Nat Hazards
DOI 10.1007/s11069-013-0759-y

ORIGINAL PAPER

Susceptibility evaluation and mapping of China's landslides based on multi-source data

Chun Liu · Weiyue Li · Hangbin Wu · Ping Lu · Kai Sang · Weiwei Sun · Wen Chen · Yang Hong · Rongxing Li

Received: 25 October 2012 / Accepted: 11 June 2013

Abstract Landslides are occurring more frequently in China under the conditions of extreme rainfall and changing climate, according to News reports. Landslide hazard assessment remains an international focus on disaster prevention and mitigation, and it is an important step for compiling and quantitatively characterizing landslide damages. This paper collected and analyzed the historical landslide events data of the past 60 years in China. Validated by the frequencies and distributions of landslides, nine key factors (lithology, convexity, slope gradient, slope aspect, elevation, soil property, vegetation coverage, flow, and fracture) are selected to construct landslide susceptibility (LS) empirical models by back-propagation artificial neural network method. By integrating landslide empirical models with surface multi-source geospatial and remote sensing data, this paper further performs a large-scale LS assessment throughout China. The resulting landslide hazard assessment map of China clearly illustrates the hot spots of the high landslide potential areas, mostly concentrated in the southwest. The study implements a complete framework of multi-source data collecting, processing, modeling, and synthesizing that fulfills the assessment of LS and provides a theoretical basis and practical guide for predicting and mitigating landslide disasters potentially throughout China.

A major outcome of this work is landslide historical data collection over 60 years and the development of LS in China, through multi-factor empirical method.

C. Liu · W. Li (✉) · H. Wu · P. Lu · W. Chen · R. Li
Center for Spatial Information Science and Sustainable Development Applications, Tongji University, Shanghai, China
e-mail: 326lwy@tongji.edu.cn

C. Liu · W. Li · H. Wu · P. Lu · K. Sang · W. Sun
College of Surveying and Geo-Informatics, Tongji University, Shanghai, China

Y. Hong
School of Civil Engineering and Environmental Science, National Weather Center, University of Oklahoma, Norman, OK, USA

R. Li
Mapping and GIS Lab, The Ohio State University, Columbus, OH, USA

Published online: 25 June 2013

Keywords Landslide susceptibility · Empirical model · Historical landslide events · ANN · Hot spots

1 Introduction

Natural disasters are complex and are closely connected with environmental degradation and human activities (Henderson 2004). Landslides, which often result in substantial causalities and property losses, are among the most common natural disasters (Manunta et al. 2011). Recently, landslides have occurred even more frequently, resulting in several catastrophic events (Dunning et al. 2007; Wang et al. 2009). More than 90,000 hazards associated with landslides have been reported in 70 cities and counties of several southern and northwestern provinces in China, and tens of millions of people currently live under the high-risk threat of landslides. Landslides have caused thousands of deaths and tens of billions of RMB in property losses (Wibowo et al. 2008). Therefore, landslide monitoring, risk assessment, and prediction are urgent subjects for the international landslide disaster research community. Particularly in China, there is a lack of maps or guidelines for evaluating the relative potential of landslide hazards at regional and national scales; thus, a large-scale comprehensive landslide hazard assessment is critical for predicting and mitigating landslide disasters (Hong et al. 2007a, b).

Major impact factors of landslides include terrain, geology, geomorphology, weather conditions, land cover, and others (Fernandes et al. 2004; Moreiras 2005; Nefeslioglu et al. 2008; Parry 2011). It is crucial to determine the relationships between these factors and landslide occurrences for quantitative evaluation of landslide susceptibility (LS). LS mapping can provide information about the spatial distribution of the probability of regional slope instability (Mathew et al. 2009), which is the first and most important step in landslide hazard assessment to make effective measures for landslide mitigation. Reliable evaluation depends on the quality and range of the available data and the selection of a proper modeling assessment framework in order to analyze landslide formation conditions and characteristics and to show a detailed geometric description of landslides (Choi et al. 2012).

Given historical data and experiences, empirical LS evaluations usually adopt statistical approaches and pattern recognition methods to construct empirical models. The initial weights for landslide factors can be obtained from previous research and literature by evaluating the relationship between spatial factors and landslide occurrences using geographical information system (GIS) (Ghosh 2011). Some studies evaluated LS using GIS approaches with either qualitative or quantitative methods (Guzzetti et al. 1999; Qin et al. 2002; Pradhan et al. 2006; Hong et al. 2007a, b; Yalcin 2008; Mondini et al. 2011). Multifactor methods typically estimate LS using the relationships between past landslide patterns and various site features (Sidle and Ochiai 2006). Some empirical models were used for LS analysis and mapping. These models include logistic regression (LR) (statistical analysis of diagram segments and landslide distributions), probabilistic (the relationships between the probability of failure and landslide stability), analytical hierarchy process (AHP), and fuzzy logic (the weight confirmation of different factors) models. However, there are several difficulties with these methods for the identification of relevant factors used as assessment variables for different types of landslides (Wang et al. 2012). For examples, the empirical models are unsuitable for analyzing susceptibility, in a small-scale landslide evaluation with uncertainty or nonlinearity problems (Bălteanu et al. 2010). An artificial neural network

(ANN) can be used to analyze complex data, such as continuous, indexical, and binary data, at different scales (Chauhan et al. 2010). Based on learning from data with known characteristics to obtain the weights of factors, ANN has been used to recognize unknown data (Pradhan 2011). However, few studies comprehensively established a systematic theory for improving landslide prediction accuracy, especially at the national level. This paper presents the theory of national susceptibility by analyzing >60-year landslide events and frequency datasets by employing BP ANN techniques. Specifically, connection weights are determined to describe the relationship, the different factors, and susceptibility through surface multi-geospatial data and remote sensing data for susceptibility evaluation and mapping of China's landslide. Figure 1 shows the workflow of our paper.

The workflow includes six parts: (1) to collect >60-year historical landslide records from different databases; (2) to identify the impact factors by analyzing different landslide characteristics; (3) to analyze LS with BP ANN models; (4) to calculate the evaluated factors and compare with the sample data; (5) to obtain LS evaluation results; and (6) to map China LS, compared with the landslide inventory (>60 years).

