Evaluation of Landslide Susceptibility Based on Logistic Regression Model

Zhen Du a, Biao Zhang b, Hong Hu c, Junji Bao d and Wenbin Li e, *
School of Nanchang University, Nanchang 360031, China.

a 1210273795@qq.com, b 3099574850@qq.com, c 1405610693@qq.com, d 2920222475@qq.com, *, e Corresponding author e-mail: faminghuang@ncu.edu.cn

Abstract. A lot of methods can be used for landslide susceptibility evaluation, such as support vector machine model, artificial neural network, etc. These models have good modeling effect, but often have the problem of low modeling efficiency. Hence, this paper proposes a simple and effective model of landslide susceptibility evaluation - Logistic regression model. The Ningdu county of Jiangxi province in China, with 297 recorded landslides, was used as study case. The 6 environmental factors including elevation, slope, profile curvature, distance to rivers, lithology and NDVI were extracted in this study. The analysis showed that the significance of Profile curvature was greater than 0.05, and there was a collinearity problem, so it was excluded. After the establishment of the factor evaluation system, the prediction rate curve is used to evaluate the accuracy of the model. The results show that the AUC value of the prediction rate curve of logistic regression model is 0.864, indicating that the evaluation accuracy of logistic regression model is high and the modeling is reasonable. In addition, landslides in the study area are mainly distributed along both sides of the rivers, and elevation and lithology play a major role in the occurrence of landslides.

Keywords: landslide; Evaluation of susceptibility; Logistic regression model; frequency ratio.

1. Introduction
As one of the most common natural disasters, landslides often occur in mountainous and hilly areas, which often bring huge property losses and casualties to people. The average annual economic loss caused by landslides in Jiangxi province amounts to tens of millions of yuan, with 20 to 30 casualties [1-5]. Therefore, it is of great practical significance and social value to carry out the study of landslide risk assessment for disaster reduction and prevention.

Landslide susceptibility assessment refers to exploring the probability of landslide occurrence in a specific area by studying the joint effect of multiple influencing factors [6-10]. The establishment of the model is very important for the evaluation of susceptibility. The common qualitative analysis is to evaluate and determine the main factors through expert scoring [11-12]. For example, AHP has been used to evaluate the susceptibility of landslide disaster [13-14]. Quantitative models have been widely applied in landslide prone evaluation, such as wu yuan less the information model is adopted for Xiamen, collapse and landslide geological disasters, The cluster analysis and support vector machine (SVM) have
been used to evaluate the landslide susceptibility [15-18]. The binary logistic regression model of quantitative statistical model has been widely applied to study the assessment of landslide disaster [19-22].

In conclusion, logistic regression model has been used in this study to evaluate landslide susceptibility. In Ningdu county of Jiangxi province in the south, for example, this study considers the topography, hydrological environment, the basic geology and the surface cover factor, finally, success rate curve of the model is employed to test precision, logistic regression model is subsequently found to be a more accurate and effective liability evaluation model of landslide, its evaluation results can provide theoretical guidance for the related department to disaster prevention and mitigation and help.

2. Logistic regression model

Logistic regression analysis model is a mathematical analysis method based on Logit transformation. It is widely used because of its simplicity and relative accuracy in landslide prediction. It takes the selected factor as the independent variable and the occurrence of landslide (occurrence = 1, non-occurrence = 0) as the dependent variable. The nonlinear classification statistical method is often used in regression analysis of the dependent variable of dichotomies. The calculation formula is as follows:

\[
Y = B_0 + B_1x_1 + B_2x_2 + \ldots + B_nx_n \tag{1}
\]

\[
P = \frac{\exp(Y)}{1 + \exp(Y)} \tag{2}
\]

\[
P_t = \frac{N^*_t / M^*}{N^*_i / M} \tag{3}
\]

Where \(x_n\) represents the NTH factor, and \(P_t\) represents the frequency ratio and serves as the model index. \(B_0\) represents logistic regression constant term, \(B_1\) to \(B_n\) is logistic regression coefficient; \(P\) is landslide prediction probability. A total of 2087 landslide units were selected in the study area. The same number of non-landslide data units were randomly selected and imported into SPSS 22 as the training data to calculate the constant term and regression coefficient of logistic regression, among which the frequency ratio was taken as the specific value of the training data. Finally, according to the equation (3), the value of each unit was calculated to represent the landslide susceptibility.

