The probability of coseismic landslides in different peak ground accelerations: a case of 1927 Gulang M8.0 event

Jixiang Xu\textsuperscript{1,2}, Chong Xu\textsuperscript{1,3}\textsuperscript{*}

1. National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China;
2. School of Engineering and Technology, China University of Geosciences (Beijing), Beijing 100083, China;
3. Key Laboratory of Compound and Chained Natural Hazards Dynamics (Under Construction), Ministry of Emergency Management of China, Beijing 100085, China;

*Corresponding to Chong Xu (chongxu@ninhm.ac.cn, http://orcid.org/0000-0002-3956-4925)

Abstract. A catastrophic Ms 8.0 earthquake struck Gulang County, Gansu Province on May 23, 1927, and triggered numerous coseismic landslides. Based on a high-resolution remote sensing image from Google Earth Platform, we delineated 936 coseismic landslides with an area of 58.6 km\textsuperscript{2}. Further, ten factors were selected as the impact factors for earthquake-triggered landslide hazard mapping of the earthquake. We propose a sampling method that selects the sliding and nonsliding samples established on the ratio of the study area to the landslide area based on the Bayesian theory, and a total of 247080 samples are selected. We establish a multi-factor impact model for landslide probabilities using the logistic regression (LR) model, calculate the regression coefficients based on LR models, and obtain the occurrence probability of coseismic landslides. The model is then applied to calculate the probability of earthquake landslides under different PGAs. The study areas are divided into extremely high-risk, high-risk, medium-risk, low-risk, and extremely low-risk areas. The results of this study provide scientific references for mitigating earthquake-related landslides in earthquake-prone areas.

Key words: Gulang earthquake, Probability of landslides, Seismic landslide hazard assessment

1. Introduction

The population is becoming increasingly concentrated as the economy and society develop, major and lifeline projects become more intense, and the sensitivity and vulnerability to earthquake disasters
grow. Thus, the risk of earthquake disasters cannot be ignored. On May 23, 1927, an earthquake of magnitude 8 occurred near Gulang County, Gansu Province. The seismic intensity reached XI degree, and the earthquake hypocenters macroscopical depth was 12 km\[1\]. About 40,000 people were killed and injured by the earthquake, and caused numerous landslides, collapses, and other geological disasters. Some of these landslides resulted in significant casualties and property losses. The Dengshanzhuang landslide, for example, buried an entire village and killed 104 people\[2\].

Strong earthquakes occur frequently in the northeastern margin of the Tibetan Plateau, with severe coseismic landslide disasters. Historically, the coseismic landslide of the 1718 Tongwei M7.5 earthquake resulted in a considerable number of deaths\[3\]. Nearly half of the more than 200,000 casualties in the 1920 Haiyuan M8.5 earthquake was caused directly by coseismic loess landslide\[4\]. Further, two remote loess landslides triggered by the Minxian Ms6.7 earthquake in 2013 resulted in at least 12 deaths\[5\]. The Gulang earthquake was also one of the most destructive earthquakes on the Tibetan Plateau’s northeastern margin. In addition, the Qilian-Haiyuan fault, which runs between the Haiyuan and Gulang earthquake source areas, with a length of 280 km has no historical records of strong earthquakes in the past 800 years. This area is called the Tianzhu earthquake void zone and is considered the most vulnerable to future severe earthquakes\[6\]. This study applies modern remote sensing and geographic information system technology to the risk assessment of an earthquake-triggered landslide. The findings of this study have significant practical application value and can provide a scientific reference for future disaster prevention and mitigation in this area, as well as theoretical support for some major projects.

2. Study area
This study focuses on the 1927 Gulang earthquake high-intensity region (VIII–XI), with a total area of 12354.1 km\(^2\), located in the Wuwei City, Gansu Province. Wuwei is a prefecture-level city in the central and western parts of Gansu Province, located in the eastern section of the Hexi Corridor. The elevation in the study area ranges from 1420–4586 m, and the terrain is generally high in the south and west and low in the north and east. There are 8 active faults with strong tectonic activity and a series of NNE trending drainage systems in this study area (Fig. 1).

3. Data and methods
3.1. Seismic landslide data
Based on the Google Earth platform’s high-resolution satellite remote sensing image, this study adopts the method of visual interpretation to conduct landslide visual interpretation in the Gulang earthquake high-intensity area. A total of 1516 landslides are interpreted, including about 600 glacial freeze-thaw landslides, which are not sure were triggered by the 1927 Gulang earthquake. Therefore, this study excludes such landslides, and 936 landslides are finally determined (Fig. 2). The scale of these landslides is mainly small and medium-sized. The total landslide area is 58.55 km\(^2\), the density of landslide point is 0.075 km\(^{-2}\), and the percentage of landslide area is 0.47%.
Figure 1. Active tectonic map of the area of the Gulang earthquake.