2 Study area

China lies in the east of the Asia–Europe continent on the western shore of the Pacific Ocean, and it covers approximately 9.6 million km^2 of land area (between latitudes 3.85°N and 53.56°N, between longitudes 73.55°E and 135.08°E). The terrain of China is low in the east but high in the west. Mountains, plateaus, and hills cover approximately 67 % of the land area, whereas basins and plains cover approximately 33 %. China is among the countries most vulnerable to landslide disasters. Landslide disasters mainly occur in

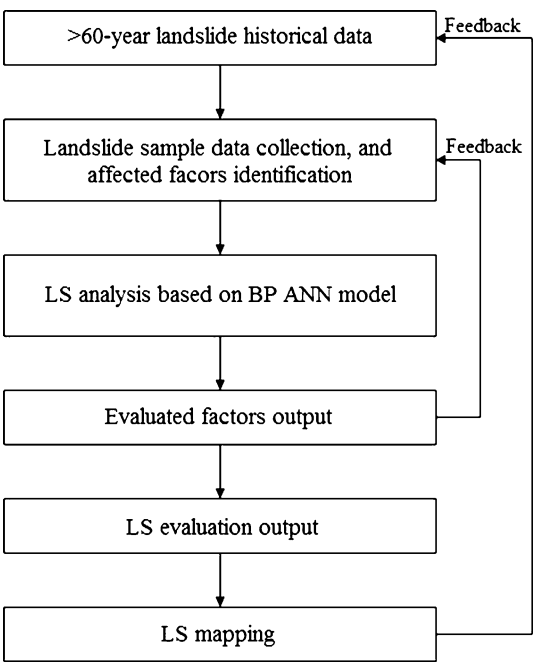

Fig. 1 LS evaluation workflow

regions with steep slopes, such as along the bank slope zones of rivers and streams or gorges with large elevation differences. Many landslides occur along geological tectonic belts (fractures or structural zones), while others occur in areas with soft rock-soil (loose covering layer, loess, mudstone, shale, coal beds). Most of landslides are induced by heavy rainfall (He et al. 2008; Li et al. 2011; Iwahashi et al. 2012).

3 Data acquisition and processing

3.1 Historical landslide data collection and database compilation

The data of the 1,200 >60-year landslide events were obtained mainly from online news reports, local yearbooks, and hazard databases in China, including major landslide reports in China (from newspapers and media), the China Statistical Yearbook (1949–2011), the China Geological Environmental Information Network (CGEIN; http://www.cigem.gov.cn/), the China Risk Network (CRN; http://www.irisknet.cn/), and the Geostress and Geological Disaster Querying Database (GGDQD; http://www.geomech.ac.cn/geo0503/). A total of 1,200 records have been gathered and shown as points on the provincial boundary map with a 10-year interval (Fig. 2).

Figure 3 shows the number of landslide events collected over 60 years. From 1949 to the late 1970s, the reported landslide events are fewer because there was a lack of monitoring techniques, and not enough importance is placed on these reports by the government. After the 1980s, the reports and occurrences start to increase; it is the result of the high disaster occurrence frequency and the government's monitoring. Geographically, landslide events are mainly distributed in the south and southwest of China, where the main landslide-inducing factors include rainfall, earthquake, and flood. These regions can also be characterized by the drastic topographical relief of gorges, rivers, bank slopes, the distribution of geological tectonic belts (fault lines, earthquake zones), and geotechnical factors (loose soil, loess, mudstone, shale, coal beds).

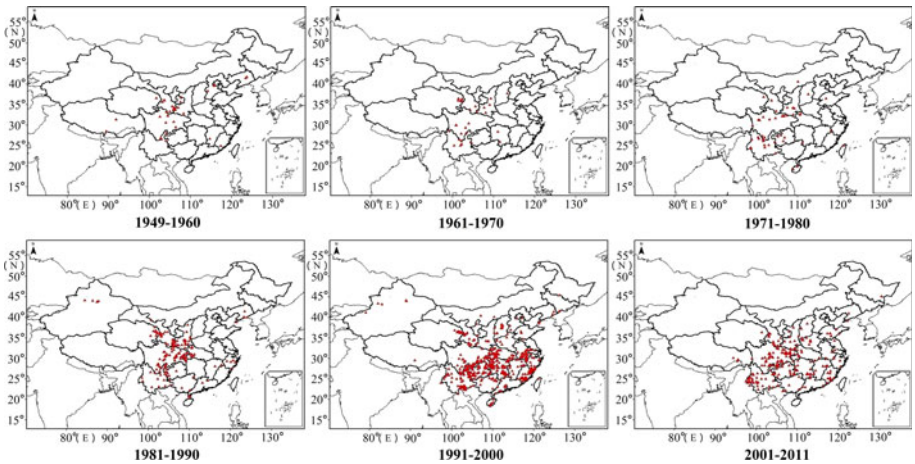

Fig. 2 Landslide locations and occurrences (red dots) between 1949 and 2011 in China

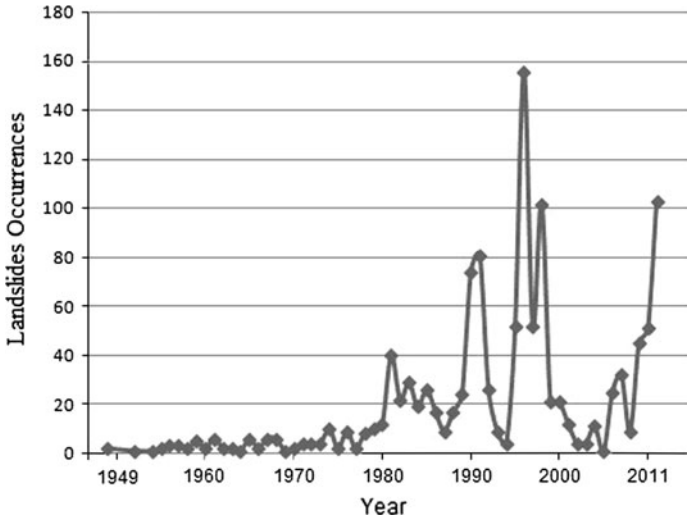

Fig. 3 The number of landslide events over 60 years

Table 1 The structure of the collected landslide events' data in China from 1949 to 2011

No.	Time (YY/MM/DD)	Location			Cause	Fatalities	Economic loss (RMB)
		Name	Longitude (°)	Latitude (°)			
1	1949/5/4	Xining, Qinghai	101.80	36.60	Loess	280	–
2	1949/7/1	Open-cast coal mines of Fushun	123.90	41.80	Excavation	–	300 million yuan
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1,200	2011/9/29	Yanbian, Sichuan	101.51	26.40	Rainfall	–	–

The collected landslide data were compiled in a geodatabase based on GIS, which is structured as illustrated in Table 1. The spatial accuracy of longitude and latitude is 0.01°, equivalent to 1 km.

