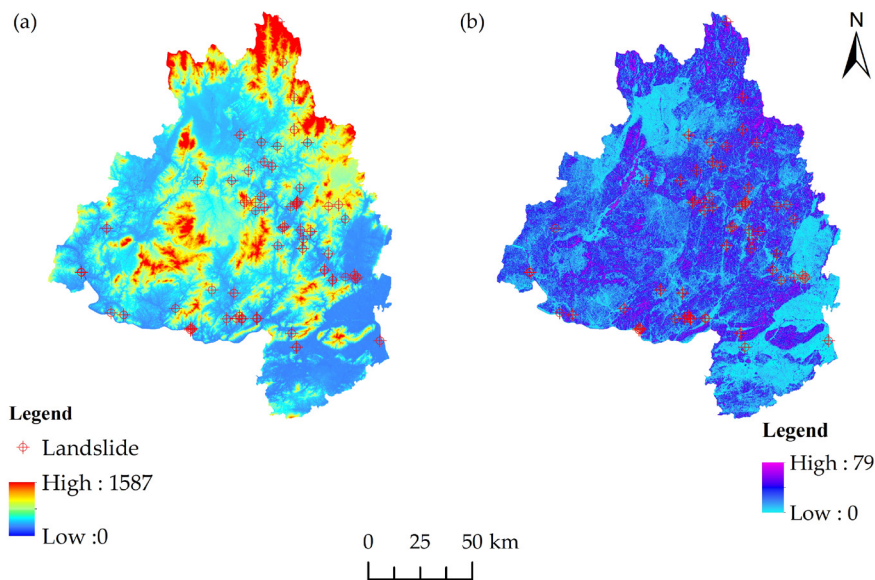


Fig. 3. Environmental variables in the study area: (a) evi01, altitude; (b) evi02, slope degree.

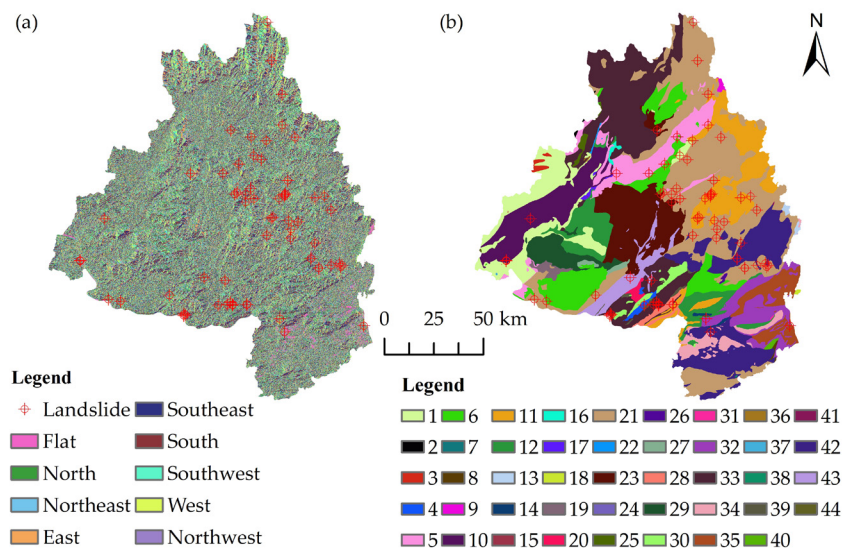


Fig. 4. Environmental variables in the study area (continuation): (a) evi03, slope aspect. Slope aspect was assigned to be: Flat-1, North-2, Northeast-3, East-4, Southeast-5, South-6, Southwest-7, West-8, and Northwest-9; and (b) evi04, lithology, refer to Table 2 for the lithology corresponding to the values..

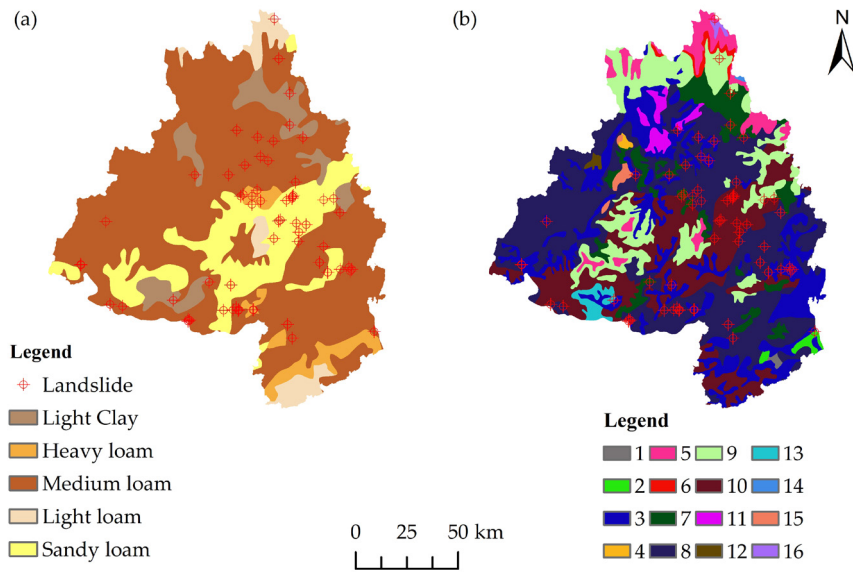


Fig. 5. Environmental variables in the study area (continuation): (a) evi05, soil texture. Soil texture was assigned to be: Light Clay-1, Heavy loam-2, Medium loam-3, Light loam-4, and Sandy loam-5. (b) evi06, soil type, refer to Table 3 for the soil types corresponding to the values.

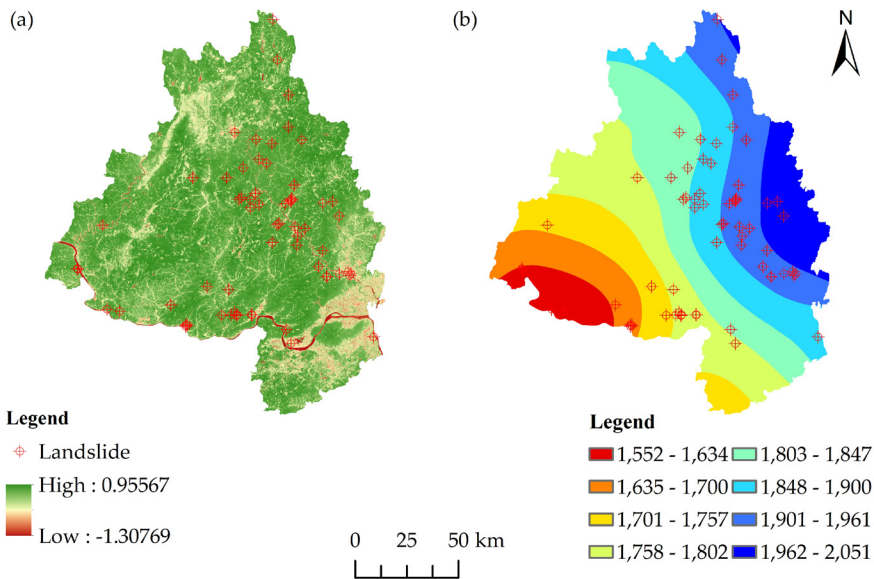


Fig. 6. Environmental variables in the study area (continuation): (a) evi07, NDVI; (b) evi08, annual precipitation.

