A GIS-based statistical model for rapid landslide susceptibility mapping in the Beichuan-Pingwu area, Sichuan, China

Y Chen, Q J Wang
Key Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China

E-mail: yuchen@ceode.ac.cn

Abstract. The 2008 Wenchuan earthquake, with a magnitude of Mw 8.0, induced numerous landslides. Remote sensing planes were sent out to take high resolution aerial photographs, from which the geologic hazards could be instantly interpreted. However, aerial images covering all of the study area could not be obtained in a short time because of the limitations of the planes and the influence of weather conditions. This study establishes a statistical model based on the landslide interpretation results of one photographic strip inside the Beichuan-Pingwu area. It has strong applicability and can be applied to other places without such data. Finally, we produced a landslide susceptibility map, which provides scientific support for the instant evaluation of disaster information and post-disaster reconstruction.

1. Introduction
The Wenchuan earthquake, measured at Ms 8.0 according to the China Earthquake Administration, occurred at 14:28 on 12 May 2008 in Sichuan Province of southwestern China. It brought overwhelming destruction to eight provinces and cities. Landslides triggered by the earthquake, which disrupted transportation, power, and communications, were distributed along the fault rupture zone and river channels. Remote sensing planes were sent out to take high resolution aerial photographs, from which the geologic hazards could be instantly interpreted. However, aerial images covering all of the study area could not be obtained in a short time because of the limitations of the planes and the influence of weather conditions.

In this study, we established a statistical model based on the landslide interpretation results of one airborne photographic strip inside the Beichuan-Pingwu area. It has strong applicability and can be applied to other places without such data. Finally, we produced a landslide susceptibility map, which provides scientific support for the instant evaluation of disaster information and post-disaster reconstruction.

2. The study area
Our study area, with a total area of 4,630 km², is situated 167 km northeast of the earthquake's epicenter (Figure 1). The area includes a part of Beichuan County, Jiangyou County, and Pingwu County in Sichuan Province. The study area is situated in the transitional mountainous belt between the Sichuan Basin and the Western Sichuan Plateau, characterized by rugged mountains with elevations between 300 m and 4,700 m as well as deeply incised valleys. The epicenter of this strong earthquake was just in the medium to high mountains west of the Sichuan Basin, where the geological environment...
is quite fragile, hence numerous geo-hazards were triggered, including slope collapses, debris flows, and landslides\cite{1}.

Aerial photographs acquired on 28 May 2008 inside the Beichuan-Pingwu area were used for landslide interpretation because of their high 0.5-m spatial resolution. The landslide interpretation result is shown as Figure 2. The extent covered by the airborne photographic strip was used to establish a GIS-based statistical model.

3. Methods

A variety of approaches have been used for landslide susceptibility mapping, and they can be classified into qualitative and quantitative methods. Most qualitative methods tend to be subjective since they depend on expert opinions and portray hazard levels in descriptive terms. Quantitative methods are based on the numerical expression of the relationship between instability factors and landslides, which can be divided into deterministic and statistical\cite{2}. Deterministic methods depend on engineering principles of slope instability, expressed in the factor of safety, and are therefore mostly used for individual slope limited to a local site\cite{3, 4}. Statistical methods based on the assumption that the past occurrence of landslides in a specific site is indicative of the potential for future landslides to occur in sites with similar characteristics have been widely used in regional scale landslide susceptibility mapping\cite{5, 6, 7}.

The most important step in forming a statistical model is to establish relationships between variables based on statistics of mapped landslide inventory data. These variables refer to factors leading to landslides and include continuous data (such as altitude, slope angle, and slope aspect) and discrete data (such as lithology). Most previous statistical models transformed continuous factor data into

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Location of the study area (yellow box).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Landslide interpretation result from aerial photographs acquired on 28 May 2008.}
\end{figure}
categorized classes\(^8\). Such conversions inevitably result in loss of information. In this study, a GIS-based statistical model was proposed based on single factor landslide susceptibility curve fitting and multivariate logistic regression. The data processing is shown below.

3.1. **Choosing evaluation factors**
Landslide occurrence depends on complex interactions among a large number of factors. As shown in Figure 3, six factors influencing landslide occurrence, including altitude, slope, aspect, distance from fault, lithology, and normalized difference vegetation index (NDVI), were extracted in this study from digital terrain models (DTM), remote sensing images, and geologic data.

![Figure 3. The six evaluation factors for landslide prediction are altitude (A), slope (B), aspect (C), distance from fault (D), lithology (E), and NDVI (F).](image)

3.2. **Calculating the susceptibility of each single factor**
For continuous evaluation factors such as altitude, slope, aspect, distance from fault, and NDVI, a curve fitting method to fit the landslide ratio and each single factor was used. The curve fitting function was chosen depending on how the factor influences the landslide. For instance, a quadratic or cubic function was selected to fit the slope, aspect, and altitude, whereas a logarithmic or linear function was used to fit the distance from fault and NDVI. Some factors were abandoned once the precision of curve fitting was low, for example, the altitude in this study. For a discrete evaluation factor of lithology, the landslide ratio of each lithology was calculated directly.