3.2 Landslide data uncertainty

It is a challenge to compile a national landslide database due to the limited number of sources for collecting landslide events that cover more than 60 years. Thus, the landslide database may lack accuracy and completeness. Common problems exist in similar studies about the regionalization and evaluation of landslide reports (Refice and Capolongo 2002; Che et al. 2012; Peruccacci et al. 2011).

We identify four main difficulties that affect the assessment results when compiling the landslide database using unmapped events. First, this research involves landslides triggered by all factors. The uncertainty in triggering factors of some collected reports renders that a landslide event correlated with multiple triggering sources. Second, Landslides are a complex process associated with many inducing factors, and therefore, extensive

misreports in landslide might exist. Third, the data sources of landslide may be inadequate or even vacant in many sparsely inhabited areas. In the southwest, there are many conditions of landslide occurring, including plentiful rainfall, occasional earthquake, undulating terrain, and loose soil. However, the number of landslide events in the same region of Fig. 2 is not the largest. The reason is that the sparse population distribution and little attention bring about fewer landslide reports. Finally, these events are of different scales. The accuracies diverge significantly among different districts, counties, and provinces.

4 Multi-factor empirical LS assessment

4.1 Identification of landslide-inducing factors

Multi-factor analysis methodologies can typically assess landslide hazards using the relationships between historical landslide events and various site characteristics. In such cases, the weights of different factors affecting slope stability are important. In this study, the landslide-inducing factors can be divided into two main categories (Wu and Sidle 1995):

1. Internal factors: those that have decisive effects on landslides, including geology (lithology), geomorphology (convexity), DEM (slope gradient, slope aspect, elevation), soil property, vegetation cover, flow distribution, and fractures.
2. External factors: those that trigger landslides suddenly, such as rainfall, earthquake, and flood.

Generally, the variation in external factors is a short and usually unpredictable process, and the distribution of internal factors affects LS.

The main data deriving the internal factors were achieved from the National Geomatics Center (NGC) and other government sources (https://modaps.nascom.nasa.gov:8499/, http://srtm.csi.cgiar.org/selection/inputCoord.asp, http://www.eq-icd.cn/webgis/download.htm) (Table 2). Using an interpolation technique, the different data layers are converted to grid maps with the same cell size of 1 km × 1 km. The factors are described in the following section.

Table 2 Data sources and specific use

Data types	Description			Specific use
	Satellite/sensor	Spatial resolution	Acquired year	
Remote sensing data	Terra, Aqua/MODIS	250 m	2007	Vegetation cover
Digital elevation model (DEM)	SRTM	90 m	2000	Slope gradient, aspect, elevation
Geological maps	N/A	Scale: 1:500,000	2007	Lithology
Soil type maps	N/A	Scale: 1:3,000,000	2000	Soil property
Geomorphological maps	N/A	Scale: 1:3,000,000	2000	Convexity
River distribution maps	N/A	Scale: 1:3,000,000	2000	Flow
Seismic zone distribution in China	N/A		2008	Fracture

4.1.1 Lithology (Lith)

Landslides have been associated with many different geologic materials around China. In some specific areas, there are relationships between many lithologies and landslide occurrences. The structural features of bedrock promote landslide initiation in several ways: (1) by forming weak surfaces that are prone to sliding; (2) by facilitating the introduction of groundwater into the overlying soil mantle; and (3) by creating an opportunity to destabilize the regolith because of weathering.

According to CDSTM (China Digital Science and Technology Museum, http://amuseum.cdstm.cn/moudisaster/page/knowledgec.jsp?pid=3100502), the formation conditions of landslides in China mainly involve the following lithologies: Jurassic (J), Cretaceous Mudstone (M), Shale (S), Argillaceous Sandstone (A), Siltstone (Si), Coal Beds (Co), Sandy Slate (Sa), and Phyllite (P). These lithologies in China are mapped in Fig. 4a.

4.1.2 Convexity (Con)

Because slope movement is a phenomenon that occurs in a gravitational field, the higher the degree of convexity, the more susceptible the slope fails. Thus, the platform of the

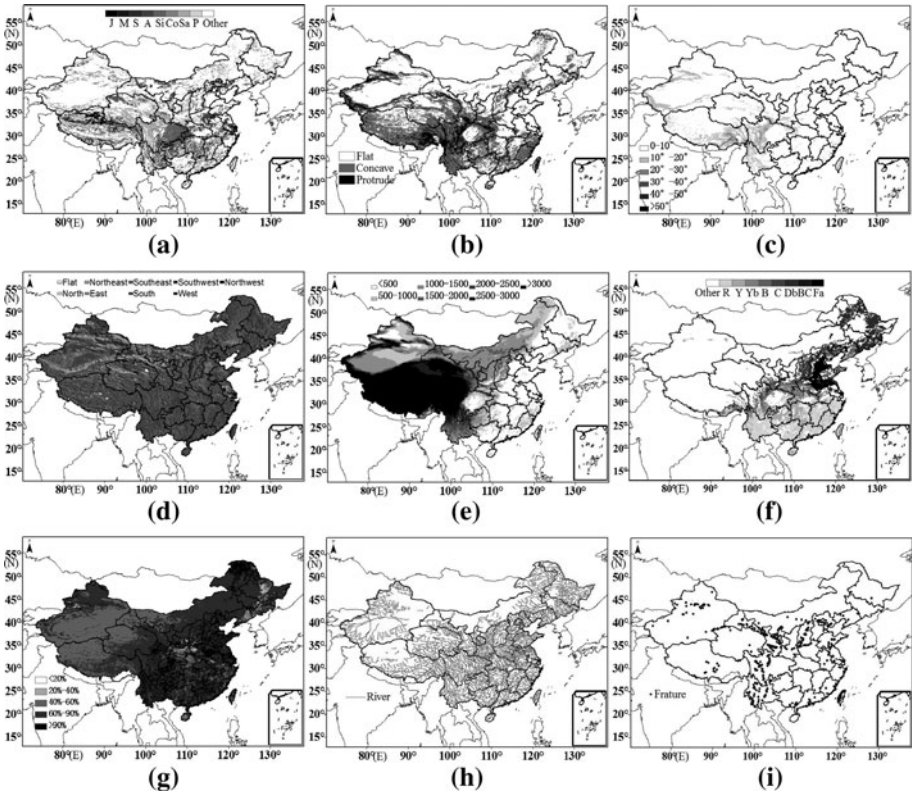

Fig. 4 Nine factors of landslides and their maps in China: lithology (**a**), convexity (**b**), slope gradient (**c**), slope aspect (**d**), elevation (**e**), soil property (**f**), vegetation coverage (**g**), flow (**h**), fracture (**i**)

slope is an important factor affecting slope stability. The types of convexity from geomorphological maps include concave, flat, and protruded (Fig. 4b).

