

# A geospatial information quantity model for regional landslide risk assessment

Yumin Tan<sup>1</sup> · Dong Guo<sup>1</sup> · Bo Xu<sup>2</sup>

Received: 29 April 2015 / Accepted: 16 July 2015 / Published online: 2 August 2015  
© Springer Science+Business Media Dordrecht 2015

**Abstract** Using geospatial technologies to assess geological hazard risk has been proved feasible and effective. In this paper, a geospatial information quantity model is proposed to assess landslide risk, which includes nine triggering factors: slope, aspect, cumulative catchment area, formation lithology, seismic intensity, distances to water, precipitation, vegetation, and land use/land cover type, in which the last three triggers are dynamic ones and need to be extracted from up-to-date remote sensing images. These triggering factors are then taken as geospatial information quantities and used to construct an information quantity-based model to assess and predict the landslides in Fuling District, Chongqing City, China, resulting in a risk distribution map. Finally, ROC curve is used to validate the model. With the AUC of success-rate ROC of 0.839 and the AUC of prediction-rate ROC of 0.807, the model is proved reliable to interpret and predict the landslide occurrences in the study area.

**Keywords** Landslide · Risk assessment · Geospatial information quantity model

## 1 Introduction

China, especially the mountainous areas with dense population, is prone to geological hazards, which threatens the safety of human life and property and damages ecological environment (Yin and Zhu 2001). Since its construction, the Three Gorges Dam has undergone frequent geological hazards including landslides, mudslides, breakdowns, unstable rocks, high cut slopes, and collapses due to geological, geomorphological, and

---

✉ Yumin Tan  
tanym@buaa.edu.cn

<sup>1</sup> Department of Civil Engineering, Beihang University, Beijing 100191, China

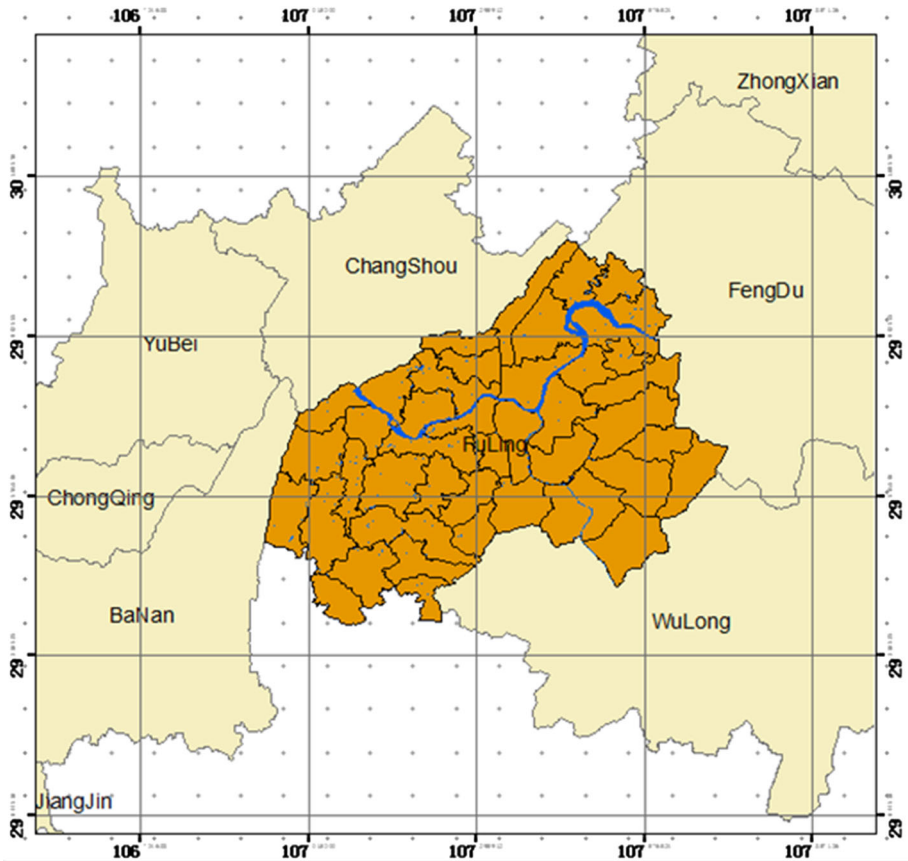
<sup>2</sup> Department of Geography & Environmental Studies, California State University, San Bernardino, CA 92407, USA

climatic reasons, among which landslide is the major hazard type. In recent years, computer technology and spatial information technology represented by geographic information sciences (GIS) and remote sensing have been extensively applied to risk assessment of landslides. The quantitative methods of risk assessment mainly include artificial neural networks (Pradhan and Lee 2010; Yilmaz 2010; Bui et al. 2012; Sdao et al. 2013; Conforti et al. 2014), information quantity-based model (Gao et al. 2006; Niu et al. 2011; Chen et al. 2013; Deng et al. 2014; Gao et al. 2014; Wang et al. 2014), multivariate regression analysis (Lee and Min 2001), and logistic regression analysis (Pradhan and Youssef 2010; Bai et al. 2011; Bui et al. 2011). Among them, the information quantity-based model is gaining increasing attention for its simple implementation and high accuracy. Information quantity model is a general model and can be used in many fields. However, few applications have been done in the field of landslide risk assessment in the past. Some applications in landslide risk assessment are: Yin and Liu (2000) extended the traditional landslides stability analysis to Monte Carlo probability simulation analysis and established a multifactor information analysis system based on information theory to predict hazard occurrences. Yuan and Huang (2001) integrated GIS with information quantity-based model and applied it to the delineation of geological hazard risk areas. Dai (2013) improved the traditional information theory by implementing GIS and remote sensing technologies.