2.4.2. Data Source of Environmental Variables

The digital elevation model (DEM) data used was GDEM V2 digital altitude data with a resolution of 30 m provided by the Geospatial Data Cloud (<http://www.gscloud.cn/>) of the Computer Network Information Center, Chinese Academy of Sciences.

Slope degree and slope aspect data were obtained through slope analysis using ArcGIS by DEM. As shown in Fig 3a, the elevation of Zhaoqing is comparatively high in the northeast and low in the southwest. This region is characterized by mountains of low-intermediate heights and hills with small plains, and it presents a mountainous geomorphology where

mountains, basins, hills, and alluvial plains are alternatively distributed.

The lithology data was obtained from the Guangdong 1:500,000 geologic map from the Digital Library of NGAC (National Geological Archives of China) (<http://geodata.ngac.cn/>). The lithologic composition is complex in study area and can be divided into 44 types (Table 2).

Soil texture and soil type data were derived from Guangdong Province Digital Soil V2.0 ([http://www.](http://www.soil.gd.cn/cn/gd.asp)

[soil.gd.cn/cn/gd.asp](http://www.soil.gd.cn/cn/gd.asp)) provided by the Guangdong Institute of Eco-Environment and Soil Sciences, with a precision of 1:1 million. See Fig. 5a for soil texture assignment. Soil textures in the study area are mainly light clay, heavy loam, medium loam, light loam, and sandy loam. Medium and sandy loam cover about 80% of the area, which contains 16 soil types. See Table 3 for various assignments.

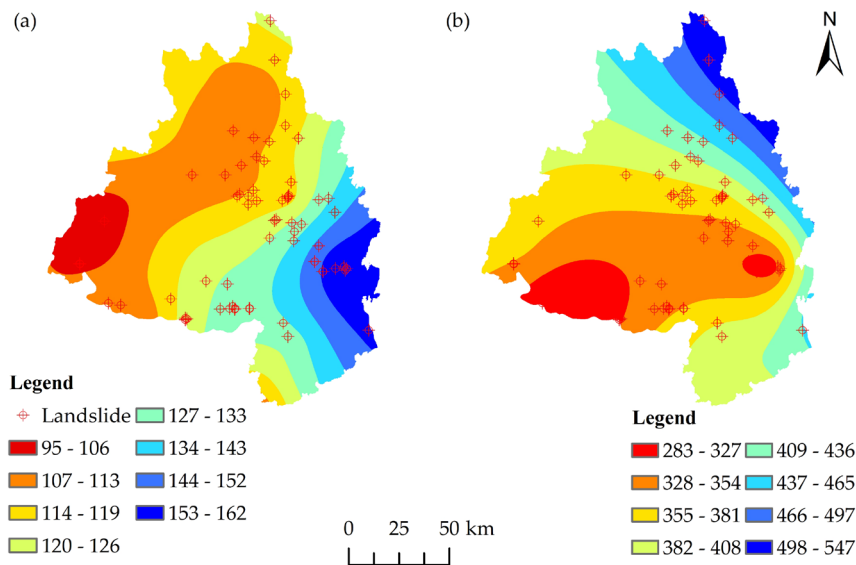


Fig. 7. Environmental variables in the study area (continuation): (a) evi09, annual maximum rainfall; (b) evi10, precipitation in the wettest month.

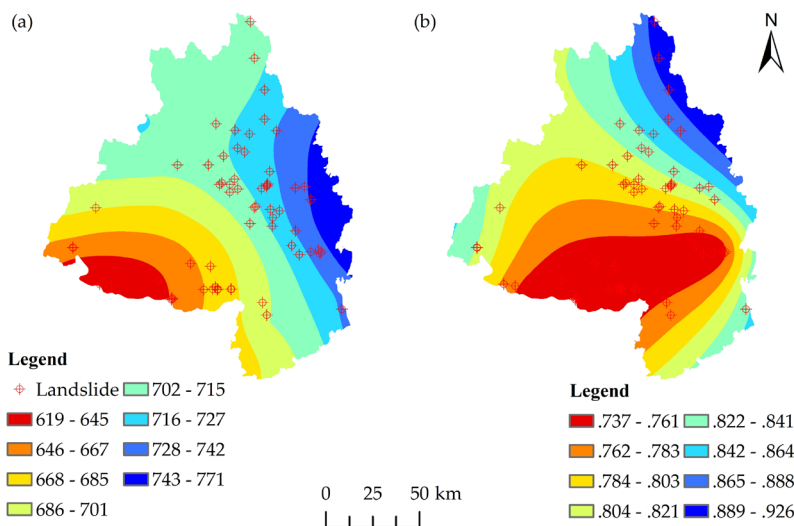


Fig. 8. Environmental variables in the study area (continuation): (a) evi11, precipitation in the wettest quarter; (b) evi12, coefficient of variation of seasonality of precipitation.

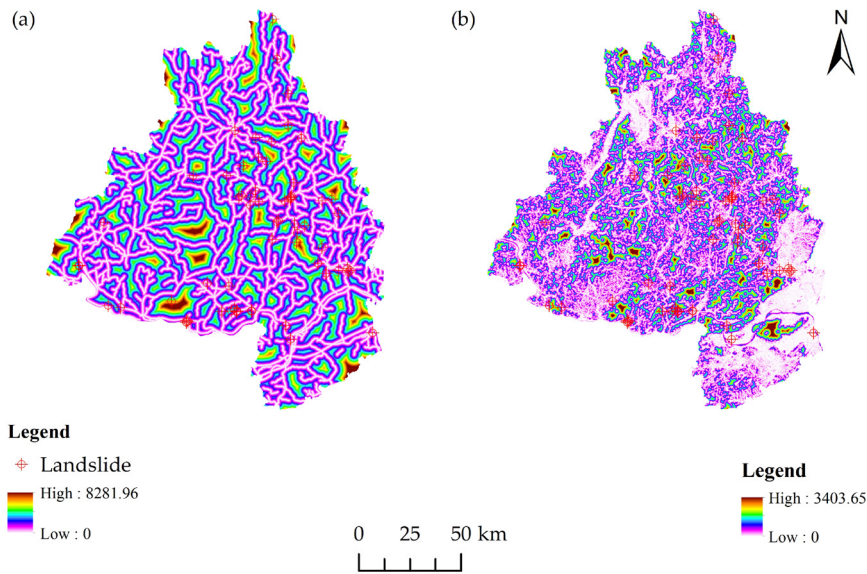


Fig. 9. Environmental variables in the study area (continuation): (a) evi13, distance to roads; (b) evi14, distance to developed land.