4.1.3 Slope gradient (SG)

Slope gradient is a crucial factor in landslide initiation. With the increase in the SG, the component force of gravity is in the slope direction, and the slope will be bound to higher landslide risk (Dai et al. 2002). This study uses the 90-m-resolution DEM with the predefined interval of slope gradient as 10°(Ercanoglu and Gokceoglu 2004) (Fig. 4c).

4.1.4 Slope aspect (SA)

Slope aspect strongly affects hydrological processes via evapotranspiration and weathering processes in a drier environment. The slope aspect can be divided into nine directions from the 90-m-resolution DEM (Fig. 4d). For example, the probability of landslide is almost zero in the 'flat' areas.

4.1.5 Elevation (Elev)

Elevation is also a critical factor in landslide assessment (Ercanoglu and Gokceoglu 2004; Park 2008). Landslide occurs in the central part of the hilly and mountainous areas. Few landslides occur on the top of hills or mountains because most rocks on the top are weathered and hard. The distinctive terrain features of China bring that most landslides occur in the regions with lower altitudes (400–800 and 2,000–3,000 m) covering the first and second level shapes (Zhang et al. 2009). We divided the elevation into 7 classes at 500-m intervals using the 90-m-resolution DEM in Fig. 4e.

4.1.6 Soil property (SP)

Similar to lithology, soil property is also a basis of landslide formation. Soils with loose structures and vulnerability to shearing and weathering are subject to high LS. In China, the distribution of soil exhibits some zonality. Soil color shows the information of organic matter in the soil, drainage, and fertility. The soil color is accustomed to identify soil types. The soil types that cause landslides on the basis of CDSTM are red soil (R), yellow soil (Y), yellow–brown soil (Yb), brown soil (B), cinnamon soil (C), dark-brown soil (Db), brown coniferous forest soil (BC), and fluvo-aquic soil (Fa) (Fig. 4f).

4.1.7 Vegetation coverage (VC)

Vegetation coverage plays a positive role in LS analysis. With remote sensing data, vegetation coverage can be calculated on a larger scale. An effective method is to use the Normalized Difference Vegetation Index (NDVI) (Eq. 1), which is defined as the ratio of the difference and sum from the near-infrared band (NIR) and red band (R).

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}} \tag{1}$$

In this study, MODIS data covering China with a spatial resolution of 250 m collected in 2007 is used, and for every pixel, an NDVI value is obtained according to Eq. (1). Subsequently, the vegetation coverage (VC) value is calculated as Eq. (2).

$$VC = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad (2)$$

$NDVI_{max}$ is the max NDVI value in the entire area of China, which may require an NDVI for a cell of dense vegetation; $NDVI_{min}$ is the NDVI value in the area that may be a cell for soil or rock (Wan 2009). VCs are divided into 5 classes at 20 % intervals.

4.1.8 Flow (Fl) and fracture (Fr) (Fig. 4h, i)

Finally, landslides correlate closely with the distribution of flows and fractures (Montgomery et al. 2002). Normally, the closer the place is to rivers or fault zones, the greater the possibility of landslide. By classifying the distances to flows and fractures, this study uses 0.5-km intervals.

4.2 ANN landslide model

Artificial neural networks are generic nonlinear functions that have been broadly used to solve complex problems such as land use classification through the establishment of processing units, network topology, and training rules. Multi-layer perception (MLP), a frequently used ANN, has one or more hidden layers between its input and output layers. The number of neurons and the hidden layers can be empirical methods according to different applications (Braspenning et al. 1995; Gong 1996). Three steps are involved in ANN implementation, including training, weights estimation, and output results. The trained data from input neurons are processed through hidden nodes to obtain output values. For instance, the input of neuron j of a single hidden layer is expressed as:

$$net_j = \sum_{i=1}^{t} w_{ij} p_i \quad (3)$$

where w_{ij} represents the weight between node i of the input layer and node j of the hidden layer; p_i is the input data at node i, which is the probability of the concerned cell having the internal factor; and t represents the number of nodes on the input layer. The output value p_i produced through the hidden neuron is usually a nonlinear sigmoid function that is applied to the weighted input data net_j before being processed to the next layer. The equation is as follows:

$$p_j = f(net_j) = \frac{1}{1 + e^{-net_j}} \quad (4)$$

Among the many mature algorithms, the back-propagation (BP) algorithm has been widely used in multi-factor model construction and is thus adopted in this study. In this algorithm, the error function E is used to determine the accuracy of the output.

$$E = \frac{1}{2} \sum_{i=1}^{n} (T_i - O_i)^2 \quad (5)$$

where T_i represents a target output and O_i is the corresponding network output. The computational process terminates until E reaches the minimum, and the adjusted weights are obtained.

In our LS ANN model, the LS values are estimated from the relative probabilities of landslide-inducing factors through the network weights. The input layer includes nine

factors (nodes), as mentioned above: lithology, convexity, gradient, aspect, elevation, soil property, vegetation cover, flow, and fracture. The output layer has one output node defined as the susceptibility of landslide occurrence (Fig. 5).

5 Results

5.1 Training of the LS ANN

5.1.1 Sample data selection

The developed ANN is implemented in MATLAB® software environment. In the ANN training, the appropriate weights connecting the network nodes are established, so that the LS map with a resolution of 1 km × 1 km can be derived as the ANN output by minimizing the error function.