This paper adopts the information quantity-based landslide risk assessment model improved by Dai (2013) but utilizes dynamic triggering factors derived from timely remote sensing images to assess landslide hazards in Fuling District, Chongqing City, China. This is the first time that information quantity model is applied in a small study area in the Three Gorges Dam Reservoir region. If it is proved to be successful, this method will be used for the entire reservoir region in the future and will greatly benefit the preparedness and management of natural hazards in this area. First, the triggering factors of landslide were extracted from different data sources in the study area. Second, hazard risk was predicted using the improved model, and a risk distribution map was generated. Finally, accuracy assessment was performed to test the fit and reliability of the information quantity model.

## 2 Study area

Fuling District, in the hinter land of the Three Gorges Dam Reservoir region, is located in the middle east of Chongqing City and is historically called the ‘gateway’ to the southeast Sichuan Province as it sits where the Yangzi River and Wujiang River converge. It covers an area of more than 2900 square kilometers, with an east–west direction extending 74.5 km and a north–south direction extending 70.8 km (Fig. 1). Fuling District belongs to the subtropical humid climatic with distinct seasons. Summer is hot and wet, while winter is cold and dry. The yearly average temperature is around 18.1 centigrade, and the yearly average rainfall is about 1070 mm, with the purple and mountain yellow soil as two main soil types. The study area is characterized by undulating low hills with elevation ranging from 200 to 800 m. Its relief gradually decreases from southeast to northwest with the maximum and minimum elevations of 1977 m and 138 m above the mean sea level, respectively. Major landform classes include hilly or table lands (54.4 %), low hills (31.3 %), mid hills (13.3 %), and plain (1.2 %). Sitting atop of the old Yantz metaplatform, crust of our study area is very stable. Bedrock exposures are mainly of upper



**Fig. 1** Study area

Paleozoic, Mesozoic, and Cenozoic strata. The major geologic structures are mainly controlled by the north–south trending Sichuan–Guizhou structure belts, which extend northward into the fault belts in the eastern Sichuan. The former structure belts are located in the south to the Yangtze River, including north–south trending and east–west trending folds. Faults are mainly reverse. Major faults in this area include Koushi and Jiaoshiba reverse faults.

In past years, project construction for resettling immigrants, the experimental water recharging of the reservoir, and human activities have exacerbated many geological hazards, bringing huge threats to local human safety and property. For example, more than 200 landslide occurrences were caused by road or house constructions in the study area from 2008 to 2014. After the phase III geological hazard and reservoir bank prevention project, the geological hazard risks have been substantially reduced. However, with the 175-m water storage of the Three Gorges Dam Reservoir, landslides still exist in Fuling District, severely interrupting the area’s stability and sustainable development.

### 3 Methodology and data

#### 3.1 The geospatial information quantity model for landslide risk assessment

In this paper, a geospatial model, based on information theory, was built to predict the probability of landslide occurrences by decreasing the entropy during the development of landslides. It is a statistical prediction model to evaluate the risk of landslide by calculating the information value of each individual triggering factor and overlaying these information values. Higher information value implies greater possibility of landslide occurrences (Wang et al. 2014). The information value was estimated using the probability of existing landslide occurrences as in Eq. (1):

$$I(y, x_1x_2 \dots x_n) = \log_2 \frac{P(y|x_1x_2 \dots x_n)}{P(y)} \tag{1}$$

Equation (1) can also be expressed as Eq. (2):

$$I(y, x_1x_2 \dots x_n) = I(y, x_1) + I_{x_1}(y, x_2) + \dots + I_{x_1x_2 \dots x_n}(y, x_n) \tag{2}$$

where  $I(y, x_1x_2 \dots x_n)$  is the information value provided by the combination of triggering factors  $x_1x_2 \dots x_n$  of landslides;  $P(x_1x_2 \dots x_n)$  is the probability of landslide occurrences under the combination  $x_1x_2 \dots x_n$  of triggering factors; and  $I_{x_1}(y, x_2)$  is the information value provided by triggering factor  $x_2$  with the existence of triggering factor  $x_1$ .