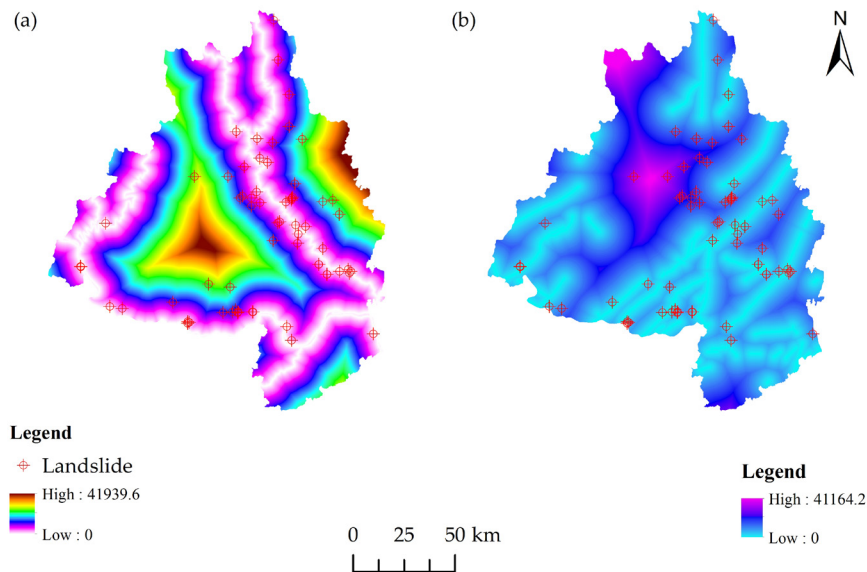


Fig. 10. Environmental variables in the study area (continuation): (a) evi15, distance to rivers; (b) evi16, distance to faults.

The NDVI data was obtained through GIS based on Landsat 8 OLI_TRIS provided by the Geospatial Data Cloud of the Computer Network Information Center, Chinese Academy of Sciences. The formula used to calculate NDVI is $NDVI = \frac{(IR - R)}{(IR + R)}$. Data for

variables concerning rainfall were generated using Kriging interpolation of GIS based on the precipitation data from 2013–2015, which were collected by the Guangdong Meteorological Service.

Data concerning roads and rivers was provided by the National Geomatics Center of China (<http://ngcc.sbsm.gov.cn/>), and fault data was provided by the Digital Library of NGAC. Developed land data

was obtained from unsupervised classification of Landsat 8 OLI_TRIS using GIS, by which land use type was divided into 6 types, i.e., river systems, forest land,

Table 2. Lithologic factor classification and assignment.

Value	Unit name	Lithology
1	Gaotan Formation	Thick-bedded blastopsammite and silt-slate, with lentoid limestone, marlstone or calcareous slate on the top as a marker bed
2	First Segment of Huangdongkou Formation	Inter-bedded layer of thick-massive quartz sandstone with arkose and shale gravel
3	Second Segment of Huangdongkou Formation	Inter-bedded layer of inequigranular sandstone, siltstone, and shale
4	Merged bed of Niujiaohe and Gaotan Formations	
5	Niujiaohe Formation	Thick-bedded blastopsammite with cinerous flag argillite that forms a rhythmite characterized by loess shale, stone coal, and siliceous phosphate lenticel and pyrite microcore.
6	Shuishi Formation	Slysch rhythmite made up of palimpsest sandstone and slate, and carbonaceous slate
7	Carboniferous Monzonite Granite	Monzonite Granite
8	Dasaiba Formation	Silty mudstone, argillaceous siltstone, limestone, marlstone, and carbonaceous mudstone
9	Merged beds of the Lianxian and Shidengzi Formations	
10	Lianxian Formation	Medium thick-bedded dolomitic limestone and dolomite, thin-bedded to medium-bedded stratified argillaceous limestone
11	Guitou Group (including Yangxi and Laohutou Formations)	Conglomerate, glutenite, and siltstone at the bottom; quartzose conglomerate, pebbly sandstone, quartz sandstone, siltstone, and silty shale
12	Laohutou Formation	quartzose conglomerate, pebbly sandstone, siltstone, and silty shale
13	Yangxi Formation	Conglomerate, glutenite with sandstone, and siltstone; characterized by polymictic conglomerate
14	Chunwan Formation	Fine sandstone, siltstone, shale with limestone, and calcareous sandstone
15	Donggangling Formation	Limestone, argillaceous limestone, marlstone, mudstone, and clastozoic limestone
16	Xindu Formation	Fine sandstone, siltstone, shale with hematite layer and carbonatite
17	Rongxian Formation	Dominated by limestone, also contains dolomite and dolomitic limestone
18	Baoyue Formation	Purplish grey siltstone, fine sandstone with conglomeratic sandstone, and mudstone
19	Middle Jurassic Granodiorite	Fine-grained granodiorite
20	Middle Jurassic Monzonite Granite	Fine-grained and medium-grained (porphyritic) biotite granite
21	Late Jurassic Biotite Granite	Coarse-grained, medium-grained, and fine-grained (porphyritic) biotite granite

Table 2. Lithologic factor classification and assignment (continuation).

Value	Unit name	Lithology
22	Early Cretaceous Biotite Granite, Mirolitic Granite	Fine-grained and medium-grained (coarse, medium, and fine-grained porphyritic) biotite granite, mirolitic granite
23	Nanxiong Group (including Zhenshui, Zhutian, and Dafeng Formations)	Conglomerate, sandy conglomerate, pebbly sandstone, inequigranular sandstone, siltstone, and silty mudstone
24	Sanshui Formation	Fuchsia, brownish-red siltstone, fine sandstone with sandy conglomerate, inequigranular sandstone, carbonaceous mudstone, and marlstone
25	Sanyajiang Formation	Volcanic breccia, sandstone, siltstone, sedimentary tuff, decite aphrolithic lava, decite, andesitic sedimentary tuff, andesite, felsite, and rhyolite
26	Ordovician Monzonite Granite	Coarse, medium, and fine-grained biotite adamellite
27	Luohong Formation	Conglomerate, sandy conglomerate, quartz sandstone with arkose gravel and pebbly phyllite, also contains medium and fine-grained sandstone, argillaceous siltstone, and sericite phyllite
28	Dongchong Formation	Fine-grained calcareous quartz sandstone, sandstone with quartz sericite micro-pinal schist
29	Ordovician-Silurian Granite	Granite
30	Permian Monzonite Granite	Medium-grained biotite adamellite
31	Beiling Formation	Clay rubble layer
32	Denglongsha Formation	Clay, yellow, fine silty sand, with sandy clay in part of the region
33	Dawanzhen Formation	Sand, gravel with sandy loam
34	Mugang Formation	Sandy clay, clay, ooze with rotten wood
35	Wanqingsha Formation	Sand, pebbly sand, and pebbly sandy silt
36	Huanggang Formation	Rubification sandy loam, gravel, and pebbles
37	Tertiary-Quaternary Gabbro	Gabbro
38	Ordovician Monzonite Granite	Fine and medium-grained (or porphyritic, amphibole) biotite adamellite
39	Liantan Formation	Dominated by grey, greyish-white, yellowish-white banded shale, with siltstone and sandstone, abundant with graptolite fossils
40	Xiaoping Formation	Greyish-white conglomerate, sand conglomerate, sandstone with black siltstone, carbonaceous shale, and thin-layered coal
41	Xiaoyunwushan Formation	Thick-bedded quartz sandstone, lithic sandstone with conglomerate, sand conglomerate, coal shale, coal line, and thin-layered coal
42	Daganshan Formation	Quartz-mica schist, mica-quartz schist, quartzite with carbon-based phyllite, silicalite, limestone, tuff and pyrite layer, with quartzite with conglomerate at the bottom
43	Bali Formation	Metamorphic feldspar quartz sandstone, fine tuff sandstone or fine clastic sedimentary limestone and siltstone, silty slate, slate, and carbonaceous slate
44	Second Segment of Peidi Formation	Fine grey sandstone, silty shale with multi-layer silicalite, and siliceous mudstone