To obtain the unknown weights of nine input factors, the calculation requires nine sample data (nonlinear correlation) at least. If selected sample data are too much and uneven, the network convergence speed may be slow and the accuracy will be not ideal. Figure 2 shows the time and spatial distribution of 1,200 significant landslide events. The result indicates most landslides uniformly spread over the southeast, middle, and southwest of China. The periods of selected factors in Table 2 are all past the 2000 year, and therefore, we select sample data from the records after 2000 to guarantee the temporal consistency. The green points represent the sample locations with ID from 1 to 20 in our study. The selecting rule guarantees the sample data are timely and spatially homogeneous and is listed as follows.

First, we divide the base map into 64 segments by 5°-interval latitudes and 10°-interval longitudes. Second, we assign successive real numbers for the intersections or the nearby landslide points from left to right and top to bottom. Third, we regard the nearest landslide point as a sample data when the distance between an intersection and the nearest landslide point is less than 100 km. Figure 6 illustrates the locations of 20 selected sample data (green points) without considering the intersections near the border. Thus, the method can solve the timely and spatially heterogeneous problem. However, in the national scale, large uncertainties exist depending on the accuracy of selected sample data. In Sect. 5.2, we will test the result of LS map (Table 3).

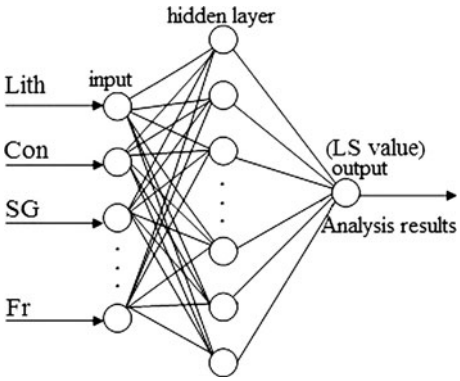

Fig. 5 The LS ANN model

5.1.2 LS ANN training

The first step in the LS ANN training is to compute the input data P_i ($i = 1, 2, \ldots, n$) as the node of the ANN in Fig. 5 or the internal factor at the location of a trained record. Because P_i is the probability that the concerned map cell would have the characteristics of the internal factors, it requires a conversion of the ith internal factor values or the ranges in Table 4 to correspond probability values. We define the relative probability according to the theory (Kumar et al. 2005).

$$RP_{ij} = \frac{P_{ij}}{P_i} \tag{6}$$

P_i is the probability of the ith factor for any landslide event, P_{ij} ($j = 1, 2, \ldots, n$) is the probability of the jth value given the ith factor, and n is the number of factors. In implementation, the factors are realized by using the corresponding frequencies computed from the trained records. The individual relative probability RP_{ij} is obtained by the total number of records with both the ith factor and the jth value dividing the total number of records by the ith factor.The occurrence frequencies reflect the number of landslide reports, and the output data were set as the summation of different factor frequencies (7).

$$Z = X_1 + X_2 + \cdots + X_n \tag{7}$$

X_n is the quantitative value from every factor; Z for the ANN trainings would be the normalized Z' between 0 and 1 with Eq. (8).

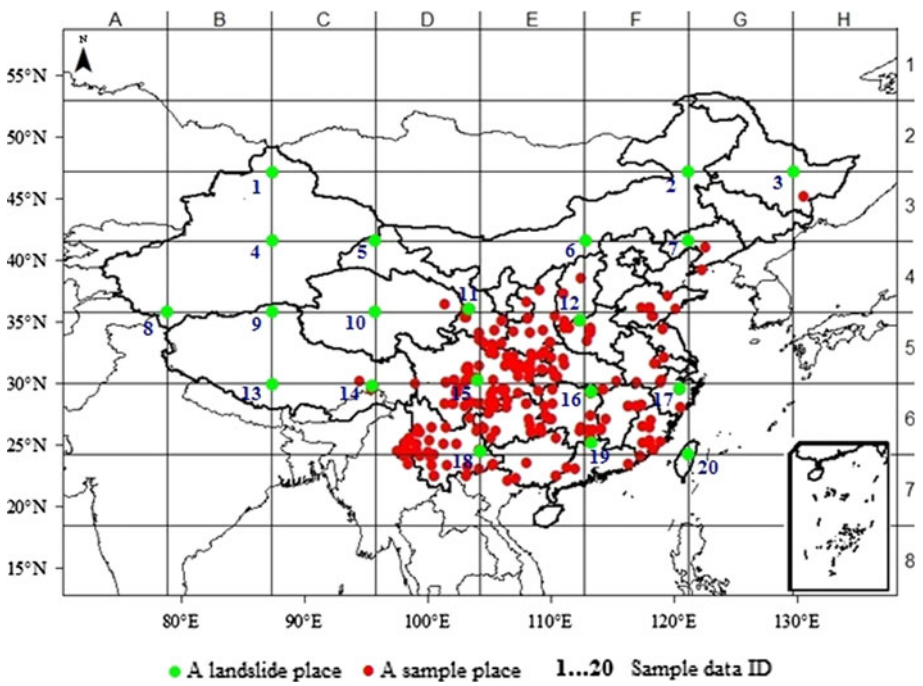

Fig. 6 The selection of trained samples

Table 3 LS evaluation factors and classification (each factor is quantified with a number of classes. The class values or ranges are listed)

Factors	Classification							
	1	2	3	4	5	6	7	8
Lith	J	M	S	A	Si	Co	Sa	P
Con	Concave	Flat	Protrude					
SG	0°–10°	10°–20°	20°–30°	30°–40°	40°–50°	>50°		
SA	East	Southeast	South	Southwest	West	Northwest	North	Northeast
Elev (km)	<0.5	0.5–1.0	1–1.5	1.5–2.0	2.0–2.5	2.5–3.0	>3.0	
SP	R	Y	Yb	B	C	Db	BC	Fa
VC (%)	<20	20–40	40–60	60–80	>80			
Fl (km)	<0.5	0.5–1.0	1.0–1.5	1.5–2.0	2.0–2.5	2.5–3.0	>3.0	
Fr (km)	<0.5	0.5–1.0	1.0–1.5	1.5–2.0	2.0–2.5	2.5–3.0	>3.0	

Table 4 Occurrence frequencies of different input factors

Factors	Classification							
	1	2	3	4	5	6	7	8
Lith	0.12	0.09	0.17	0.24	0.15	0.11	0.04	0.08
Con	0.58	0.26	0.16	0	0	0	0	0
SG	0	0.03	0.16	0.29	0.27	0.25	0	0
SA	0.14	0.16	0.18	0.17	0.15	0.09	0.09	0.08
Elev	0.01	0.14	0.17	0.22	0.19	0.19	0.08	0
SP	0.38	0.36	0.15	0.02	0.05	0.02	0	0.02
VC	0.34	0.25	0.23	0.11	0.07	0	0	0
Fl	0.30	0.25	0.18	0.12	0.06	0.04	0.02	0
Fr	0.36	0.20	0.15	0.13	0.06	0.05	0.05	0

$$Z' = \left(\frac{1}{1+e^{-Z}} - 0.5\right)/0.5 \tag{8}$$

By the methods mentioned above, 17 trained samples and 3 tested samples selected uniformly to be input data in the distribution maps of various factors are nonlinear correlations in Tables 5, 6. In the normalized summation of the different frequencies, every sample can be considered as the output data.