There were various triggering factors causing landslides. By overlaying the individual information values, we obtain the total information value provided by the combined impact of the triggering factors. If we use  $I_i$  to denote the information value of triggering factor  $i$ , we get Eq. (3):

$$I(y, x_1x_2 \dots x_n) = \sum_{i=1}^n I_i \tag{3}$$

The basic unit of regional landslide risk assessment in the study area is grid, that is, all evaluation factors should be assigned to each individual grid. Therefore, the grid size should be carefully defined, as large grid size may cause information confusion, while small grid size will affect efficiency. The appropriate grid size can be obtained from empirical formulas, such as that given by Li and Zhou (2003), which is taken to define the grid size in the paper, as Eq. (4):

$$G_s = 7.49 + 0.0006S - 2.0 \times 10^{-9}S^2 + 2.9 \times 10^{-15}S^3 \tag{4}$$

where  $G_s$  is the size of suitable grid, and  $S$  is the denominator of contour line scale (in this paper,  $S$  is 50,000).

In practice, frequency was used to represent probability as in Eq. (5):

$$I(y, x_1x_2 \dots x_n) = \sum_{i=1}^n I_i = \sum_{i=1}^n \log_2 \frac{N_i/N}{S_i/S} \tag{5}$$

where  $S$  is the total area of an evaluation unit;  $N$  is the total number of landslide sites in the study area;  $S_i$  is the area that triggering factor  $i$  occupies in the evaluation unit; and  $N_i$  is the number of landslide sites caused by triggering factor  $i$ .

### 3.2 Derivation of triggering factors of landslides

Previous research in the study area indicated that the major triggering factors were landforms, rock and soil types, geological structure, precipitation, vegetation, human activities, hydrogeological conditions, and tectonic belts, etc. (Ren et al. 2011; Wang et al. 2012; Dieu et al. 2012). Based on field work and comprehensive analysis of the spatial relationship between triggering factors and landslide, nine triggering factors were selected to assess the landslide risk. They were slope, aspect, cumulative (or total) catchment area, formation lithology, seismic intensity (earthquake intensity), normalized difference vegetation index (NDVI), stream/river buffers, land use/land cover type, and precipitation.

Slope and aspect were first derived from the 1:50,000 contour map with a 20-m contour interval using ArcGIS 10.0 software. The results were then classified based on the criteria that the information quantities computed from every categories follow normal distribution and maximize standard deviation (Figs. 2, 3). As a result, slope was classified into five categories: 0–5, 5–10, 10–15, 15–20, 20–25, and larger than 25 degrees. Aspect was classified into eight classes: flat, 0–30, 30–150, 150–200, 200–250, 250–310, 310–330, and 330–360 degrees.

Cumulative catchment area (Fig. 4) was obtained using the Hydrology toolbox in ArcGIS 10 software and then reclassified according to the number of grids. A formation lithology map was extracted from the geological map of the Three Gorges Dam Reservoir area. This map was then reclassified into 14 categories according to the geological formations in Fuling District (Fig. 5).

Streams and rivers with single or double channel were extracted from 1:2000 topographic maps. Four buffer zones (0–150, 150–250, 250–400 m, and larger than 400 m) were created along the channels. Such buffer zones were then used to generate a stream

