Evaluation of Seismic Landslide Susceptibility by Integrating Statistical Learning Model and Newmark Model—A Case Study of the Wenchuan Earthquake

Zelang MIAO a, Minghui PU b, Yueguang HE b,1, Ke LI a, Renfeng PENG a, Shuai CHEN a, Changbo JIANG b, Bin DENG b and Zhibin HE c

a School of Geo-Sciences and Info-Physics, Central South University, Changsha, China
b School of Hydraulic Engineering, Changsha University of Science and Technology, Changsha, China
c School of Civil Engineering, Changsha University of Science and Technology, Changsha, China

Abstract. Whether it can quickly and effectively predict the susceptibility of regional earthquake landslides to achieve rapid rescue, loss assessment and post-disaster reconstruction has always been a difficult problem. However, the traditional high-precision evaluation of seismic landslide susceptibility often relies heavily on the complete or incomplete landslide inventory, which is poor in timeliness and cannot effectively evaluate the target area before or shortly after the earthquake. In most cases, the Newmark model relies on experts’ experience to select model parameters, therefore the evaluation result of this method is unstable and it lacks strong generalization ability. A fused model is proposed to classify the positive and negative training samples of the study area through the evaluation results of the Newmark model under the slope units, and it applies a variety of statistical learning models to evaluate the landslide susceptibility of the Wenchuan earthquake based on the classification results of the Newmark model. The results show that the evaluation of the statistical learning model fused with the Newmark model has higher accuracy. This method can overcome the inherent shortcomings of a single Newmark model to obtain better evaluation results without relying on obtaining the complete landslide inventory. Meanwhile, the model can be applied to quickly obtain the evaluation results of regional landslide susceptibility before or shortly after the earthquake, thereby effectively reducing human and economic losses caused by earthquake landslides.

Keywords. Earthquake landslide, landslide susceptibility, newmark model, statistical learning model, slope units

1. Introduction

Earthquakes are one of the environmental geological disasters with the longest existence and greatest impact in history. As a common destructive natural phenomenon, landslide is one of the major components of seismic secondary disasters. Earthquake
landsides tend to occur in clusters in space, and potential landslide occurrence sites within the disaster area can exist for years or even decades after the earthquake. For this reason, effectively assessing and predicting the susceptibility of landslides before and after earthquakes in a timely and effective manner is an important foundation for cooperating with disaster area to achieve the risk assessment and post-disaster reconstruction [1].

There are many research methods for earthquake landslide susceptibility at home and abroad, among which the typical representatives are the Newmark model and the statistical analysis method based on different statistical learning models [2-3]. The Newmark model is one of the most significant research methods in the field of the evaluation of earthquake landslide susceptibility. For example, it has been frequently applied to analyze the 2008 Wenchuan earthquake [4], the 2013 Lushan earthquake [5], the 2015 Nepal earthquake [6] and the 2017 Jiuzhaigou earthquake [7]. However, there are obvious flaws in this model. Many parameters in the model need to rely on manual empirical selection, and the overall model calculation requires more regional geological engineering data. Therefore, its accuracy is largely dependent on the manually selected parameters, and the model calculation results are less stable and have poor generalization ability.

The evaluation and mapping of landslide susceptibility based on statistical learning models have received much attention in related research fields in recent years. Commonly used methods include Neural network model, Hierarchical analysis model [8], and Logistic regression model [9-11]. These statistical learning models are superior in terms of evaluation stability and generalization ability compared to the Newmark model. However, the time efficiency of earthquake landslide susceptibility mapping based on statistical analysis model tends to be rather low due to the difficult access to the complete or incomplete landslide inventory. It is difficult for the statistical learning models to be applied to the potential risk prediction of landslide hazardous zones before the earthquake and the rapid assessment of landslide susceptibility in the short time after the earthquake.

We intend to select the Longmenshan fault zone in Sichuan Province, which is the epicenter of the Wenchuan earthquake, as the study area, and classify the potential landslide hazardous zones by the results of the landslide cumulative displacement calculation of the Newmark model. Then use the two types of zones as the positive and negative sample selection area for several types of statistical learning models, and derive the evaluation results of landslide susceptibility and the corresponding accuracy evaluation are finally carried out. The experimental results show that this fused model has higher landslide prediction accuracy and reliability than the Newmark model alone, and can effectively overcome the destabilization problem of the Newmark model in the evaluation accuracy of seismic landslide susceptibility.