The number of neurons of the hidden layer is more than twice that of the input data. The numbers of input, hidden, and output layers are 9, 19, and 1, respectively. The BP learning algorithm, with a network training error of 10^{-3} and 1,000 training runs (Fig. 7), is implemented to train the various ANN architectures. When the training reaches 146 times, the mean squared error is below 10^{-3}. The values of weight for each factor can be calculated:

$$\begin{aligned}&\text{Weight (Lith, Con, SG, SA, Elev, SP, VC, Fl, Fr)}\\&\quad = (0.1020, 0.1650, 0.1473, 0.1405, 0.1028, 0.1481, 0.0083, 0.1318, 0.0542)\end{aligned} \tag{9}$$

With the trained network, the other 3 records are assessed and compared with the actual results (Table 6).

Table 5 Relative probability estimated by the training process

ID	Lith	Con	SG	SA	Elev	SP	VC	Fl	Fr	Z	Z'	Status
1	0	0.26	0	0.16	0.14	0	0.23	0	0	0.79	0.37	Stable
2	0	0.26	0	0.14	0.17	0	0.23	0	0	0.80	0.38	Stable
3	0	0.26	0	0.16	0.01	0	0.11	0	0	0.54	0.26	Stable
5	0	0.26	0	0.15	0.14	0.02	0.25	0	0	0.82	0.39	Stable
6	0	0.26	0	0.14	0.01	0.02	0.11	0	0	0.54	0.26	Stable
7	0.07	0.58	0.03	0.09	0.05	0.05	0.11	0.25	0	1.23	0.55	Unstable
8	0.04	0.58	0	0.09	0.05	0	0.23	0	0	0.99	0.46	Stable
9	0.04	0.16	0	0.16	0.08	0	0.11	0	0	0.55	0.27	Stable
11	0.17	0.58	0	0.16	0.22	0.38	0.07	0.3	0	1.88	0.74	Unstable
12	0.15	0.58	0	0.16	0.14	0.36	0.07	0	0	1.46	0.62	Unstable
13	0	0.58	0	0.09	0.08	0	0.23	0	0	0.98	0.45	Stable
14	0.24	0.26	0	0.08	0.01	0.38	0.11	0	0	1.08	0.49	Unstable
15	0	0.26	0	0.15	0.17	0.38	0.11	0.12	0	1.19	0.53	Unstable
16	0	0.26	0	0.14	0.14	0.38	0.11	0.18	0.36	1.57	0.66	Unstable
17	0	0.58	0	0.14	0.01	0.38	0.23	0.06	0	1.40	0.60	Unstable
18	0.17	0.58	0	0.09	0.14	0.36	0.11	0.24	0	1.69	0.69	Unstable
20	0.15	0.58	0.03	0.08	0.17	0.15	0.07	0	0	1.23	0.55	Unstable

Table 6 The evaluated results of the tested samples

ID	Lith	Con	SG	SA	Elev	SP	VC	Fl	Fr	Estimated results	Actual results
4	0	0.26	0	0.17	0.08	0	0.11	0	0	0.27	Stable
10	0.11	0.26	0	0.09	0.17	0.05	0.11	0	0	0.46	Stable
19	0.11	0.58	0	0.08	0.01	0.38	0.11	0.25	0	0.61	Unstable

5.1.3 LS mapping and analysis

The output data show the LS indices values of the pixels. The higher the LS value is, the more susceptible a pixel to the occurrence frequencies of landslides is. Based on the 1 km $\times$ 1 km grid, the LS map of China is then produced by the ANN learning and multiplying the weighted factor layers (Fig. 8). As shown in Fig. 8, the southwest of China has the higher susceptibility indices than those in the other areas, which could also be consistent with landslide reports from the news.

The values Z' of output results in Table 5 are normalized between 0 and 1, and Table 7 shows the normalized values in five categories using the 0.2 intervals. LS map in China then has the following classifications (Fig. 9a): 1—very low; 2—low; 3—moderate; 4—high; 5—very high susceptibility. High-susceptibility regions of China account for 4.15 % (4—high plus; 5—very high), while low-susceptibility (1—very low plus, 2—low) areas account for more than 77 %.

Figure 9a demonstrates the hot spots of the high landslide potential provinces, where Yunnan, Guizhou, Fujian, Taiwan, parts of Zhejiang, Sichuan, Guangxi, Guangdong, Jiangxi, Shaanxi, Hubei, Hunan, and Hainan are revealed to be landslide-prone areas. Many