Take the continuous factor slope angle as an example. First, slope angle was divided into more than 32 classes between 0 and 90 degrees as its value range. Second, the landslide frequency of each class was calculated. Third, several kinds of profiles such as linear, logarithmic, quadratic, and cubic curves were used to fit the landslide frequency. The curve fitting result is shown in Table 1. The best curve fitting equation was chosen to calculate the landslide susceptibility against slope angle. In this study, the cubic equation was chosen because its R-squared reached 0.76, which was much more than the R-squared of the quadratic equation. So the function below was used to calculate the landslide susceptibility against slope angle. The result is shown as Figure 4. The locations where landslides have occurred fit well with the high susceptibility value (red color shown in Figure 4).

\[
S_{slope} = 0.004 + 2.54 \times 10^{-5} \cdot slope \cdot slope - 2.49 \times 10^{-7} \cdot slope \cdot slope \cdot slope
\]
Table 1. Curve fitting coefficients of landslide frequency (percentage) against slope (degree).

| Equation   | Model Summary | Parameter Estimates |
|------------|---------------|---------------------|
|            | R-squared     | F       | df1 | df2 | Sig.  | Constant | b1      | b2      | b3      |
| Linear     | 0.126         | 4.048   | 1   | 28  | 0.054 | 0.003    | 8.73E-05|         |         |
| Logarithmic| 0.174         | 5.898   | 1   | 28  | 0.022 | -0.003   | 0.003   |         |         |
| Quadratic  | 0.453         | 11.168  | 2   | 27  | 0     | -0.005   | 0.001   | -5.81E-06|         |
| Cubic      | 0.76          | 27.475  | 3   | 26  | 0     | 0.004    | 0       | 2.54E-05| -2.49E-07|

Figure 4. The result of landslide susceptibility against slope angle.

3.3. Logistic regression of multiple factors

The principle of logistic regression (LR) rests on the analysis of a problem in which a result measured with dichotomous variables such as 0 and 1 or true and false is determined from one or more independent factors. In this study, LR was used to find the best-fitting model to describe the relationship between the presence or absence of landslides (dependent variable) and a set of independent parameters as susceptibility of each single factor (quantified in Step 2). The region of the modeling area was divided into 20*20-m pixels and there were 5,488 pixels considered to be landslide occurrence, while 5,488 pixels were selected without landslides to be included in LR together with the pixels of landslide occurrence. In order to implement this regression method, SPSS software was employed to calculate the correlation between the landslide and each factor. The LR function was built as follows.

Table 2. LR coefficients of multiple factors.

| Factor      | B      | S.E.   | Wald   | df | Sig. |
|-------------|--------|--------|--------|----|------|
| NDVI (n)    | 2.129  | 0.124  | 294.455| 1  | 0.00 |
| Slope angle(s) | 8.328  | 0.315  | 697.126| 1  | 0.00 |
| Aspect(a)   | 1.316  | 0.059  | 498.678| 1  | 0.00 |
| Fault(f)    | 4.716  | 0.399  | 139.768| 1  | 0.00 |
| Lithology(l) | 0.631  | 0.114  | 30.783 | 1  | 0.00 |
The created landslide susceptibility map was classified into four classes based on the statistic of susceptibility, including very low (0.00-0.17), low (0.18-0.37), high (0.38-0.57), and very high (0.58-1.00). The majority of landslides occurred within areas designated as very high susceptibility. The resultant landslide susceptibility map, together with the landslide inventory, is shown in Figure 5. However, we also observed some landslides in high- to low-susceptibility areas.

4. Model evaluation
To evaluate the accuracy of the result quantitatively, a cumulative proportion diagram was used (Figure 6). It was found that when the most susceptible areas cover 10.48% of the total modeling area the cumulative percent of landslide occurrence is 22.60%. The area under curve (AUC) was calculated to quantitatively represent the prediction accuracy of the proposed model in the area. The area under curve (0.7073) showed good predictive capacity.

5. Discussion and conclusion
The first results of our model were the single factor fitting coefficients of each factor (Table 1), which were useful for influencing factor selection. In this study, the altitude factor was abandoned because of its poor single factor curve fitting accuracy calculated in Step 3.2. The slope angle, aspect, lithology, distance from fault, and NDVI were chosen as influential factors. The second results were the logistic regression model statistics and coefficients (Table 2), which were useful in assessing the accuracy of the regression function and the role of influential factors in the presence or absence of landslides. In this study, the slope angle factor was found to have the strongest relationship with landslides followed by distance to fault and NDVI. Landslides will easily and frequently occur in the quake-hit area in the following years. Large landslide events may block river channels and roads. The methods described in this paper could be used for rapid landslide susceptibility mapping. The predicted susceptibilities generated from the model within the GIS were used to produce a map of relative landslide susceptibility for a broad area.

Acknowledgements
This research was financially supported by the National Science and Technology Support Program (No. 2012BAK15B05), National Natural Science Foundation of China (No. 41240012 and 41171280), and the National Science and Technology Support Program (No. 2012BAH27B05).

[1] Huang R Q and Li W L 2009 Analysis of the geo-hazards triggered by the 12 May 2008 Wenchuan Earthquake, China B Eng Geol Environ 68 363-71
[2] García-Rodríguez, Malpica J A, Benito B and Díaz M 2008 Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression Geomorphology 95 172-91
[3] Ablay G and Hurlimann M 2000 Evolution of the north flank of Tenerife by recurrent giant landslides J Volcanol Geoth Res 103 135-59
[4] Wan Y and Kwong J 2002 Shear strength of soils containing amorphous clay-size materials in a slow-moving landslide Eng Geol 65 293-303
[5] Hong Y, Adler R and Huffman G 2007 Use of satellite remote sensing data in the mapping of global landslide susceptibility Nat Hazards 43 245-56
[6] Lee S and Min K 2001 Statistical analysis of landslide susceptibility at Yongin, Korea Environ Geol 40 1095-113
[7] Wang J and Peng X-g 2009 GIS-based landslide hazard zonation model and its application Procedia Earth and Planetary Science 1 1198-204
[8] Davis J C, Chung C-J and Ohlmacher G C 2006 Two models for evaluating landslide hazards Comput Geosci-Uk 32 1120-7
[9] Ayalew L and Yamagishi H 2005 The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan Geomorphology 65 15-31