3. Overview and data source of the study area

3.1. Overview and classification of landslide evaluation unit in ningdu area

In this paper, the study area is located in Ningdu county of Ganzhou city, Jiangxi province, which lies between latitudes 26°05'N and 26°31'N, and longitudes 115°40'E and 116°17'E. The landslide is relatively developed in the area. As shown in figure 1, the elevation of the study area is between 154.906 m and 1059.74 m, the terrain is uneven. This area belongs to the subtropical monsoon humid climate, the average annual rainfall ranged from 1500mm to 1700 mm. The water system in Ningdu area is all over and the rock types in the study area are abundant.

3.2. Landslide evaluation unit division

Landslides are widely distributed in the study area. There are 297 landslide points with a total landslide level of about 1.88km², and are divided into 2,087 landslide grids. The middle and upper part of the landslide is relatively less, and the landslide is more concentrated in the south and more common in the
low terrain, the landslide points show the characteristics of distribution along the water system. The landslide body is mainly composed of quaternary accumulation layer, and the movement mode is mainly tractive integral sliding. Continuous heavy rainfall and human engineering activities are two important factors inducing landslides.

Evaluation unit division is an important link in the evaluation process of landslide susceptibility. Grid and slope unit are widely used by many scholars. Among them, grid unit is more operable in division and applicable to small and medium-sized scale topographic maps. In this study, grid unit is adopted as the evaluation unit [6-7]. In addition, choosing the right size grid is also crucial to the accuracy and rationality of the study. In this paper, 30 m×30 m grids are selected for analysis and research based on the actual situation of the research area. There are 1,898,935 grids in the research area.

3.3. Selection of basic environmental factors

3.3.1. Selection of basic environmental factors. Data sources of this study are as follows:(1) digital elevation data (DEM) with spatial resolution of 30m;(2) field investigation related data and landslide catalogue information;(3) geological map of Ningdu area with 1:50,000 measuring scale for extracting lithologic factor information;(4) remote sensing data in Ningdu area is used to extract Normalized Difference vegetation Index (NDVI).Landslide susceptibility assessment is to obtain the possibility of regional landslide in the same engineering geological environment by analyzing the spatial distribution of basic environmental factors [23-28]. Among them, basic environmental factors are mainly classified into four categories: topographical and geomorphic, basic geology, hydrologic environment and surface cover factors. In this paper, the geographical environment conditions of the study area were considered comprehensively, the strong correlation factors with the occurrence of landslide were selected to establish the evaluation index system, as shown in table 1.