Faults: F1: The fault along the southern margin of the Wuwei Basin; F2: Huangcheng–Shuangta fault; F3: Wuwei–Tianzhu buried fault; F4: Gulang–Shuangta buried fault; F5: Lenglongling fault; F6: Gulang fault; F7: Jinqianghe fault; F8: Maomaoshan–Laohushan fault

Locations: DSZ: Dengshanzhuang; XSP: Xishanpu; HYC: Huangyangchuan

3.2. A sample selection method based on the Bayesian thinking

The coseismic landslide probability of a single seismic event is the ratio of the total area of all landslides to the size of the study area based on the Bayesian probability theory, which is the prior probability of the occurrence of the coseismic landslide corresponding to the seismic event. Thus, the probability of the occurrence of a coseismic landslide in a region ($P_{cols}$) can be defined as the total area of the coseismic landslide in the region divided by the total area of the region.

Therefore, the selection of training samples should ensure that the proportion of landslide and nonslip samples is consistent with the proportion of landslide and nonlandslide area in the actual area. In most prior investigations of this model, the percentage of sliding and nonsliding samples was equal. This sampling method artificially exaggerated the proportion of landslide samples in the study area, and such evaluation results only considered the relative risk, not the true probability of landslide occurrence, resulting in the probability generated by the model overestimating the actual probability of landslide occurrence[7]. Thus, we expect the sliding sample generation probability to be proportional to the spatial range.

Sample points are randomly generated in the study area; the points falling into the landslide area are sliding samples, whereas the rest are nonsliding samples. This ensures that the ratio of sliding to nonsliding in the study area is equal to the probability of a coseismic landslide. After experimenting with various sampling intensities, 20 points km$^{-2}$ was chosen, and the minimum distance between each sample point was 100 m, and a total of 247,080 samples were selected in the study area. The total area of the landslide was 58.55 km$^2$, the study area was 12354.1 km$^2$, and the probability of a landslide associated with this earthquake ($P_{cols}$) was about 0.47%. For the entire training sample database, the sliding sample was 1163, the nonsliding sample was 245917, and the ratio of landslide to nonlandslide
was equal to the landslide probability ($P_{cls}$), which is also 0.47%. This indicates that the sampling intensity of 20 km$^{-2}$ satisfies the requirements.

### 3.3. Logistic regression model

The logistic regression (LR) model, which is one of the most widely used machine learning models, is the mathematical model used in this study. To predict the probability of an event occurring in a specified area, LR analysis mainly forms a multiple regression relationship between a dependent variable (whether an event occurs) and multiple independent variables (impact factors). The dependent variable in this study represents the occurrence of coseismic landslides. Its values of 1 and 0 represent the occurrence and absence of coseismic landslides, respectively. The independent variable is each impact factor. The relationship between the probability of landslide occurrence and impact factors can be expressed as follows:

$$Z = B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n$$  \hspace{1cm} (1)

$$P = \frac{1}{1 + e^{-Z}}$$  \hspace{1cm} (2)

Where $P$ is the probability of landslide forecast, the output range of 0–1; $Z$ represents the sum of linear weights after the superposition of factors; $B_i$ is the regression coefficient of each factor; $X_i$ represents each impact factor.

#### Table 1. Influence factors and their classes.

| Influence factor               | Reclassification standard                                      |
|-------------------------------|----------------------------------------------------------------|
| Elevation/m                   | 1.1420–2000; 2.2000–2400; 3.2400–2800; 4.2800–3200; 5.>3200      |
| Slope/$^\circ$                | 1. $<5^\circ$; 2. $5^\circ$–$10^\circ$; 3. $10^\circ$–$15^\circ$; 4. $15^\circ$–$20^\circ$; 5. $20^\circ$–$25^\circ$; 6. $25^\circ$–$30^\circ$; 7. $30^\circ$–$35^\circ$; 8. $35^\circ$–$40^\circ$; 9. $40^\circ$–$45^\circ$; 10. $45^\circ$–$50^\circ$; 11. $>50^\circ$ |
| Aspect                        | 5.SE($112.5^\circ$–$157.5^\circ$); 6.S($157.5^\circ$–$202.5^\circ$); 7.SW($202.5^\circ$–$247.5^\circ$); 8.W($247.5^\circ$–$292.5^\circ$); 9.NW($292.5^\circ$–$347.5^\circ$) |
| TPI                           | 1.Ridge; 2.Upper slope; 3.Middle slope; 4.Flat; 5.Lower slope; 6.valley |
| Distance from river/km        | 1. $<1$; 2. $1.1$–$2$; 3. $2.2$–$3$; 4. $3.3$–$4$; 5. $4.4$–$5$; 6. $5.5$–$6$; 7. $>6$ |
| Distance from the epicenter/km | 1. $<5.2.5$–$10$; 3. $10.1$–$15$; 4. $15.2$–$20$; 5. $20.25$–$25$; 6. $25.3$–$30$; 7. $30.35$; $8.35$–$40.9$ |
| Seismic intensity             | 1.VIII; 2.IX; 3.X; 4.XI |
| Strata age                    | 1.O; 2.N; 3.E; 4.K; 5.J; 6.T; 7.P; 8.C; 9.D; 10.S; 11.O; 12.E |
| PGA                           | 1.0.1g; 2.0.2g |

Ten factors, including elevation, slope, slope aspect, TPI, distance from rivers, stratigraphic age, distance from the epicenter, seismic intensity, distance from faults, and peak ground acceleration are included in the model after comprehensive consideration of various influencing factors of Gulang earthquake landslide. The topographic data of this study’s impact factor analysis come from the ASTER GDEM V1 digital elevation model (DEM) with a spatial resolution of one arc second in the Geospatial Data Cloud Website (http://www.gscloud.cn). We further process it into 20 m resolution DEM data with ArcGIS software to increase the accuracy. The DEM data are used to calculate topographic factors such as TPI, slope, slope aspect, and river. The strata age data come from the...
1:500,000 national geological map published by the China Geological Survey. PGA data are obtained from the U.S. Geological Survey (https://www.usgs.gov/). The fault information is obtained from Active Tectonics Map of China. The results of the reclassification of impact factors are shown in Table 1.