Table 3. Soil type and assignment.

Value	Soil Type
1	Bleached paddy soil
2	Gleyed paddy soil
3	Waterloggogenic paddy soil
4	Yellow soil from shale
5	Yellow soil from gneiss
6	Yellow-red soil from gneiss
7	Red soil from shale
8	Latosolic red soil from shale
9	Red soil from gneiss
10	Latosolic red soil from gneiss
11	Acidic purple soil
12	Basic purple soil
13	Latosolic red clay soil
14	Black calcareous soil
15	Calcium lithosol
16	Mountain shrubby meadow soil

cultivated land, grassland, developed land, and bare land. Distance factors, including distance to roads, rivers, faults, and developed land, were calculated using the GIS Euclidean distance tool.

2.5. Analysis Process

In this study, the analytical flow diagram (Fig. 11) includes data integration and analysis, landslide susceptibility modeling using the MaxEnt model, model verification, optimal model selection based on the ROC curve, and reclassification of the model to determine the final distribution simulation diagram.

3. Results

Most previous studies have failed to take distance to developed land, which is the manifestation of human activities on the Earth’s surface, into consideration. However, the author of this paper argues that this factor is one of the important factors contributing to major landslide disasters that threatened over 100 people in Guangdong Province. To illustrate the difference including this factor makes, a comparative analysis of

the simulated results (with and without the distance to developed land factor) was conducted for landslide susceptibility zoning, prediction precision, and leading environmental variables.

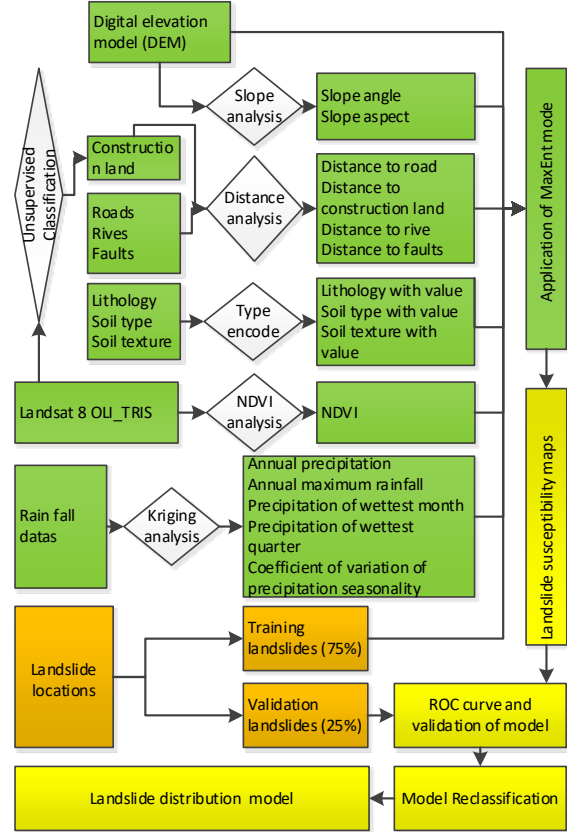


Fig. 11. Analysis process flow diagram of Maximum entropy-based model of high-threat landslide disaster distribution.

3.1. Distribution Model of Landslide Disaster

An ASCII result layer was produced along with the MaxEnt analysis. This layer was imported to ArcGIS for reclassification of landslide susceptibility assessment levels: $P < 0.1$, low susceptibility zone; 0.1–0.3, middle susceptibility zone; 0.3–0.6, high susceptibility zone; and $P > 0.6$, ultra-high susceptibility zone. As shown in Fig. 12, whether distance to developed land is included as an environmental variable or not, Sihui City, Guangning County and the area north of Gaoyao City in Zhaoqing are the main regions with ultra-high landslide susceptibility.

In the model with and without distance to developed land taken into account, the zones of ultra-high susceptibility mentioned above were 450 km² (Table 4) and 480 km², respectively, which indicates a difference

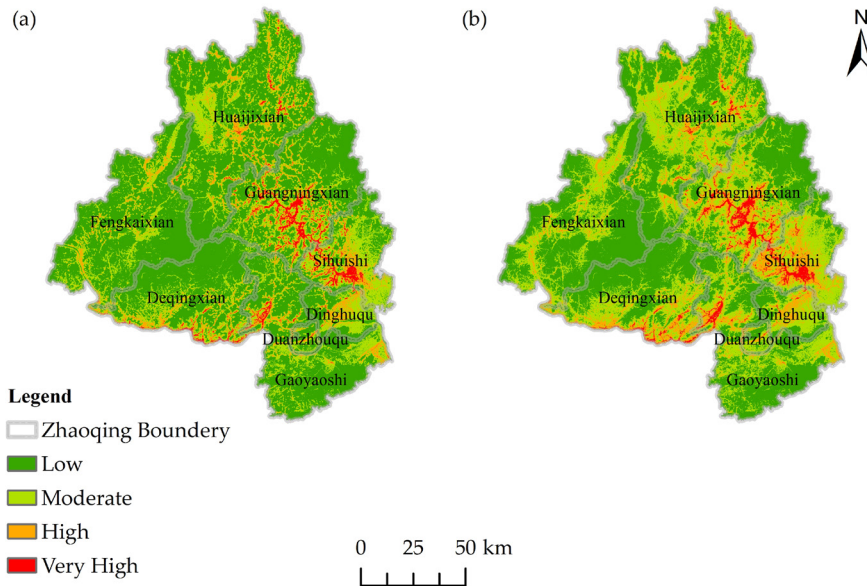


Fig. 12. Landslide susceptibility distribution model in Zhaoqing, Guangdong Province. (a) With distance to developed land model; (b) without distance to developed land model. low susceptibility zone: <math><0.1</math>, middle susceptibility zone: 0.1–0.3, high susceptibility zone: 0.3–0.6, ultra-high susceptibility zone: >0.6.

of 33 km² between them. The area classified as a high susceptibility zone was 1577 km² with this variable and 2116 km² without it, which indicates a difference of 539 km². It is evident that the areas of high and ultra-high susceptibility predicted by the model with the distance to developed land variable included were smaller.