![Fig.1 Overview of the study area and landslide catalog](image-url)
| Environmental factors | A variable's value | Total grid number | Interval grid proportion (%) | Landslide grid number | Landslide grid proportion (%) | Frequency ratio |
|-----------------------|--------------------|-------------------|-------------------------------|-----------------------|-------------------------------|-----------------|
| Height (m)            |                    |                   |                               |                       |                               |                 |
| 154.906~247.164       | 745,670            | 0.393             | 1,001                         | 0.480                 | 1.221                         |                 |
| 247.164~342.969       | 604,542            | 0.318             | 779                           | 0.373                 | 1.172                         |                 |
| 342.97~463.613        | 335,827            | 0.177             | 204                           | 0.098                 | 0.553                         |                 |
| 463.614~644.58        | 158,919            | 0.084             | 79                            | 0.038                 | 0.452                         |                 |
| 644.581~1059.738      | 53,977             | 0.028             | 24                            | 0.011                 | 0.405                         |                 |
| Slope (°)             |                    |                   |                               |                       |                               |                 |
| 0 ~ 4.555             | 598,985            | 0.315             | 248                           | 0.119                 | 0.377                         |                 |
| 4.556 ~ 9.506         | 557,047            | 0.293             | 891                           | 0.427                 | 1.455                         |                 |
| 9.507 ~ 14.853        | 401,427            | 0.211             | 632                           | 0.303                 | 1.433                         |                 |
| 14.854 ~ 21.785       | 247,353            | 0.130             | 261                           | 0.125                 | 0.960                         |                 |
| 21.786 ~ 50.502       | 94,123             | 0.050             | 55                            | 0.026                 | 0.532                         |                 |
| Profile curvature     |                    |                   |                               |                       |                               |                 |
| 0 ~ 1.894             | 692,157            | 0.364             | 645                           | 0.309                 | 0.848                         |                 |
| 1.895 ~ 4.144         | 600,636            | 0.316             | 783                           | 0.375                 | 1.186                         |                 |
| 4.145 ~ 6.748         | 366,666            | 0.193             | 405                           | 0.194                 | 1.005                         |                 |
| 6.749 ~ 10.537        | 186,825            | 0.098             | 202                           | 0.097                 | 0.984                         |                 |
| 10.538 ~ 30.071       | 52,651             | 0.028             | 52                            | 0.025                 | 0.899                         |                 |
| Distance to rivers (m)|                    |                   |                               |                       |                               |                 |
| 0 ~ 150               | 202,840            | 0.107             | 543                           | 0.260                 | 2.436                         |                 |
| 150 ~ 300             | 192,217            | 0.101             | 433                           | 0.207                 | 2.050                         |                 |
| 300 ~ 450             | 181,869            | 0.096             | 176                           | 0.084                 | 0.881                         |                 |
| 450 ~ 600             | 172,483            | 0.091             | 132                           | 0.063                 | 0.696                         |                 |
| >600                  | 1,149,526          | 0.605             | 803                           | 0.385                 | 0.636                         |                 |
| Lithology             |                    |                   |                               |                       |                               |                 |
| Magmatic rocks class  | 26,598             | 0.014             | 32                            | 0.015                 | 1.468                         |                 |
| Clastic rock class    | 515,919            | 0.272             | 412                           | 0.197                 | 0.741                         |                 |
| Carbonate rocks       | 590,728            | 0.311             | 532                           | 0.255                 | 0.836                         |                 |
| Metamorphic rock class| 763,219            | 0.402             | 1111                          | 0.532                 | 1.346                         |                 |
| Water area            | 2,471              | 0.001             | 0                             | 0.000                 | 0.000                         |                 |
| NDVI                  |                     |                   |                               |                       |                               |                 |
| -0.187 ~ 0.111        | 50,558             | 0.027             | 63                            | 0.030                 | 1.138                         |                 |
| 0.112 ~ 0.216         | 174,270            | 0.092             | 272                           | 0.130                 | 1.420                         |                 |
| 0.217 ~ 0.279         | 468,870            | 0.247             | 587                           | 0.281                 | 1.139                         |                 |
| 0.280 ~ 0.334         | 694,635            | 0.366             | 797                           | 0.382                 | 1.044                         |                 |
| 0.335 ~ 0.516         | 510,802            | 0.269             | 368                           | 0.176                 | 0.656                         |                 |

3.3.2. Frequency ratio analysis of basic environmental factors. In general, the basic environmental factors are obtained, the frequency ratio method can be used to reflect the quantitative statistics of the correlation between the attributes of the basic environmental factors and the landslide susceptibility. In this study, ArcGIS 10.2 have been applied to divide the selected basic environmental factors into 5 grades (in which, stratigraphic lithologic factors are divided according to stratigraphic combination) by natural break point method, and then the frequency ratio can be obtained by formula (3). The results are shown in table 1. When the frequency ratio is greater than 1, it indicates that the landslide occurs in this condition; when the frequency ratio is less than 1, it indicates that the landslide occurs in this condition [6-7]. The presence of collinearity between factors should be determined to prevent duplicate information. In this study, the same non-landslide data as landslide data were selected and imported.
into the statistical software SPSS 22 to analyze the correlation between environmental factors, and the results show that the correlation coefficient of factors are less than 0.3, as the weak correlation, the 10 factors can be used to establish evaluation index system of landslide liability.

(1) Topographic and geomorphic factors
Elevation is an important factor of landslide. As shown in figure 2 (a) and table 1, when the elevation is 154.906 ~ 342.969 m, the frequency ratio is greater than 1, and the landslide is more likely to occur. Profile curvature is defined as the slope of slope, which represents the change degree of slope in the vertical direction. As shown in figure 2 (c) and table 1, when the section curvature is between 1.895 and 6.748, the corresponding frequency ratio is greater than 1, indicating that when the section curvature is in this section, landslides are easy to occur.