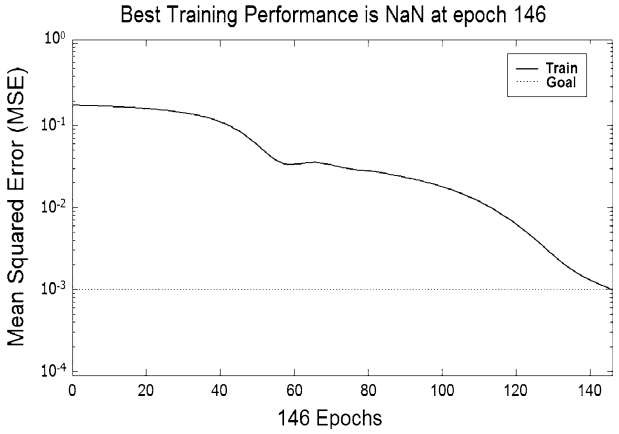

Fig. 7 Error and training times of ANN

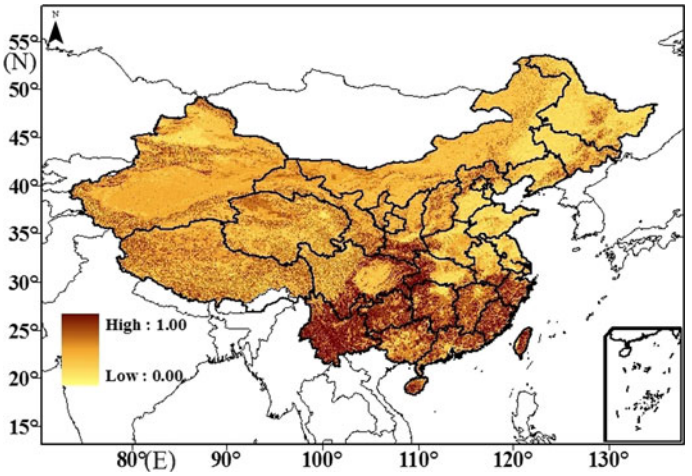

Fig. 8 LS map of China based on multi-factors

exterior factors including rainfall, earthquake or flood and so on induce the occurrences of landslides in these regions.

A 2.5-min grid global map of landslide hazard distribution by rainfall is available at (http://sedac.ciesin.columbia.edu/data/set/ndh-landslide-hazard-distribution/metadata). The hazard mapping incorporates a range of data on March 29, 2005, including slope gradient, soil property, soil moisture conditions, precipitation, seismicity, and temperature (Nadim et al. 2006). As shown in Fig. 9b, the rainfall landslide hazard areas mostly belong to the high-susceptibility areas (parts of Sichuan, Yunnan, Taiwan, and Zhejiang), and there are some small-scale landslides in parts of Tibet and Xinjiang. The map reflects China's LS locations and provides a useful method for studying and evaluating landslide occurrences.

Table 7 The LS classification

ID	LS value	Level	Percentage (%)
1	0.00–0.20	Very low	29.61
2	0.20–0.40	Low	47.68
3	0.40–0.60	Moderate	18.56
4	0.60–0.80	High	4.04
5	0.80–1.00	Very high	0.11

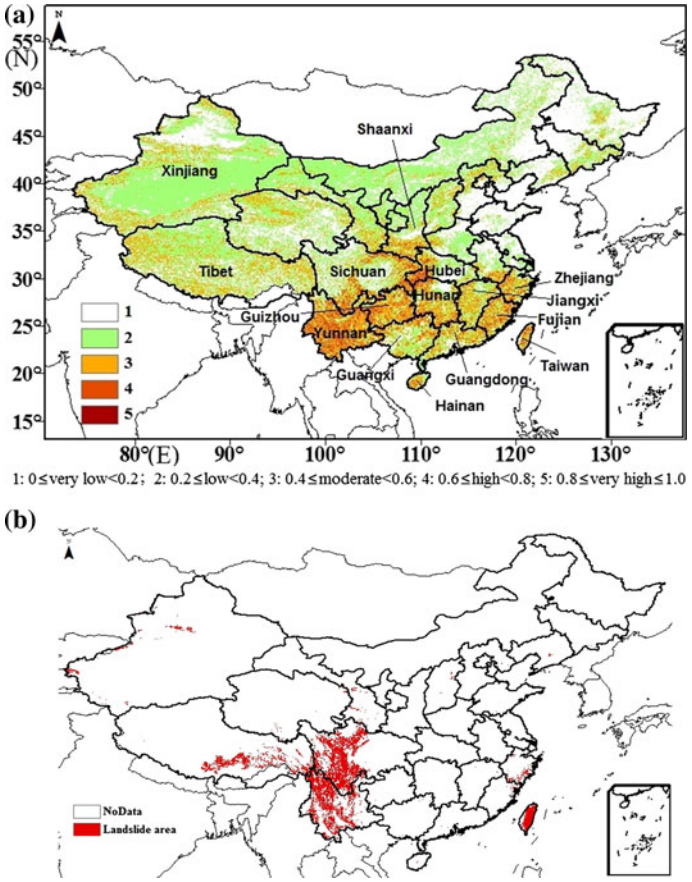

Fig. 9 **a** Chinese landslide susceptibility classification map derived from surface multi-geospatial data; **b** China's landslide hazard hotspot map based on the NGI's work in 2005 (Nadim et al. 2006)

5.2 LS accuracy test

Many papers assigned the weights subjectively caused inaccuracy (Hong et al. 2007a, b; Günther et al. 2012). The ANN methodology solves the problem of factor weights selection in a nonlinear constitutive relation. Therefore, the quality of the LS map obtained relies heavily on accuracy and scale of occurrence frequencies of different input factors and

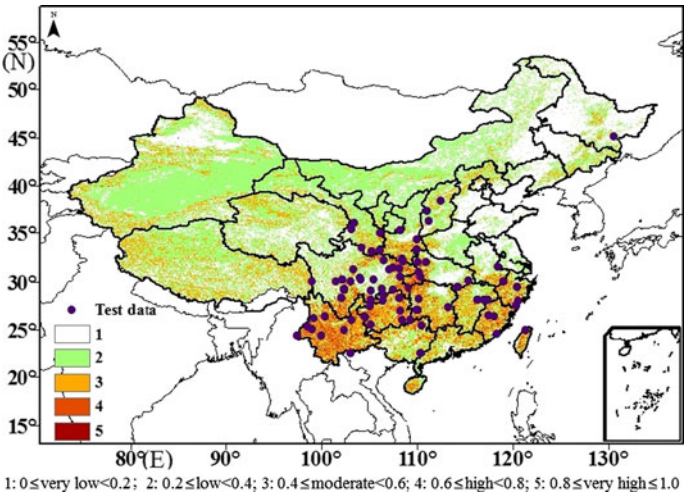

Fig. 10 The overlay of the landslide test data on the LS map

Table 8 LS map accuracy evaluation with 112 worldwide check data

	The number of landslide events	Percentage (%)	The areas in LS map
	3	2.68	Very low
	6	5.36	Low
	5	4.46	Moderate
	44	39.29	High
	54	48.21	Very high
Total	112	100.00	

sample data derived from the landslide inventory. We devise the following method to test the accuracy. If the result does not meet the requirements, we will return to reselect the sample data uniformly to retrain the ANN.