Table 4. Landslide susceptibility area of the model.

Susceptibility	A1*	A2*
Low	11536	9046
Moderate	4213	6112
High	1557	2116
Very High	450	483

*A1: Area for model with distance to developed land variable (km²); A2: Area for model without distance to developed land variable (km²)

3.2. Prediction Precision

ROC curve (model predictability accuracy index) analysis was used for precision verification of potential landslide distribution results predicted by the MaxEnt software by determining the AUC value (i.e. area under the ROC curve), which ranged from 0.5–1. The closer this value was to 1, the better the prediction results and

the more accurate the model prediction result. The AUC value was divided into the following categories: 0.50–0.60 (failed), 0.60–0.70 (poor), 0.70–0.80 (common), 0.80–0.90 (good), and 0.90–1.00 (excellent)³⁰. As shown in Fig. 13, the AUC value of the model with distance to developed land included was 0.845, which indicates a good model precision and a higher prediction accuracy; whereas, the AUC value of the model without this variable was 0.769 (Fig. 14), which indicates a normal model precision and a comparatively lower prediction precision.

3.3. Leading Environmental Variables

Major environmental variables affecting landslide distribution can be identified by the contribution of the environmental variables that were used in the construction of the maximum entropy model. Table 5 shows that when distance to developed land is taken into consideration, it can be seen that it is a major factor that affects landslide disasters in Zhaoqing, Guangdong Province. The contribution rate of this factor was 43.6%, followed by lithology with a contribution rate of 13.5%, distance to rivers with a contribution rate of 10.4%, rainfall variable coefficient with a contribution rate of 7%, distance to roads with a contribution rate of 4.9%, annual rainfall with a contribution rate of 4.2%,

and altitude with a contribution rate of 3.2%, all of which add up to a total contribution rate of 86.6%.

Table 5. Analysis of variable contributions.

Variable	PC1*	PC2*
evi01	3.2	8.1
evi02	1.4	1.2
evi03	1.8	2.3
evi04	13.5	15.6
evi05	1.5	2.3
evi06	2.5	3.5
evi07	0.8	22.4
evi08	4.2	5.9
evi09	1.4	2.5
evi10	2.5	2.5
evi11	0.1	0.5
evi12	7	8.1
evi13	4.9	11.3
evi14	43.6	
evi15	10.4	12.3
evi16	1.1	1.6

*PC1: Percent contribution with distance to developed land variable; PC2: Percent contribution without distance to developed land variable

When distance to developed land is not taken into account, the major impact factor is NDVI, with a contribution rate of 22.4%, followed by lithology with a contribution rate of 15.6%, distance to rivers with a contribution rate of 12.3%, distance to roads with a contribution rate of 11.3%, rainfall variable coefficient with a contribution rate of 8.1%, altitude with a contribution rate of 8.1%, and annual rainfall with a contribution rate of 5.9%, all of which add up to a total contribution rate of 83.7%.

In summary, whether distance to developed land is taken into consideration or not, the major impact factors other than distance to developed land and NDVI always remain the same, but their contribution rates varied.

Jackknife is a statistical resampling technique that is particularly applicable to variance and bias estimation. With Jackknife, the impact of a certain factor on the model can be compared in the presence or absence of this factor. As shown in Fig. 15 and Fig. 16, whether distance to developed land is taken into account or not, single variables, such as altitude (evi01), NDVI (evi07), distance to rivers (evi15), distance to roads (evi13), and slope degree (evi02) show large contributions to the AUC of the model. With distance to developed land

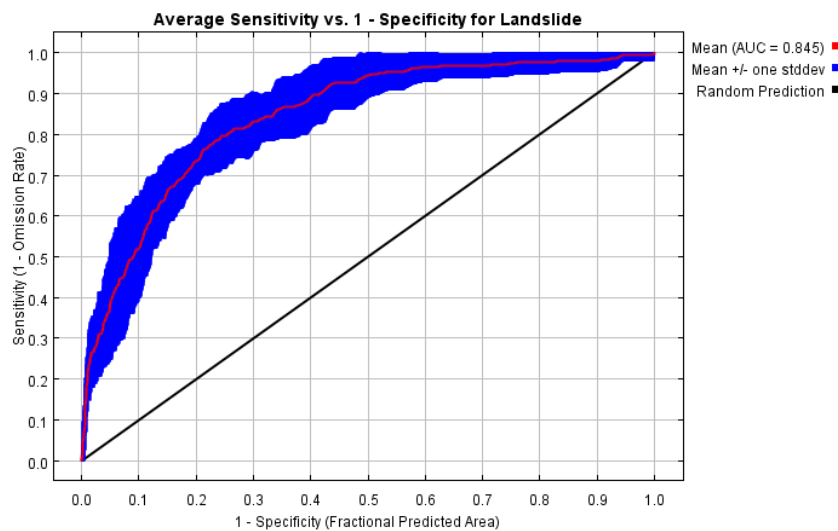


Fig. 13. ROC curve for potential landslide prediction with distance to developed land included.

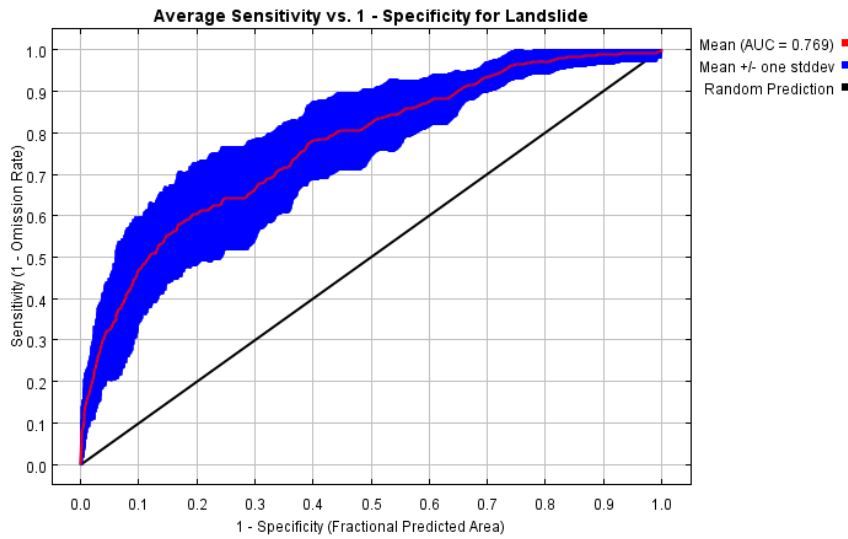


Fig. 14. ROC curve for potential landslide prediction without distance to developed land included.