(2) Hydrological environment factors
Water is an important factor of landslide. It will not only accelerate the erosion of rock soil, but also make the interlayer soil between the sliding surfaces easier to soften and rub, thus leading to the landslide more easily. In this study, through ArcGIS 10.2 software, DEM data were filled with depression, flow direction and discharge extraction and river connection to obtain water distribution map, and then polycyclic buffer analysis was conducted. As shown in figure 2 (d) and table 1, when the distance to rivers is less than 300 m, the frequency ratio exceeds 2, indicating that the closer to the water system, the more likely the landslide will occur.

(3) Basic geology and surface cover factors
Lithology is an important internal factor for the occurrence of landslides. As shown in figure 2 (e) and table 1, the corresponding frequency ratio of all magmatic rocks in the northwest direction and most metamorphic rocks in the southeast direction is greater than 1, which is conducive to the occurrence of landslides. NDVI is used to represent the extent of vegetation cover, and it ranges from (-1) to 1. As shown in figure 2 (f) and table 1, when NDVI is less than 0.334 (approximately 73.1% grid percentage), the frequency ratio is greater than 1, indicating that this section is conducive to landslide.

4. Evaluation of logistic regression model
There are 2087 landslide grid units in this study area, and 2087 non-landslide grid units are randomly selected from non-landslide areas. A value of 1 is assigned to the landslide points are , and a value of 0
is assigned to the non-landslide points. The frequency ratio can be taken as the index value, a digital matrix is formed and imported into SPSS 22 software for analysis.

Logistic regression analysis requires that there is no collinearity relationship between each sample, that is, the Sig value of significance test index should be less than 0.05, and the result indicates that the profile curvature isn't required, so it is eliminated, and the result is shown in table 2. Table 3 shows that the influence degree of each factor on the occurrence of landslide is from large to small: lithology (1.72), distance to rivers (1.593), elevation (1.299), NDVI (0.621) and slope (0.614). Among them, the weight of lithology and distance to rivers is the largest, indicating that these two factors affect the occurrence of landslide to a large extent. The corresponding regression coefficients of NDVI and slope were both less than 1, indicating that they had the least influence on the occurrence of landslide.

The regression coefficient after eliminating section curvature was substituted into equations (2) and (3) to calculate the probability value of landslide occurrence corresponding to each grid, and the probability value was used to characterize the landslide susceptibility. The natural discontinuity point method was also used for grading, and the susceptibility partition diagram was shown in figure 3. As shown in table 2, it is divided into five grades: high-prone area (0.792–0.993), high-prone area (0.584–0.792), medium-prone area (0.386–0.584), low-prone area (0.209–0.386) and low-prone area (0.007–0.209), with the proportions of each grade being 15.9%, 14.9%, 19.0%, 24.0% and 26.2%, respectively.

**Table 2.** Distribution of disaster in all landslide susceptibility partitions

| Model           | Susceptibility level | Prone zoning grid number | Landslide grid scale | Total grid number | Interval grid scale | Frequency ratio |
|-----------------|----------------------|--------------------------|----------------------|-------------------|---------------------|-----------------|
| Logistic regression model | High                 | 0.792–0.993              | 997                  | 0.478             | 300,991             | 0.159           | 3.014           |
|                 | Higher               | 0.584–0.792              | 419                  | 0.201             | 282,198             | 0.149           | 1.351           |
|                 | Medium               | 0.386–0.584              | 306                  | 0.147             | 361,686             | 0.190           | 0.770           |
|                 | The lower            | 0.209–0.386              | 225                  | 0.108             | 456,330             | 0.240           | 0.449           |
|                 | Low                  | 0.007–0.209              | 140                  | 0.067             | 497,730             | 0.262           | 0.256           |

**Table 3.** Significance results of logistic regression model

| Landslide factor | Logistic regression model B value | Significant |
|------------------|-----------------------------------|-------------|
| Elevation        | 1.299                             | 0           |
| Slope            | 0.614                             | 0           |
| Distance to rivers | 1.593                           | 0           |
| Lithology        | 1.72                              | 0           |
| NDVI             | 0.621                             | 0.001       |
| Constant         | -10.595                           | 0           |
5. Discussion
From table 2, it shows that the frequency ratio gradually decreases from the high-prone area to the low-prone area, and the frequency ratio of logistic regression model in the high-prone area is 3.014, respectively. It can be concluded that the evaluation results of logistic regression model are good. As shown in figure 3, on the whole, the areas prone to high and high landslides show centralized distribution along the water system, which is consistent with the actual situation in table 1 and previous studies, further confirming the important role of water in the process of landslide.