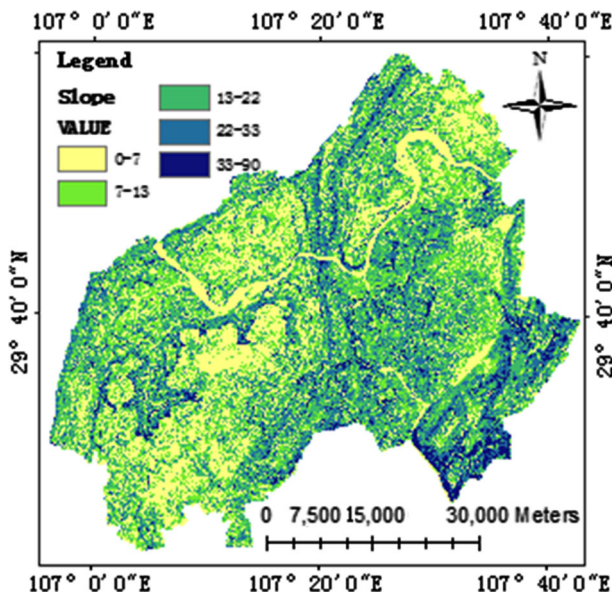


Fig. 2 Slope

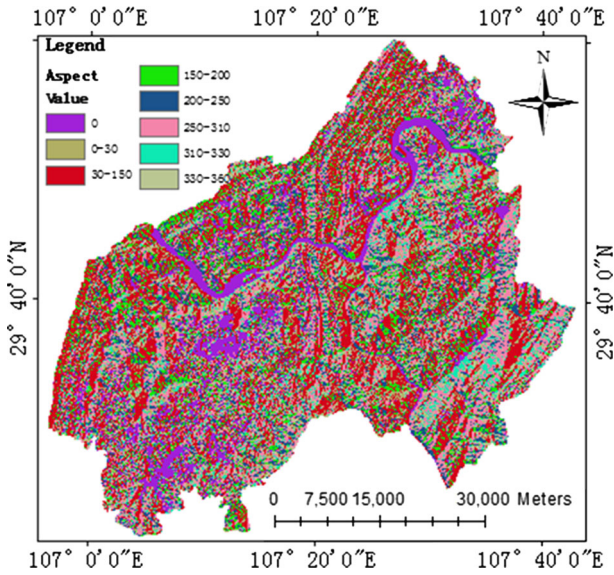


Fig. 3 Aspect

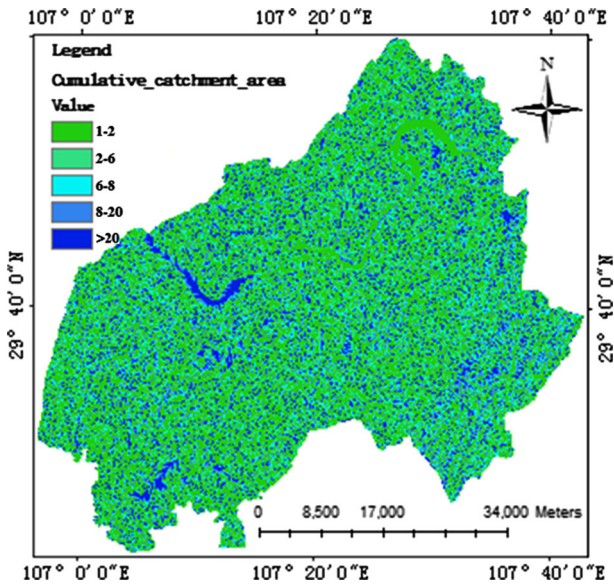


Fig. 4 Cumulative catchment area

and river buffer map, showing the influences of the surface stream and river network (Fig. 6).

A pan-sharpened image was created from a 2-m panchromatic band (0.45–0.90) and 4- to 8-m multispectral bands (0.45–0.52, 0.52–0.59, 0.63–0.69, 0.77–0.89) of a Chinese GF-

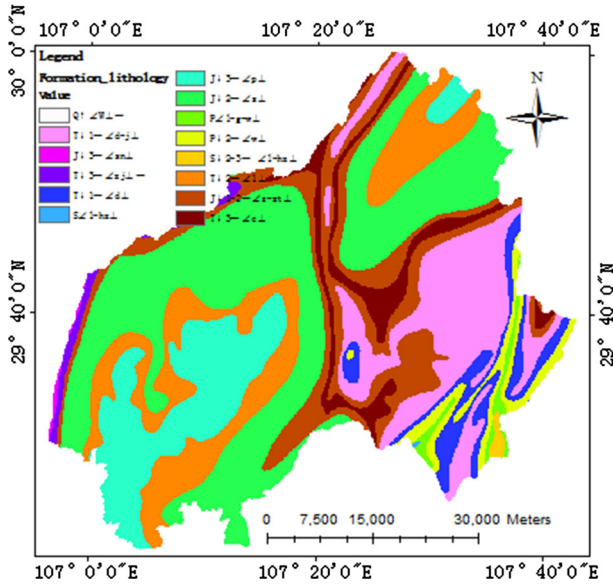


Fig. 5 Formation lithology

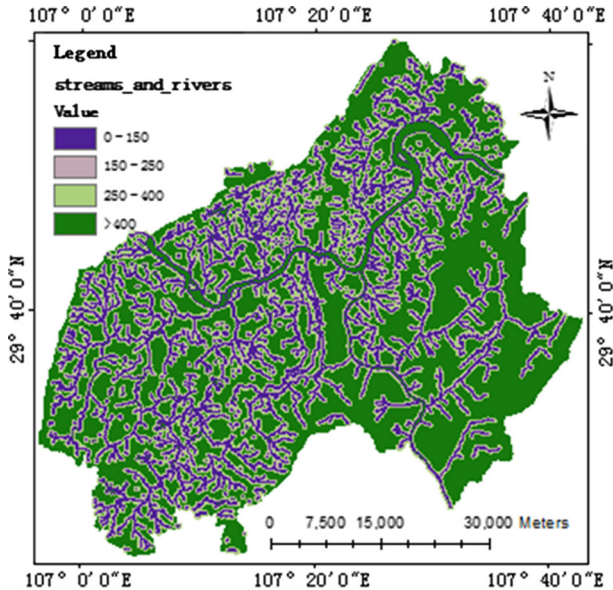
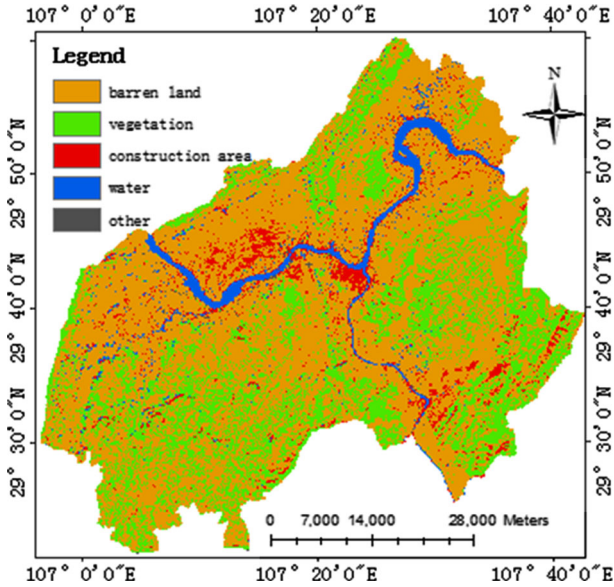
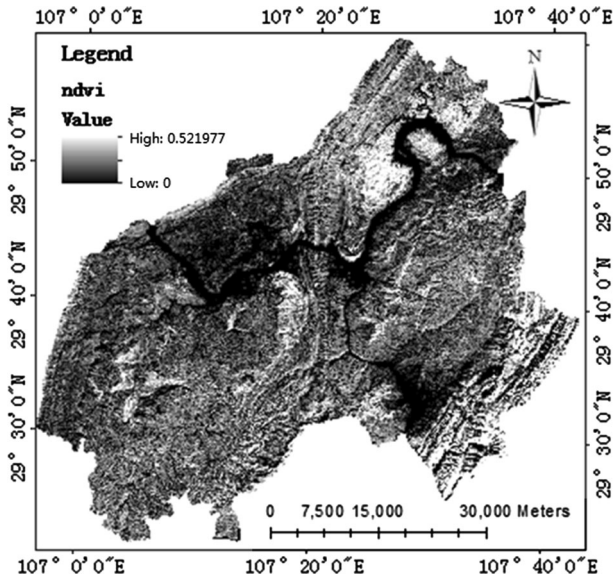