2. Study Area and Data Sources

2.1. Study Area

The main body of the Longmenshan fault zone is located in the subtropical region, in the territory of Sichuan Province in southwest China. It is located at the eastern boundary of the Bayankara Massif, near the Aba Tibetan and Qiang Autonomous Prefecture in western Sichuan Province, and resides in the northwest of the Sichuan
Basin, with an overall northeast-southwest trend. Due to the fact that the overall geological conditions of the Longmenshan fault zone are extremely complex and the topography in the region is undulating to a large extent, several strong earthquakes have occurred in the region historically, including the 2008 Wenchuan earthquake and the 2013 Lushan earthquake, of which the 5.12 Wenchuan earthquake is the main representative. Figure 1 shows the geographical location of the study area.

![Figure 1. The location map of the Longmenshan fault zone in Sichuan Province.](image)

2.2. Data Sources

The data sources used in this paper include: 1) a detailed and complete landslide inventory of the Wenchuan earthquake in the region of Longmenshan fault zone, which is used as the basis for spatial selection of positive and negative samples of landslides for the comparison experiments, as well as the accuracy assessment of the combined model for the evaluation results of landslide susceptibility. 2) a digital elevation model of the Longmenshan region with a data resolution of 30m, which is used to extract the slope units as the minimum partition of the study, and calculate the average elevation within the slope units for the establishment of the influence factors database in the study area. 3) The cumulative displacement map of seismic landslides derived from the Newmark model as the basis for selecting the positive and negative sample partition of several statistical learning models. 4) Other data sources used to obtain the distribution information of influence factors such as rivers and faults in the study area.

3. Methodology

The traditional Newmark model and statistical analysis methods based on mathematical models respectively have unique advantages and independently insurmountable disadvantages in the evaluation of seismic landslide susceptibility. In this paper, the Newmark model is used to obtain the permanent cumulative landslide displacement $D_n$ of the slope units in the area, and the study area is divided into two types: landslide hazardous zones and non-hazardous zones. At the same time, two types of positive and
negative training samples are selected in the two types of zones mentioned above, which will later be applied to several statistical analysis models: OSCVM, SVM and RF model. According to the above methods, the mapping of regional landslide susceptibility is obtained and the accuracy is finally evaluated by the ROC (Receiver Operating Characteristic) curve model. The algorithm flow is shown in figure 2.

3.1. Selection of Landslide Evaluation Units, Landslide Training Samples and Influence Factors

3.1.1. Selection of Landslide Evaluation Unit

A correct selection of model evaluation units is the basis for successful experiments. At present, the selection of evaluation units for earthquake landslide at home and abroad can be mainly divided into raster units, slope units and geomorphological units [12]. The slope unit based on geomorphology can effectively reflect the topographic relief and water system characteristics of a single region, and landslide itself is a geological hazardous phenomenon developed at the slope location. Therefore, it is more reasonable to choose the slope unit as the evaluation unit for landslide susceptibility research [13]. The current methods of dividing slope units based on high-resolution remote sensing images include hydro-logical analysis method [14] and surface curvature watershed method, etc. However, the selection of threshold parameters of

![Figure 2. The flow chart of earthquake landslide susceptibility evaluation.](image-url)
these methods relies on empirical determination, which leads to the poor overall generalization ability. Meanwhile, the results obtained after automatic computer step-by-step extraction are poorly partitioned and require more refined manual finishing, which is inefficient to utilize. Therefore, these several methods are often applied in limited study area with less expected partitioning. Considering the wide distribution of the Longmenshan fault zone, the R.Slopeunits method based on Linux environment and GRASSGIS software is adopted to obtain slope units [15], which is a fully automated iterative extraction of slope units under the condition of setting parameters. The method can realize the re-division and unit merging for slope units with too large or too small area after partitioning, and is suitable for the extraction in an expected large study area. A total of about 11,000 partitions are obtained by the R.Slopeunits method, and parts of the partition results are shown in figure 3.

Figure 3. Parts of the partition results of the slope units in the study area.