In this study, the accuracy of logistic regression model is evaluated by the prediction rate curve, and the degree of conformity between model prediction and actual situation is verified by the proportion of landslide grids in each prone area. The higher the AUC value, the higher the model accuracy and the better the prediction effect. The prediction rate curve of logistic regression model is shown in figure 4, and the AUC value is 0.864 respectively. The evaluation accuracy shows that logistic regression model is high. The proportion of landslide grid in each prone area is obtained by superposition analysis of historical landslide distribution and prone area map. As shown in table 2, the proportion of landslide grids in the high-prone areas and low-prone areas of the logistic regression model is 47.8% and 6.7%, respectively, indicating that the logistic regression model is in line with the actual situation, and the prediction of the model is relatively accurate.
6. Conclusion

The AUC accuracy of the landslide susceptibility of logistic regression model in Ningdu area is 0.864 respectively, indicating that the logistic regression model reflects the landslide comprehensively and has high evaluation accuracy. According to the logistic regression model, the proportion of landslide grids in the highly vulnerable zones is 47.8%, while that in the low-vulnerable zones is 6.7%, which indicates that the logistic regression model is in good agreement with the actual situation.

The exploitable zoning map in this study shows that the exploitable zoning map corresponding to logistic regression model is mainly distributed along the water system, that the results show that water plays an important role in the occurrence of landslides, and elevation and lithology also have a great influence on the occurrence of landslides.

References

[1] Su Chenxu, Tian Qin, Liu Benchao, et al. Regional Landslide Susceptibility Assessment for Longman County in Jiangxi Province [J]. Science Technology and Engineering, 2019, 19(17):91-99.

[2] Faming Huang, Kunlong Yin, Guirong Zhang, et al. Landslide displacement prediction using discrete wavelet transform and extreme learning machine based on chaos theory[J]. Environmental Earth Sciences, 2016, 20(75): 1-18.

[3] Faming Huang, Xiaoyan Luo, Weiping Liu. Stability Analysis of Hydrodynamic Pressure Landslides with Different Permeability Coefficients Affected by Reservoir Water Level Fluctuations and Rainstorms[J]. Water, 2017, 9(7): 450.

[4] Faming Huang, Jinsong Huang, Shuihua Jiang, et al. Landslide displacement prediction based on multivariate chaotic model and extreme learning machine[J]. Engineering Geology, 2017, 218: 173-186.

[5] Zhu A X, Wang R, Qiao J, et al. An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic[J]. Geomorphology, 2014, 214: 128-138.

[6] Faming Huang, Chi Yao, Weiping Liu, et al. Landslide susceptibility assessment in the Nantian area of China: a comparison of frequency ratio model and support vector machine[J]. Geomatics, Natural Hazards and Risk, 2018, 9(1): 919-938.

[7] Faming H, Kunlong Y, Shuihua J, et al. Landslide susceptibility assessment based on clustering
analysis and support vector machine[J]. Chinese Journal of Rock Mechanics and Engineering, 2018.

[8] Li, D.; Huang, F.; Yan, L.; Cao, Z.; Chen, J.; Ye, Z. Landslide Susceptibility Prediction Using Particle-Swarm-Optimized Multilayer Perceptron: Comparisons with Multilayer-Perceptron-Only, BP Neural Network, and Information Value Models. Appl. Sci. 2019, 9, 3664.

[9] Huang Faming, Jing Zhang, Chuangbing Zhou, Yuhao Wang, Jinsong Huang & Li Zhu*. A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction[J], Landslides, 2019. DOI: 10.1007/s10346-019-01274-9

[10] Wu Xueling, Shen Shaoqing, Niu Ruiqing. Landslide Susceptibility Prediction Using GIS and PSO-SVM[J]. Geomatics and Information Science of Wuhan University, 2016, 41(05):665-671.