(evi14) included in the model, evi14 turned into the single variable with the highest contribution to the AUC, indicating that these variables are the most effective ones for the landslide prediction model of the study area. By contrast, factors, such as slope aspect (evi03), soil type (evi06), and distance to faults (evi16) show smaller contributions to the AUC. Without them the prediction performance of the model might be enhanced, suggesting that they might not be suitable for landslide simulation in the study area.

3.4. Impact of the leading variables on landslides

3.4.1. Model with distance to developed land variable included

When distance to developed land is added to the model, this variable plays a dominate role in the model prediction because its contribution rate to the prediction result was 43.6%, which indicates a close relationship between potential landslide disasters and developed land

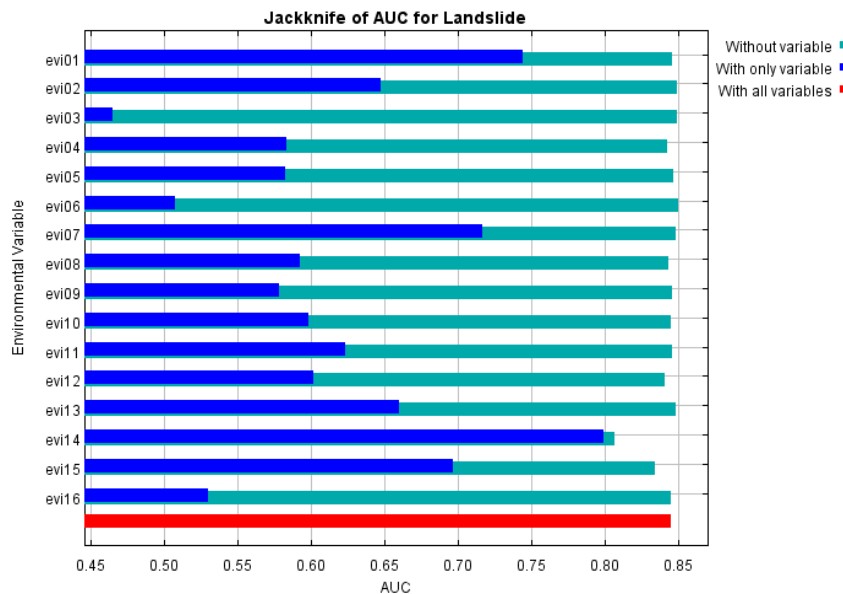


Fig. 15. Jackknife of AUC for Landslide with distance to construction land variable.

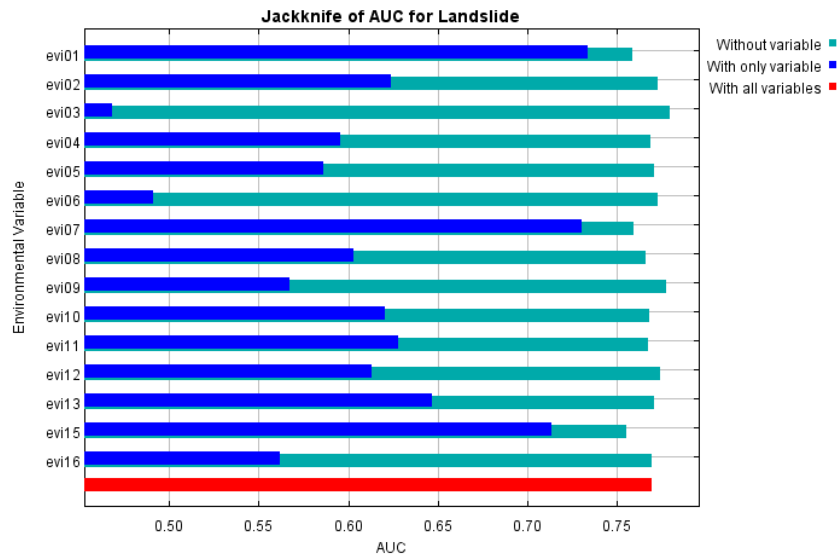


Fig. 16. Jackknife of AUC for Landslide without distance to construction land variable.

in the study area. As shown in Fig. 17, the closer the location is to developed land, the higher the probability of a landslide occurring. The fact that developed land is the type of land use with the highest level of excavation and the largest variation of land cover suggests that human activity is the major cause of potential landslide disasters and that reducing the impact of human activity would lead to a considerable decrease in landslide occurrence.

Lithology contributed 13.5% to the prediction model. Landslides responded significantly to various sub-factors of this factor, regions where the lithologic characteristics were lentoid limestone, marlstone, calcareous slate, quartz conglomerate, pebbly sandstone, quartz sandstone, siltstone and silty shale, sericite-quartz schist, and sandstone were more vulnerable to landslide disasters.

In this model, landslide response to distance to rivers and distance to roads also followed the rule that the closer it was to a certain variable, the more significant the landslide response. Contribution rates of the rainfall variable coefficient and the annual rainfall to landslide disasters were 7% and 4.2%, respectively, and 11.2% in

total, indicating that rainfall exerted a significant influence on landslides in this region and that a higher rainfall variation reflected a higher likelihood of landslide disasters.

3.4.2. Model without distance to developed land variable included

In the model without the land use variable, the contribution rate of NDVI to the prediction result was 22.4%. Fig. 18 shows that when NDVI is between 0–0.2, a landslide is most likely to occur. The NDVI of bare rock and sandy soil are usually low (≥ 0.1); whereas, that of sparse vegetation, shrub, and grassland are about 0.2–0.5³¹, which means uncovered surfaces are more prone to landslides.

In this model, the contribution rate of the lithology to the prediction model was 15.6%. Landslides also responded significantly to various sub-factors of this factor. The landslide's responses to other major variables, including distance to rivers, the rainfall variable coefficient, distance to roads, annual rainfall, and altitude are similar to the response in the model with distance to developed land included.

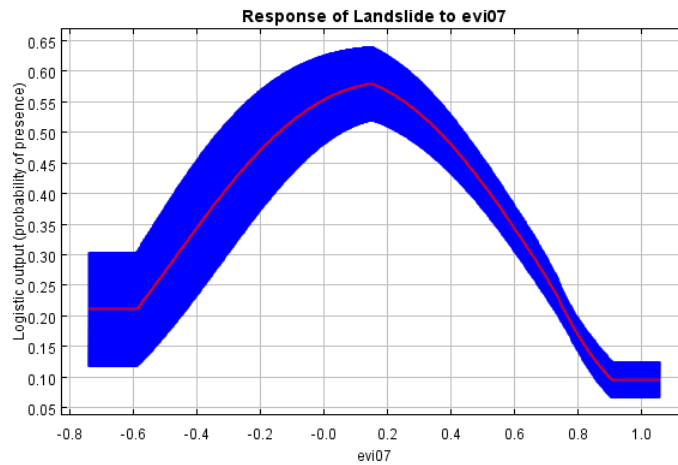


Fig. 18. Response of landslide to NDVI (evi07) without evi14 variable (The curves show the mean response of the 15 replicate Maxent runs (red) and the mean \pm one standard deviation (blue, two shades for categorical variables)).