3.1.2. Selection of Landslide Training Samples

To successfully establish a statistical analysis model for the evaluation of landslide susceptibility, it is necessary to reasonably select samples of landslide hazardous and non-hazardous zones within the study area, thus replacing landslide inventory as two types of training samples for different machine learning models in this experiment. In this paper, the partition statistics are based on the permanent cumulative displacement $D_n$ of the landslide derived from the Newmark model; when the cumulative displacement $D_n$ is higher than the selected empirical threshold, the single slope unit is considered as a part of landslide hazardous zones, and vice versa as a part of landslide non-hazardous zones. By means of empirical threshold partition, about 4400 pieces of landslide hazardous zones are obtained based on the slope units for the selection of positive samples of landslide hazardous zones. In the problem of training sample selection for several types of statistical analysis models, the following approach is chosen.

1) Randomly select the positive samples applied to the SVM and RF model in the slope units of landslide hazardous zones derived from the Newmark model, the number of which is 70% of the landslide hazardous zones in the Newmark model (about 3000 pieces of landslide hazardous zones), and randomly select about an equal number of negative samples in the corresponding landslide non-hazardous zones, and use these two types of samples as the statistical learning model's positive and negative training samples.
2) To test the single-class sample selection area of OCSVM model in the evaluation of landslide susceptibility, this paper not only considers randomly selecting positive samples of OCSVM model in the landslide hazardous zones to obtain the evaluation results of regional landslide susceptibility, but also considers selecting positive samples of OCSVM model in the non-hazardous zones in the same way to obtain the evaluation results of the non-hazard zones of regional landslides. Unlike the other parts of the above-mentioned studies in which landslide susceptibility is used as a research index, the process of randomly selecting positive samples of the OCSVM model in landslide non-hazardous zones is equivalent to viewing landslide non-susceptibility as a research index of this method. These two different methods of sample selection enable a more comprehensive overall evaluation of the study area using a single-class classifier OCSVM model, and also enable an analysis of the effect of the selection of positive and negative landslide sample spaces on the accuracy of the results.

3) To compare the impact of this partitioning result with the partitioning result of the complete landslide inventory on the accuracy of the model evaluation, the complete landslide inventory is used to determine whether landslides have occurred in a single slope unit zone based on the partitioning statistics of the slope units in the study area. A total of 5,500 pieces of landslide area are obtained. Based on this, positive and negative training samples are selected in the same way and proportion in the landslide and non-landslide zones, and compared with the above study as the basis for the accuracy evaluation of the above study methods.

3.1.3. Selection of Landslide Influence Factors

Scientific selection and reasonable classification of the influencing factors is an important basis for applying statistical analysis models to the study of landslide susceptibility. Considering the complex development and occurrence process of seismic landslides, the comprehensive influence of regional geological topography needs to be considered. In this paper, a total of six influence factors, including distance to the fault, distance to rivers, DEM, slope, PGA (Peak Ground Acceleration) and TWI (Terrain Wetting Index), are selected and incorporated into the evaluation system. The state classification and reclassification of the above-mentioned influence factors are carried out according to the actual situation of the study area, and the classification results are shown in figure 4.
4. Results and Analysis

4.1. Results

The predicted probabilities of landslides from the OCSVM, SVM and RF models are used as the seismic landslide susceptibility indices. Meanwhile, the results of the Newmark model and the slope unit data of the remaining 30% of landslide hazardous zones classified in the complete landslide catalog are used as test data to test the three models. In order to evaluate the regional landslide susceptibility more intuitively, the natural break-point classification method is used to classify the study area into five categories based on the prediction results of the statistical analysis models: very high susceptibility, high susceptibility, medium susceptibility, low susceptibility, and very low susceptibility. The ROC curve and the corresponding AUC (Area Under Curves) curve models related to the prediction results of the three statistical learning models are calculated based on the existing complete landslide inventory of the Wenchuan earthquake. The predicted landslide susceptibility results of the three models based on the Newmark model results are shown in figure 5, and the results based on the complete landslide inventory are shown in figure 6, and the corresponding ROC curve evaluation results are shown in figure 7.
Figure 5. Mapping of the landslide susceptibility: a) OCSVM model result with the hazardous zones of the Newmark model as positive samples; b) OCSVM model result with the non-hazardous zones of the Newmark model as positive samples; c) SVM model result with the sample partitioning results of the Newmark model; d) RF model result with the sample partitioning results of the Newmark model.