4. Results and analysis

4.1. LR coefficient calculation

The 247080 training samples and impact factors were combined, resulting in sample classification information for each impact factor, then the training samples were imported into SPSS software for LR analysis, yielding the regression coefficients of each grade under each impact factor. The regression coefficient indicates the size of the impact factor in this classification corresponding to the dangerous degree or importance. Finally, the regression coefficient of each grade under all ten impact factors (Fig. 3) was obtained, as well as its constant value, $B_0$, as $-24.663$, when the regression coefficient value of the last grade of each factor was set as 0.
4.2. Seismic landslide hazard assessment map

A corresponding weight value was assigned to each grade under each impact factor based on the regression coefficient of the LR model obtained from SPSS software. Based on Formulas 1 and 2, each influence factor layer was overlaid on the ArcGIS platform, and the distribution map of landslide risk index in the area of the Gulang earthquake was finally obtained (Fig. 4).

![Figure 4. Landslide risk index map in the area of the Gulang earthquake.](image)

Landslide hazard index was divided into 1 grade every 0.2, and the study area was divided into the extremely low-risk area (0–0.2), low-risk area (0.2–0.4), medium-risk area (0.4–0.6), high-risk area (0.6–0.8) and extremely high-risk area (0.8–1). Figure 5 shows the landslide risk assessment map in the area of the Gulang earthquake.

![Figure 5. Landslide risk assessment map in the area of the Gulang earthquake.](image)
in the Gulang earthquake area. The figure shows the extremely dangerous areas for landslides near Dengshanzhuang and the 25-km long section between Xishanpu and Huangyangchuan, as well as the actual intensive areas of landslides. This indicates that the Gulang earthquake’s landslide intensive area remains a high-risk area for large-scale landslides and collapse in the future.

4.3. Seismic landslide risk at a given PGA value

The landslide occurrence probability under different PGA in the study area was predicted in this research to better deal with the earthquake disaster with different damage degrees, and the PGA was
set as the trigger factor and substituted into the existing model. Figure 6 shows the landslide risk assessment map in the study area when the PGA was at five levels: 0.1, 0.2, 0.5, 0.8, and 1.0 g. Similarly, the study area was divided into five levels: extremely low-risk, low-risk, medium-risk, high-risk, and extremely high-risk areas, based on the landslide risk index.

5. Conclusions
We selected 247,080 sample points in the study area and ten impact factors for LR analysis based on the Bayesian probability theory and real landslide data from the Gulang earthquake. We conducted probability modeling of coseismic landslide occurrence and obtained a landslide risk assessment map in the Gulang earthquake area. Finally, the occurrence probabilities of coseismic landslides in this area under different peak ground accelerations (0.1, 0.2, 0.5, 0.8, and 1.0 g) were obtained. The findings can help prevent and mitigate landslide disasters, as well as provide a technical and methodological reference for future study in other areas.

Acknowledgments
This study was supported by the National Key Research and Development Program of China (2018YFC1504703).

References
[1] Gu G X 1983 China Earthquake Catalogue (1983 BC–1969 AD) (Science Press, Beijing)
[2] Zou J C, Shao S M and Jiang R F 1994 Distribution and tectonic implications on the Gulang seismic landslide Earthquake Research In China 2 168-174
[3] Sun P, Li R J, Jiang H, Igwe O and Shi J S 2017 Earthquake-triggered landslides by the 1718 Tongwei earthquake in Gansu Province, northwest China Bulletin of Engineering Geology and the Environment 76 1281-95
[4] Earthquake Team of Ningxia Hui Autonomous Region 1980 The Great Haiyuan Earthquake in 1920 (China Earthquake Administration Lanzhou Institute of Seismology)
[5] Tian Y Y, Xu C, Xu X W and Chen J 2016 Detailed inventory mapping and spatial analyses to landslides induced by the 2013 Ms 6.6 Minxian earthquake of China Journal of Earth Science 27 1016-26
[6] Gao W 2018 Late quaternary activity of the Tianqiaogou-Huangyangchuan fault: Implication for the Tectonic Movement mechanism at the Northeastern Tibet Institute of Geology, China Earthquake Administrator
[7] Jessee M A N, Hamburger M W, Allstadt K, Wald D J, Robeson S M and Tanyas H 2018 A global empirical model for near-real-time assessment of seismically induced landslides Journal of Geophysical Research: Earth Surface 123 1835-59