4. Discussion

With distance to developed land included in the landslide disaster distribution model, high and ultra-high susceptibility zone areas that the model covered was smaller, which might be because the landslide data used in this study includes samples that threatened hundreds of people. Landslide cases that threatened over 100 people were those closer to developed land; therefore, with this variable included, high susceptibility zones were more precisely determined, and it can be seen that they are located around developed land. This would also explain why the precision of the model increased with distance to developed land included.

Landsat8 image data was used to extract two variables, i.e., NDVI and developed land. Both of these variables played a significant role in the model construction. In the model without distance to developed land included, a landslide was more likely to occur on bare rock and sandy soil when the NDVI, which contributed 22.4% to the model, was between 0–0.2. Whereas, in the model with distance to developed land included, developed land replaced NDVI and became the leading environmental factor with a contribution rate of 43.6%, as if it masked the role of NDVI in the model, causing the contribution rate of NDVI to decrease to 0.8%. This is most likely because the area encompassed by the distance to developed land

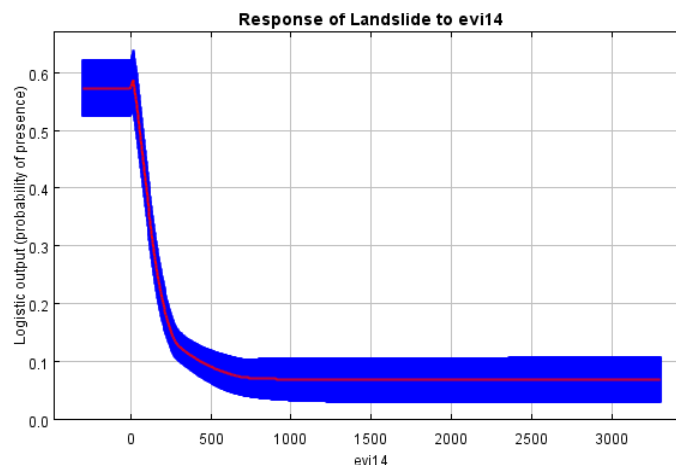


Fig. 17. Response of landslide to distance to construction land (evi14) (The curves show the mean response of the 15 replicate Maxent runs (red) and the mean \pm one standard deviation (blue, two shades for categorical variables)).

variable (with a distance of less than 373.73 units) overlaps the area encompassed by the NDVI variable with values in the range of 0–0.5, as described in section 2.4.1. Thus, the linear distance to developed land variable replaces the influence of the non-linear NDVI variable in the model, and more accurately expressed the relationship between landslides and environmental variables in the study area, and increased the overall accuracy of the model. The results of this study indicate that the distance to developed land variable is more suitable for landslide disaster distribution models than NDVI.

NDVI showed a low contribution to the model's precision with distance to developed land taken into consideration, but Fig. 15 suggests that NDVI, second only to distance to developed land and altitude, was another single variable that contributed significantly to the model, indicating that the NDVI was still of significance to landslide disasters in the study area.

Rainfall always served as a leading factor in other landslide distribution models or susceptibility studies^{16–18}; whereas, in this study, the impact of NDVI, lithology, distance to rivers, and distance to roads seem to exceed that of rainfall, which might be attributed to the fact that landslide sampling locations didn't include smaller landslide cases that threatened less than 100 people. In this paper, the results show that the MaxEnt model can be used in landslide distribution models; however, whether small landslides are also related to the distance to developed land or to rainfall is the main issue, and further studies are needed to determine this. As there is not sufficient data for small landslides in the study area, the same analysis using the same method and variables could be performed in other areas such as Wuhua²⁸, Guangdong, for which sufficient data is available.

Distances to developed land, rivers, and roads were the leading factors in this study, which agrees with the research on landslide disasters in Guangdong Province conducted by Xiongye Deng et al. who argued that side slope area around newly constructed roads, open-pit mines, deforestation, building construction that includes cutting slopes, and slopes at the banks of rivers and reservoirs were the governing factors of landslide disasters in Guangdong Province²⁷. They also suggest that major landslides are mainly caused by surface disturbances by human activities.

5. Conclusions

In this study, landslide disaster locations that threatened over 100 people were taken as samples, and the MaxEnt model was used for simulation analysis of landslide distribution. The conclusions of this study are as follows:

- In the model without distance to developed land included, NDVI, lithology, distance to rivers, distance to roads, the rainfall variable coefficient, altitude, and annual rainfall were the leading environmental variables contributing to landslide occurrence. In the model with distance to developed land included, distance to developed land, which contributed 43.6% to landslide occurrence, lithology, distance to rivers, the rainfall variable coefficient, distance to roads, annual rainfall, and altitude were the leading environment variables.
- The closer high-threat landslide locations were to developed land, rivers, and roads in the study area, the higher the likely hood of a landslide occurring. Landslide locations responded most significantly to NDVI values of 0–0.2, suggesting that areas with low vegetation coverage were more vulnerable to high-threat landslides. Regions whose surface lithologic characteristics were lentoid limestone, marlstone, calcareous slate, quartz conglomerate, pebbly sandstone, quartz sandstone, siltstone and silty shale, sericite-quartz schist, and sandstone were more vulnerable to landslide disasters. Annual rainfall and the rainfall variable coefficient played a significant role in landslide disaster samples in this study area.
- The MaxEnt model performed favorably in simulating high-threat landslide disasters in Zhaoqing, China. In the model without the distance to developed land variable, the AUC of the model prediction precision reached 0.769, a value rated as “normal.” In the model with the distance to developed land variable, the AUC of the model prediction precision reached 0.845, a value rated as “good.” These results indicate that the distance to developed land was applicable to simulation of landslide disasters that threatened more than 100 people.
- High-threat landslide disasters are highly correlated to human activities in the study area. As a result, particular attention must be paid to potential landslides during the preparation of geo-disaster prevention and urban planning schemes in places

where building construction results in cutting slopes, deforestation, road cuts, slopes around the banks of rivers and reservoirs. The results of our study show that the MaxEnt model can be used to aid in planning safe development of areas and in landslide prevention and prediction in areas that have already been developed.

Funding

This study was funded by the Natural Science Foundation of China (41271029, 41671506), the Science and Technology Project in Guangdong Province (2013B060500068, 2015B070701020), and the Special Fund Project for Environment and Capacity Construction of Guangdong Academy of Science Platform of Scientific Research (2016GDASPT-0103).