Figure 6. Mapping of the landslide susceptibility: a) OCSVM model result with the landslide zones of the complete landslide inventory as positive samples; b) OCSVM model result with the non-landslide of the complete landslide inventory as positive samples; c) SVM model result with the sample partitioning results of the complete landslide inventory; d) RF model result with the sample partitioning results of the complete landslide inventory.
Figure 7. Evaluation results of the ROC curve model: a) OCSVM and SVM models taking the Newmark model as the basis for the selection of positive and negative samples; b) OCSVM and SVM models taking the complete landslide inventory as the basis for the selection of positive and negative samples; c) RF model based on two kinds of positive and negative sample selection criteria.

4.2. Analysis

As can be seen from figures 5, 6, and 7, the three different statistical learning models and the different selection of positive and negative sample spaces can lead to some differences in the evaluation results and accuracy of landslide susceptibility.

1) From figure 5 and figure 7, it can be seen that the method of partitioning the slope units in the study area into hazardous and non-hazardous zones based on the results of the Newmark model, and using the two types of area as the basis for sample space selection for several types of statistical analysis models, is better than the analysis results of the Newmark model in terms of overall accuracy. Among them, the RF model has the highest evaluation accuracy, with an AUC value of 0.8777; in contrast, the OCSVM model that takes the non-hazardous zones of the Newmark model as its positive sample selection space performs the worst in this respect. Therefore, it can be concluded that the overall model proposed in this paper has better performance in landslide susceptibility prediction as a whole. The evaluation of the statistical analysis models fusing the Newmark model not only can be independent on the complete or incomplete landslide inventory derived from the post-earthquake studies, but also can achieve the prediction of regional landslide susceptibility. It can also show better evaluation results in terms of overall accuracy, which can overcome the problem of unstable evaluation results and accuracy of the traditional Newmark model to a certain extent.

2) Compared with figure 5(a) and figure 5(b), the two differ in the single-class positive sample selection space of the OCSVM model. The distribution area of the hazardous zones (i.e., the set of area of medium-prone, high-prone and very high-prone zones) of the OCSVM model in figure 5(a) is larger, while the distribution area of the model hazardous zones in figure 5(b) is smaller, and the medium-prone and high-prone zones are mostly located at the edge of the study area, which is different from the susceptibility evaluation results of several other types of models and sample selection methods. Combining figure 7(a) and figure 7(b), it can be seen that the accuracy of the OCSVM model method with the landslide non-hazardous zones as the positive sample selection space is limited and slightly lower than the analysis results of the independent Newmark model in the landslide susceptibility evaluation problem. This is also confirmed by the accuracy evaluation results in figure 6(a) and figure 6(b), which differ in the positive sample space selection for the complete landslide inventory. Therefore, when a single-class sample-based statistical analysis model is used to conduct a study
in the field of landslide susceptibility, the selection space of the positive samples of the model should be consistent with the indicators to be obtained for the study objective: when a landslide susceptibility study is conducted, the selection space of the positive sample of the model should be the landslide hazardous zones.

3) The landslide susceptibility partitioning results of the Newmark model are not a well-established substitute for the complete landslide inventory as the positive and negative sample selection space of the statistical analysis models. Combined with the accuracy of the ROC curve evaluation results in figure 7, although the Newmark model results show a bit of advantage of result accuracy on a single-class classifier like the OCSVM model, however, the accuracy limitation of this method is more obvious when combining the SVM model and RF model considering positive and negative samples. The accuracy difference between the SVM model evaluation of the two positive and negative sample space selection methods reaches 0.11, and that of the RF model is about 0.08. Therefore, in the statistical analysis models considering positive and negative samples, using the complete landslide inventory as the basis of positive and negative sample space selection can obtain the evaluation results of regional landslide susceptibility with higher evaluation accuracy, while the method proposed in this paper can achieve the advantage of accuracy over the Newmark model alone to a certain extent, but it cannot be the alternative to the complete landslide inventory.