Fig. 6 Multiple buffer zones of surface stream and river network

1 image, obtained on December 24, 2013, in ENVI 5.1 using Gram–Schmidt spectral sharpening method. This pan-sharpened image was subsequently classified into five categories: construction area, vegetation, water, barren land, and other, (Fig. 7) based on



**Fig. 7** Land use/land cover



**Fig. 8** NDVI

maximum likelihood classification method with an overall accuracy of 95.276 % and a Kappa index of 0.9241.

NDVI was derived from the above-mentioned GF-1 fusion image. The NDVI ranges from  $-1$  to  $1$ , with higher positive numbers representing more vegetation (Fig. 8).



The Chinese seismic intensity zoning map, published by the China Earthquake Administration in 1990, was scanned, georeferenced, and digitized using ArcGIS 10.0 software. The map showed that over 99 % of the study area, as well as all the landslide sites, were located in the area with a seismic intensity of scale V (Fig. 9).

Average precipitation data from December 1 to December 24, 2013, were obtained from the precipitation stations located throughout the study area. And the inverse distance weighting (IDW) spatial interpolation method was employed to create a precipitation map.

### 3.3 Methodology flowchart

The basic steps of the methodology in this research were listed below (Fig. 10):

1. Derive the triggering factors of landslides from remote sensing images and other sources;
2. Randomly select a subset of all the landslides sites as the input data to the model, and use the remaining subset as testing data for later model validation;
3. Calculate the information value for each individual triggering factor, and generate the combined information value map by overlaying the individual information values;
4. Reclassify the combined information value to get the final landslide risk distribution map;
5. Test and validate the information model;

Every computed information value from above-mentioned triggering factors was rasterized. The final assessment map was generated by overlaying these rasterized triggering factors and reclassifying the overlaid result.