[11] He Peng, Tong Liqiang, Guo Zhaoccheng, et al. Evaluation Research on the Landslide Disaster Liability in Zhada Region of Tibet[J]. Science Technology & Engineering, 2016, 16(25):193-200.

[12] Nicu I C. Application of analytic hierarchy process, frequency ratio, and statistical index to landslide susceptibility: an approach to endangered cultural heritage[J]. Environmental earth sciences, 2018, 77(3): 79.

[13] Wu Y, Li W, Liu P, et al. Application of analytic hierarchy process model for landslide susceptibility mapping in the Gangu County, Gansu Province, China[J]. Environmental Earth Sciences, 2016, 75(5): 422.

[14] Dou J, Yamagishi H, Pourghasemi H R, et al. An integrated artificial neural network model for the landslide susceptibility assessment of Osado Island, Japan[J]. Natural Hazards, 2015, 78(3): 1749-1776.

[15] Zizheng Guo, Kunlong Yin, Lei Gui, Qingli Liu, Faming Huang, Tengfei Wang. Regional Rainfall Warning System for Landslides with Creep Deformation in Three Gorges using a Statistical Black Box Model[J], Scientific Reports, 2019, 9,8962

[16] Chen W, Xie X, Wang J, et al. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility[J]. Catena, 2017, 151: 147-160.

[17] Huang Faming, Wang Yang, Dong Zhiqiang, Wu Lizhou, Guo Zizheng, Zhang Taili. Regional Landslide Susceptibility Mapping Based on Grey Relational Degree Model[J]. Earth Science, 2019,44(02):664-676.

[18] Faming Huang, Kunlong Yin, Jinsong Huang, et al. Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine[J]. Engineering Geology, 2017, 223: 11-22.

[19] Wang Rui, Zhang Taili. Evaluation of landslide susceptibility based on GIS and logistic regression method [J]. Cryogenic building technology,2018,40(06):95-97.

[20] Xu Chong, Lv Chunlai. Comparative study on landslide susceptibility evaluation model of wenchuan earthquake in 2008 [J]. World earthquake translation, 2018, 49(3): 268-286.

[21] Faming Huang, Kunlong Yin*, Tao He, et al. Analysis of Influence Factors and Displacement Prediction of Reservoir Landslide-A Case Study of Three Gorges Reservoir, China[J]. Technical Gazette, 2016, 23(02):617-626.

[22] Faming Huang, Jinsong Huang, Shui-Hua Jiang, et al. Prediction of groundwater levels using evidence of chaos and support vector machine[J]. Journal of Hydroinformatics, 2017, 19(4): 586-606.

[23] Yuanyao Li, Jinsong Huang, Shui-Hua Jiang, et al. A web-based GPS system for displacement monitoring and failure mechanism analysis of reservoir landslide[J]. Scientific reports, 2017, 7(1): 17171.

[24] Huang F, Chen L, Yin K, et al. Object-oriented change detection and damage assessment using high-resolution remote sensing images, Tangjiao Landslide, Three Gorges Reservoir, China[J]. Environmental Earth Sciences, 2018, 77(5):183.
[25] Jiang S H, Huang J, Huang Faming*, et al. Modelling of spatial variability of soil undrained shear strength by conditional random fields for slope reliability analysis[J]. Applied Mathematical Modelling, 2018, 63:374~389.

[26] Huang F, Yin K, Yang B, et al. Step-Like Displacement Prediction of Landslide Based on Time Series Decomposition and Multivariate Chaotic Model[J]. Earth Science, 2018, 43(3):887-898.

[27] Zhang J, Yin K, Wang J, et al. Evaluation of landslide susceptibility for Wanzhou district of Three Gorges Reservoir[J]. Chinese Journal of Rock Mechanics & Engineering, 2016, 35(2):284-296.

[28] YANG Beibei, YIN Kunlong, HUANG Faming, et al. Stability analysis of Baijiabao landslide under reservoir water level fluctuation and rainfall[J]. Journal of China Institute of Water Resources & Hydropower Research, 2016(6):460-467.