5. Conclusion and Discussion

In this paper, a total of six influence factors, namely distance to the fault, distance to rivers, DEM, slope, PGA and TWI, are selected as the landslide influence factors for this experiment to carry out the evaluation of seismic landslide susceptibility. The slope unit in terrain evaluation units is selected as the data partitioning unit for this experiment, and the slope unit partitioning results with better partitioning results are obtained by the method of automatic computer extraction. Since the output of complete landslide inventory often requires a lot of internal and external work after the earthquake, in the absence of data related to it, the Newmark model evaluation results are creatively selected as the basis for positive and negative sample space selection of various statistical learning models. Then, the landslide susceptibility evaluation is carried out by various positive and negative sample selection methods, and the evaluation accuracy of several statistical learning models is compared. The results show that the fused model evaluation results are better than the traditional Newmark model alone for landslide susceptibility evaluation in terms of overall accuracy when the Newmark model evaluation results are used as the basis for partitioning the positive and negative sample selection space of statistical analysis models. In summary, the method proposed in this paper can effectively overcome the disadvantages of limited accuracy and instability of the traditional Newmark model evaluation, and has the advantages of stable overall accuracy and high prediction efficiency. At the same time, the method propose does not depend on the establishment and acquisition of a complete landslide inventory, and can achieve rapid evaluation of regional earthquake landslide susceptibility, which can provide an important reference basis for the prevention and control of potential landslide disasters before earthquakes, post-earthquake emergency relief and reconstruction.
Acknowledgement

This work was supported in part by the National Key R&D Program of China under Grant 2018YFC15035, in part by the Innovation-Driven Project of Central South University under Grant 2020CX036, and in part by the Fundamental Research Funds for the Central Universities of Central South University under Grant 2020ZZTS179 and 2021ZZTS0846.

References

[1] Huang RQ, Li WL. Post-earthquake landsliding and long-term impacts in the Wenchuan earthquake area. Engineering Geology. 2014 Jul; 182: 111-120.
[2] Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F. A review of statistically-based landslide susceptibility models. Earth-Science Reviews. 2018 May; 180:60-91.
[3] Xu Ch, Xu XW, Dai FCh, Saraf AK. Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan Earthquake in China. Computers and Geosciences. 2012 Sept; 46: 317-329.
[4] Ma SY, Xu Ch. Applicability of two newmark models in the assessment of coseismic landslide hazard and estimation of slope-failure probability: An example of the 2008 Wenchuan M_w 7.9 Earthquake Affected Area. Journal of Earth Science. 2019 Oct; 30(5): 1020-1030.
[5] Ma SY, Xu Ch. Assessment of co-seismic landslide hazard using the Newmark model and statistical analyses: a case study of the 2013 Lushan, China, Mw6.6 earthquake. Natural Hazards. 2019 Mar; 96(1).
[6] Gallen SF, Clark MK, Godt JW, Roback K, Niemi NA. Application and evaluation of a rapid response earthquake-triggered landslide model to the 25 April 2015 M w 7.8 Gorkha earthquake, Nepal. Tectonophysics. 2017 Sept; 714-715.
[7] Yue XL, Wu ShH, Yin YH, Gao JB, Zheng JY. Risk identification of seismic landslides by joint newmark and rockfall analyst models: A case study of roads affected by the Jiuzhaigou Earthquake. International Journal of Disaster Risk Science. 2018 Sept; 9(3): 392-406.
[8] Yi YN, Zhang ZH, Zhang WC, Xu Q, Deng C, Li QL. GIS-based earthquake-triggered-landslide susceptibility mapping with an integrated weighted index model in Jiuzhaigou region of Sichuan Province, China. Natural Hazards and Earth System Sciences. 2019 Sept; 19(9): 1-25.
[9] Xu Ch, Xu XW, Dai FCh, Wu ZhD, He HL, Shi F, Wu XY, Xu SN. Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake of China. Natural Hazards. 2013 Sept; 68(2): 883-900.
[10] Wang Q, Wang Y, Niu RQ, Peng L. Integration of information theory, K-Means cluster analysis and the logistic regression model for landslide susceptibility mapping in the Three Gorges Area, China. Remote Sensing. 2017 Sept; 9(9): 938.
[11] Lin L, Lin QG, Wang Y. Landslide susceptibility mapping on a global scale using the method of logistic regression. Natural Hazards and Earth System Sciences. 2017 Sept; 17(8): 1-20.
[12] Guzzetti F, Carrara A, Cardinali M, Reichenbach P. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology. 1999 Jun; 31(1): 181-216.
[13] Erener A, Düzgün HSB. Landslide susceptibility assessment: What are the effects of mapping unit and mapping method?. Environmental Earth Sciences. 2012 Jun; 66(3): 859-877.
[14] Xie MW, Esaki T, Zhou GY. GIS-based probabilistic mapping of landslide hazard using a three-dimensional deterministic model. Natural Hazards. 2004 Oct; 33(2): 265-282.
[15] Alvioli M, Marchesini I, Reichenbach P, Rossi M, et al. Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling. Geoscientific Model Development. 2016 Nov; 9(11): 3975.