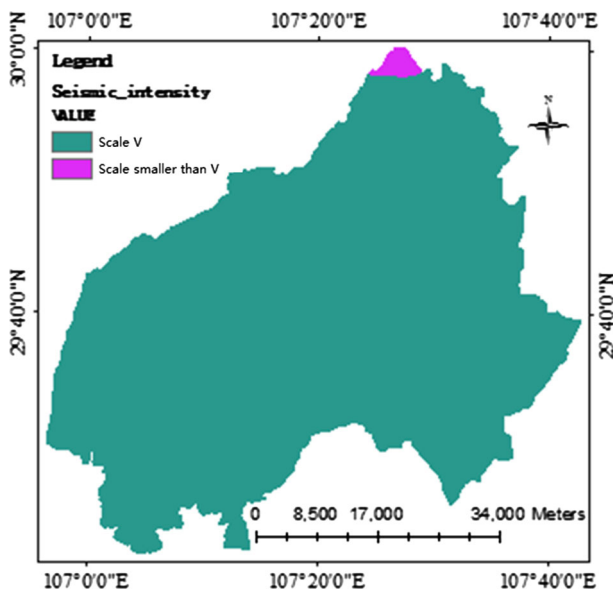
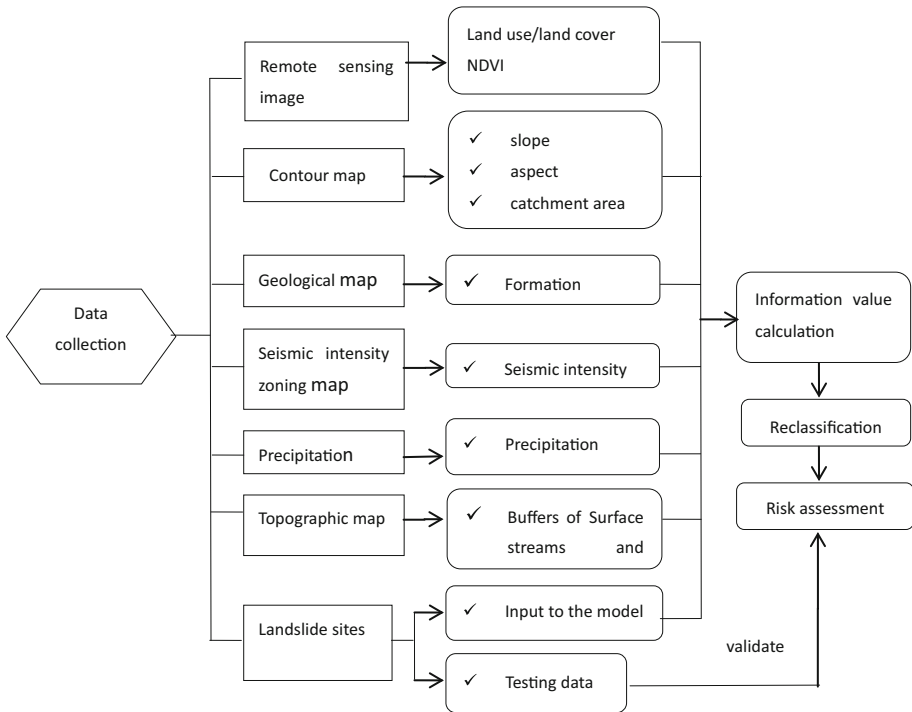


Fig. 9 Seismic intensity



**Fig. 10** Methodology flow chart

**Table 1** Example information values for individual triggering factors

| Triggering factors  | Category     | Number of landslide sites | Information value | Rank of information values |
|---------------------|--------------|---------------------------|-------------------|----------------------------|
| Slope (degree)      | 0–5          | 18                        | −1.529255         | 39                         |
|                     | 5–10         | 33                        | −0.277137         | 31                         |
|                     | 10–15        | 87                        | 0.516223          | 9                          |
|                     | 15–20        | 56                        | 0.652874          | 8                          |
|                     | 20–25        | 4                         | −1.259605         | 38                         |
| Land use/Land cover | Other        | 11                        | 4.095245          | 1                          |
|                     | Water        | 13                        | 0.809629          | 5                          |
|                     | Construction | 36                        | 2.015316          | 2                          |
|                     | Vegetation   | 2                         | −4.643001         | 43                         |
|                     | Barren land  | 136                       | 0.053346          | 16                         |

### 4 Results and discussion

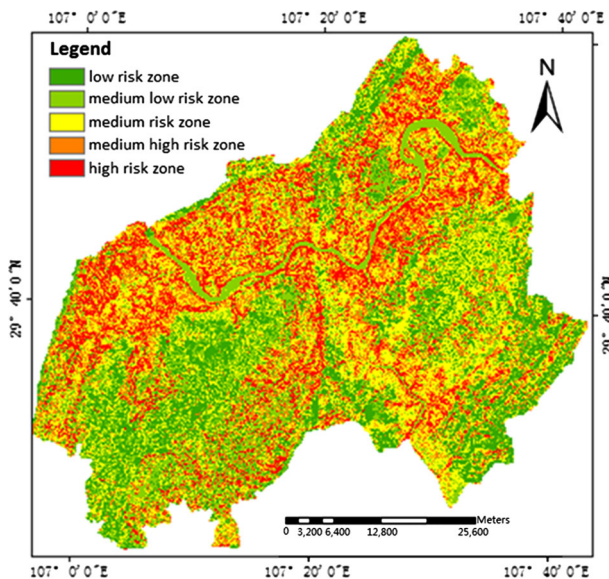
By the end of 2013, there were 284 landslide sites in the study area, including nine unstable rocks, 49 deformation bodies, 33 reservoir bank collapses, 189 landslides, and four slope hazards. Seventy percent (198) of these hazard sites were randomly selected as the input to

the information model, and the rest 30 % (86) hazard sites were used to test the results and validate the model.

The calculated information values of some triggering factors are listed in Table 1. The combined information values were reclassified into five categories based on the natural breaks method: low-risk zone, medium- to low-risk zone, medium-risk zone, medium- to high-risk zone, and high-risk zone. The landslide risk distribution map was generated from these categories (Fig. 11). Table 1 and Fig. 11 showed that the medium- to high- and high-risk zones had the following characteristics:

- (a) Landslides mainly occur in the areas with slopes between 5 and 20 degree, aspects of northeast and northwest, and stratum of Jurassic system on Shaximiao or Jurassic artesian well under the EC group;
- (b) Most landslides occur within 150 m of surface rivers/streams, the closer the surface rivers/streams, the more likely do the landslides occur. Generally, landslides occur on both banks of the reservoir as the variation and intensity in water level of the reservoir affect the development of landslides. The rise and fall of the water in the reservoir influence the moisture content of the soils, thereby changing the shear strength and causing landslides.
- (c) Construction area and other land use/land cover are prone to landslides. These land use/land cover types are also areas with an NDVI ranging from 0 to 0.04, which indicates human activities have relatively high impacts on landslides occurrence. The unreasonable and excessive exploitation and production destruct the ecological and geological environment, leading to the occurrence of landslides.