Acknowledgements

Acknowledgement: We would like to extend our gratitude for the help and guidance offered by Mr. Liu Wei, senior engineer of the Guangdong Meteorological Service with respect to rainfall data processing and analysis. We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

1. ÁM. Felicísimo, A. Cuartero, J. Remondo, et al. Mapping Landslide Susceptibility with Logistic Regression, Multiple Adaptive Regression Splines, Classification and Regression Trees, and Maximum Entropy Methods: a Comparative Study[J]. *Landslides*, 2013, 10(2): 175-189.
2. J. Davis, L. Blesius. A Hybrid Physical and Maximum-entropy Landslide Susceptibility Model[J]. *Entropy*, 2015, 17(6): 4271.
3. A. Carrara, M. Cardinali, F. Guzzetti. Uncertainty in Assessing Landslide Hazard and Risk[J]. *Itc Journal*, 1992, 2: 172-183.
4. S. Q. Wang, D. J. Unwin. Modelling Landslide Distribution on Loess Soils in China: An Investigation[J]. *International Journal of Geographical Information Systems*, 1992, 6(5): 391-405.
5. C. V. Westen, I. V. Duren, H. Kruse. Gissiz. Training Package for Geographic Information Systems in Slope Instability Zonation[J]. *Itc Publication*, 1993, 2(15): 2.
6. C. F. Chung, A. G. Fabbri. Probabilistic Prediction Models for Landslide Hazard Mapping[J]. *Photogrammetric Engineering and Remote Sensing*, 1999, 65(12): 1389-1399.
7. L. Ermini, F. Catani, N. Casagli. Artificial Neural Networks Applied to Landslide Susceptibility Assessment[J]. *Geomorphology*, 2005, 66(1): 327-343.
8. B. Pradhan, E. A. Sezer, C. Gokceoglu, et al. Landslide Susceptibility Mapping By Neuro-fuzzy Approach in a Landslide-prone Area (Cameron Highlands, Malaysia) [J]. *Ieee Transactions on Geoscience and Remote Sensing*, 2010, 48(12): 4164-4177.
9. B. Neuhäuser, B. Damm, B. Terhorst. Gis-based Assessment of Landslide Susceptibility on the Base of the Weights-of-evidence Model[J]. *Landslides*, 2012, 9(4): 511-528.
10. C. V. Westen, T. V. Asch, R. Soeters. Landslide Hazard and Risk Zonation—why Is It Still so Difficult?[J]. *Bulletin of Engineering Geology and the Environment*, 2006, 65(2): 167-184.
11. L. Zhang. Maximum Entropy Modeling[EB/OL]. [2017-2-24]. <http://homepages.inf.ed.ac.uk/lzhang10/maxent.html>.
12. S. J. Phillips, R. P. Anderson, R. E. Schapire. Maximum Entropy Modeling of Species Geographic Distributions[J]. *Ecological Modelling*, 2006, 190(3): 231-259.
13. S. J. Phillips, M. Dudík. Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation[J]. *Ecography*, 2008, 31(2): 161-175.
14. R. Baldwin. Use of Maximum Entropy Modeling in Wildlife Research[J]. *Entropy*, 2009, 11(4): 854.
15. S. Kumar, T. J. Stohlgren. Maxent Modeling for Predicting Suitable Habitat for Threatened and Endangered Tree *Canacomyrica Monticola* in New Caledonia[J]. *Journal of Ecology and the Natural Environment*, 2009, 1(4): 94-98.
16. M. Convertino, A. Troccoli, F. Catani. Detecting Fingerprints of Landslide Drivers: a Maxent Model[J]. *Journal of Geophysical Research: Earth Surface*, 2013, 118(3): 1367-1386.
17. M. O’banion, M. Olsen. Predictive Seismically-induced Landslide Hazard Mapping in Oregon Using a Maximum Entropy Model (maxent)[C]//Tenth U.s. National Conference on Earthquake Engineering, 2014.
18. H. Kim, D. Lee, Y. Mo, et al. Prediction of Landslides Occurrence Probability Under Climate Change Using Maxent Model[J]. *Journal of Environmental Impact Assessment*, 2013, 22(1): 39-50.
19. H. Hong, S. A. Naghibi, H. R. Pourghasemi, et al. Gis-based Landslide Spatial Modeling in Ganzhou City, China[J]. *Arabian Journal of Geosciences*, 2016, 9(2): 112.
20. H. Liu, A. Pan, Y. Wang, L. Luo. 2004. Geological disasters in Guangdong Province and their prevention and control. *J. Nat. Disasters*. 13(2): 101–105.
21. J. Elith. *Quantitative Methods for Conservation Biology*[M]: Springer New York, 2000: 39-58.
22. S. J. Phillips. A Brief Tutorial on Maxent[EB/OL]: Center for Biodiversity and Conservation, American

- Museum of Natural History [2017-5-31]. <https://www.cs.princeton.edu/~schapire/maxent/>.
23. F. Guzzetti, A. Carrara, M. Cardinali, et al. Landslide Hazard Evaluation: a Review of Current Techniques and Their Application in a Multi-scale Study, Central Italy[J]. *Geomorphology*, 1999, 31(1): 181-216.
 24. A. Brenning. Spatial Prediction Models for Landslide Hazards: Review, Comparison and Evaluation[J]. *Natural Hazards and Earth System Science*, 2005, 5(6): 853-862.
 25. M. E. Dickson, G. L. Perry. Identifying the Controls on Coastal Cliff Landslides Using Machine-learning Approaches[J]. *Environmental Modelling & Software*, 2016, 76: 117-127.
 26. V. Moosavi, Y. Niazi. Development of Hybrid Wavelet Packet-statistical Models (wp-sm) for Landslide Susceptibility Mapping[J]. *Landslides*, 2016, 13(1): 97-114.
 27. X. Deng, S. Yi. 2008. Space-time distribution characteristics of rockfall hazard activities in Guangdong. *Chin. J. Eng. Geophys.* 3: 356–363.
 28. Y. Luo. 2008. Preliminary study on the genesis, characteristic and the preventive and improving countermeasures of the geological hazards in Wuhua County, Guangdong Province. *Chin. J. Geol. Hazard Control*. 19(3): 96–101.
 29. Y. Zha, J. Gao, S. Ni. Use of Normalized Difference Built-up Index in Automatically Mapping Urban Areas from Tm Imagery[J]. *International Journal of Remote Sensing*, 2003, 24(3): 583-594.
 30. J. J. Wang, B. Cao, C. Bai, L. Zhang, et al. 2014. Potential distribution prediction and suitability evaluation of *fritillaria cirrhosa* D. Don based on maxent modeling and GIS. *Bull. Botanical Res.* (5): 642–649.
 31. USGS. NDVI, the Foundation for Remote Sensing Phenology[EB/OL]: USGS (2015-1-12) [2017-2-24]. https://phenology.cr.usgs.gov/ndvi_foundation.php.