From Table 2, we can see that the area of high-risk and medium- to high-risk zones is 929.45 km<sup>2</sup>, accounting for 31.72 % of the total area in the study area. The number of landslide sites in high-risk and medium- to high-risk zones is 134 and 40, respectively, accounting for 87.9 % of the total number of landslide sites in the study area. The hazard



**Fig. 11** Landslide risk distribution map

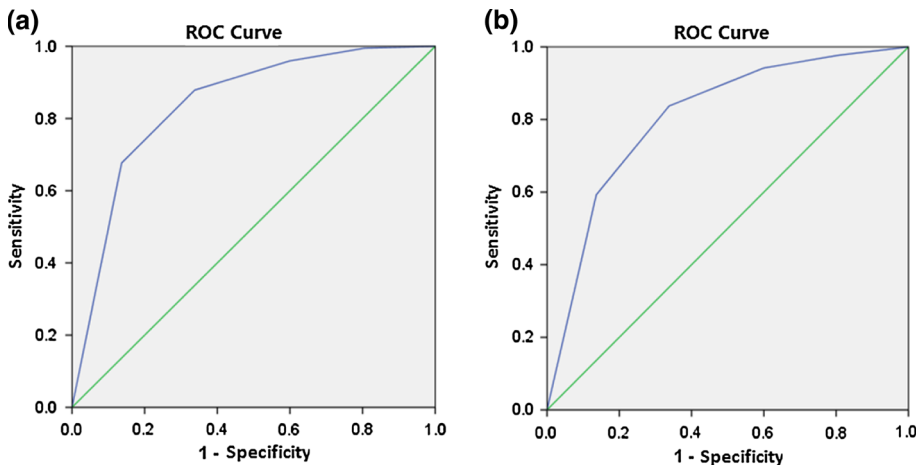
**Table 2** Comparison of risk zones and landslide sites

| Classified risk zones | Area (km <sup>2</sup> ) | Percentage of the total area (a, %) | Number of landslide sites | Percentage of the total number of landslide sites (b, %) |
|-----------------------|-------------------------|-------------------------------------|---------------------------|--|
| Low                   | 572.686                 | 19.55                               | 1                         | 0.5  |
| Medium low            | 695.799                 | 23.75                               | 7                         | 3.5  |
| Medium                | 732.046                 | 24.98                               | 16                        | 8.1  |
| Medium high           | 526.282                 | 17.96                               | 40                        | 20.2   |
| High                  | 403.167                 | 13.76                               | 134                       | 67.7   |

sites are rarely located in the medium-risk, medium- to low-risk, and low-risk zones. In general, the distribution of landslide sites is well associated with the classified risk zones, with more hazardous sites in high- and medium- to high-risk zones and less hazardous sites in low- and medium- to low-risk zones.

The receiver operating characteristic (ROC) curve is widely adopted to test the accuracy of geological hazard risk assessment as it is not only simple and straightforward, but also can reflect the relationship between sensitivity and specificity, with *y*-axis indicating sensitivity and *x*-axis denoting 1-specificity (Hamid et al. 2012; Omar et al. 2012; Liang et al. 2013; Pontius and Kangpin 2014). The value range of the area under curve (AUC) is 0–1. The bigger the AUC, the higher accuracy is the model. When the AUC is 0.5, the model is regarded as being random. Only when it is bigger than 0.7, the model is considered credible.

The success-rate method and prediction-rate method were used to create the ROC curve for the validation (Chung and Fabbri 2003). In the success-rate method, the predicted landslide distribution map was compared with the input data used in the modeling (the 198 landslide sites) to test the results. It represented the correspondence between the modeling results and the input data. Figure 12a showed the ROC curve based on success-rate method. The AUC was 0.839, meaning a good fit and a relatively high accuracy of the

**Fig. 12** a Success-rate ROC curve; b Prediction-rate ROC curve

information model using the input data. The prediction-rate method compared the predicted landslide risk distribution map and the testing data (86 landslide sites) to evaluate the reliability of the information model. The ROC curve based on prediction-rate model (Fig. 12b), with the AUC of 0.807, implied that the proposed information model was reliable and was able to be used to predict future landslide occurrences in the study area.

## 5 Conclusion

In this paper, based on the geomorphological and geological characteristics in Fuling District, Chongqing City, nine triggering factors were selected and derived from remote sensing images and other sources to build the geospatial information quantity model for the assessment and prediction of landslide risk in the study area. The ROC curve was employed to validate the information model. The success-rate ROC curve indicated a good fit between the modeling results and the input data, and the prediction-rate ROC curve showed that the model was reliable to predict the future landslide occurrences in the study area. The results in the study also indicate the high possibility of applying the proposed method to the entire Three Gorges Dam Reservoir region in the future to better assist the efficient and effective management of landslide hazards in the area.

**Acknowledgments** This paper is funded by Chinese Governmental Projects in Satellite Application.

## References

- Bai S, Lv G, Wang J, Zhou P, Ding L (2011) GIS-based rare events logistic regression for landslide-susceptibility mapping of Lianyungang, China. *Environ Earth Sci* 62:139–149
- Bui DT, Lofman O, Revhaug I, Dick O (2011) Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. *Nat Hazards* 59(3):1413–1444
- Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012) Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. *Geomorphology* 171–172:12–29
- Chen Y, Guo H, Wang Q (2013) Earthquake geological hazard susceptibility assessment based on Remote Sensing and GIS in Lushan County, Sichuan Province, China (in Chinese). *Chin Sci Bull* 58:3859–3866
- Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Nat Hazards* 30(3):451–472
- Conforti M, Pascale S, Robustelli G, Sdao F (2014) Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena* 113:236–250
- Dai Y (2013) Landslide risk assessment methods based on information model in the Three Gorges Reservoir (in Chinese). Master thesis, Tsinghua University, Beijing, China
- Deng H, He Z, Chen Y, Cai H, Li X (2014) Application of information quantity model to hazard evaluation of geological disaster in mountainous region environment: a case study of Luding County, Sichuan Province (in Chinese). *J Natl Disasters* 23:67–76
- Gao K, Cui P, Zhao C, Wei F (2006) Landslide hazard evaluation of Wanzhou District, Chongqing city, China, based on GIS information value method (in Chinese). *Chin J Rock Mech Eng* 25(5):991–996
- Gao Z, Wu L, Zhao X (2014) Landslide risk mapping based on GIS in Shenzhen city, Guangdong Province, China (in Chinese). *J Catastrophol* 29(1):67–74
- Hamid RP, Majid M, Biswajeet P (2012) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: safarood Basin, Iran. *Catena* 97:71–84
- Lee S, Min K (2001) Statistical analysis of landslide susceptibility at Yongin, Korea. *Environ Geol* 40(9):1095–1113

- Li J, Zhou CH (2003) Appropriate grid size for terrain based landslide risk assessment in lantau island, Hong Kong. *J Remote Sens* 7(2):86–93
- Liang JW, Kazuhide S, Shuji M (2013) Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. *Comput Geosci* 57:81–92
- Niu J, Cheng W, Lan H, Liu Y, Jie Y (2011) Risk assessment of Secondary geological hazard caused by earthquakes based on information model in Yushu Autonomous Tibetan Prefecture, Qinghai Province (in Chinese). *J Mt Sci* 29:243–249
- Omar FA, Biswajeet P, Saro L (2012) Application of an evidential belief function model in landslide susceptibility mapping. *Comput Geosci* 44:120–135
- Pontius RG Jr, Kangpin Si (2014) The total operating characteristic to measure diagnostic ability for multiple thresholds. *Int J Geogr Inf Sci* 28(3):570–583
- Pradhan B, Lee S (2010) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environ Earth Sci* 60:1037–1054
- Pradhan B, Youssef AM (2010) Manifestation of remote sensing data and GIS for landslide hazard analysis using spatial-based statistical models. *Arabian J Geosci* 3(3):319–326
- Ren K, Mao J, Chen G (2011) A Discussion on the weights of the influence factors of geological hazard. *Chin J Geol Hazard Control* 1:80–86
- Sdao F, Lioi DS, Pascale S, Caniani D, Mancini IM (2013) Landslide susceptibility assessment by using a neuro-fuzzy model: a case study in the Rupestrian heritage rich area of Matera. *Natl Hazards Earth Syst Sci* 13:395–407
- Wang N, Peng K, Li Q (2012) Quantitative evaluation of geological disaster liability based on RS & GIS analysis: a case study of Wufeng County, Hubei Province. *Earth Sci Front* 6:221–229
- Wang J, Yin K, Xiao L (2014) Landslide susceptibility assessment based on GIS and weighted information value: a case study of Wanzhou District, Chongqing city in the three Gorges Reservoir (in Chinese). *Chin J Rock Mech Eng* 4:797–808
- Yilmaz I (2010) The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability (CP) and artificial neural networks (ANN). *Environ Earth Sci* 60:505–519
- Yin K, Liu Y (2000) Systematic studies on landslide hazard zonation. *Chin J Geol Hazard Control* 11(4):28–33
- Yin K, Zhu L (2001) Landslide hazard spatial zoning and GIS application (in Chinese). *Earth Sci Front* 8(2):279–284
- Yuan S, Huang R (2001) Application of information model based on GIS in geological hazard risk zonation. *J Chengdu Univ Technol* 28:89–92