Some of the landslide tested data are from authoritative worldwide network databases (United States Geological Survey (USGS), http://landslides.usgs.gov/recent; EM-DAT International Disaster Database, http://www.em-dat.net/; International Consortium on Landslides Website (ICL), http://iclhq.org/). Landslide tested data of 112 events, triggered mainly by earthquakes, rainfall, and weathering, are gathered to compare with LS classification map (Fig. 10).

Table 8 shows LS map accuracy evaluation with 112 worldwide check data. The number of high- and very high-susceptibility regions account for 88.39 %, revealing that the accuracy of ANN models meet landslide study demands. Meanwhile, 9 records are spread around low-susceptibility areas, mainly because of the above-mentioned landslide data uncertainty in Sect. 3.2.

6 Conclusion

The diversity, fuzziness, uncertainty, and randomness features of landslides occurrences bring great difficulty in data processing. The processing capability of traditional methods is

limited and results in much loss of useful information, degrading the reliability in results. A powerful nonlinear mapping capability allows to simulate the possibilities of landslides and the relationships among complex factors (Chacón et al. 2006).

A LS model is developed for the whole country of China by applying a scoring system with a set of relevant factors based on BP ANNs. The BP ANNs merge nonlinear elements by qualitative and quantitative indices. The results are tested by comparing with landslide data reported by authoritative worldwide Web sites, and the accuracy of the network is approximately 88.39 % and satisfies landslide study requirements in China. The high-susceptibility areas, mainly spreading over the southwest of China, are characterized by the presence of landslide-prone sedimentary rocks, high seismicity, frequent severe earthquakes and rainfall events, and significant human activities.

The LS evaluation and mapping of China in our study constitutes a preliminary step toward detailed susceptibility and hazard research, is a useful method for hazard assessment, and provides a theoretical basis for predicting and forecasting landslide disasters throughout China.

Acknowledgments The work described in this paper is funded by National Basic Research Program of China (Project No. 2013CB733204), National Natural Science Foundation of China (No.41201380) and Key Laboratory of Advanced Engineering Surveying of NASMG (TJES1010), and is also supported by the Center of Spatial Information Science and Sustainable Development Applications, Tongji University.

References

Bălteanu D, Chendeş V, Sima M, Enciu P (2010) A country-wide spatial assessment of landslide susceptibility in Romania. Geomorphology 124(3):102–112

Braspenning PJ, Thuijsman F, Weijters AJMM (1995) Artificial neural networks: an introduction to ANN theory and practice. Springer, Berlin

Chacón J, Irigaray C, Fernandez T, El Hamdouni R (2006) Engineering geology maps: landslides and geographical information systems. B Eng Geol Environ 65(4):341–411

Chauhan S, Sharma M, Arora M, Gupta N (2010) Landslide susceptibility zonation through ratings derived from artificial neural network. Int J Appl Earth Obs 12(5):340–350

Che VB, Kervyn M, Suh C, Fontijn K, Ernst G, del Marmol MA, Trefois P, Jacobs P (2012) Landslide susceptibility assessment in Limbe (sw Cameroon): a field calibrated seed cell and information value method. Catena 92:83–98

Choi J, Oh H-J, Lee H-J, Lee C, Lee S (2012) Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using aster images and GIS. Eng Geol 124:12–23

Dai F, Lee C, Ngai Y (2002) Landslide risk assessment and management: an overview. Eng Geol 64(1):65–87

Dunning SA, Mitchell WA, Rosser NJ, Petley DN (2007) The Hattian Bala rock avalanche and associated landslides triggered by the Kashmir earthquake of 8 October 2005. Eng Geol 93(3–4):130–144

Ercanoglu M, Gokceoglu C (2004) Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (west black sea region, turkey). Eng Geol 75(3):229–250

Fernandes AM, Utkin AB, Lavrov AV, Vilar RM (2004) Development of neural network committee machines for automatic forest fire detection using lidar. Pattern Recogn 37(10):2039–2047

Ghosh S (2011) Knowledge guided empirical prediction of landslide hazard. ITC, Faculty of Geo-Information Science and Earth Observation, Enschede

Gong P (1996) Geological mapping. Photogramm Eng Rem Sens 62(5):513–523

Günther A, Reichenbach P, Malet J-P, Eeckhaut M, Hervás J, Dashwood C, Guzzetti F (2012) Tier-based approaches for landslide susceptibility assessment in Europe. Landslides

Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central Italy. Geomorphology 31(1):181–216

He KQ, Li XR, Yan XQ, Dong G (2008) The landslides in the three gorges reservoir region, china and the effects of water storage and rain on their stability. Environ Geol 55(1):55–63

Henderson LJ (2004) Emergency and disaster: pervasive risk and public bureaucracy in developing nations. Public Org Rev 4(2):103–119
Hong Y, Adler R, Huffman G (2007a) Use of satellite remote sensing data in the mapping of global landslide susceptibility. Nat Hazards 43(2):245–256
Hong Y, Adler RF, Huffman G (2007b) An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets. IEEE T Geosci Remote 45(6):1671–1680
Iwahashi J, Kamiya I, Yamagishi H (2012) High-resolution dems in the study of rainfall-and earthquake-induced landslides: use of a variable window size method in digital terrain analysis. Geomorphology 153–154:29–38
Kumar K, Sati D, Goyal N, Mathur S (2005) Landslide hazard potential analysis using GIS, Patalganga Valley, Garhwal, Western Himalayan region of India. Eur J Sci Res 45(3):346–366
Li C, Ma T, Zhu X, Li W (2011) The power–law relationship between landslide occurrence and rainfall level. Geomorphology 130(3):221–229
Manunta P, Brugioni M, Casagli N, Colombo D, Deflorio AM, Farina P, Ferretti A, Goutier E, Graf K, Haeberle J, Lateltin O, Meloni E, Mayoraz R, Montini G, Moretti S, Paganini M, Palazzo F, Spina D, Sulli L, Strozzi T (2011) Slam, a service for landslide monitoring based on eo-data. http://earth.esa.int/workshops/fringe03/participants/485/paper_paper_manunta_SLAM.pdf
Mathew J, Jha V, Rawat G (2009) Landslide susceptibility zonation mapping and its validation in part of Garhwal lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. Landslides 6(1):17–26
Mondini A, Guzzetti F, Reichenbach P, Rossi M, Cardinali M, Ardizzone F (2011) Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. Remote Sens Environ 115(7):1743–1757
Montgomery DR, Dietrich WE, Heffner JT (2002) Piezometric response in shallow bedrock at cb1: implications for runoff generation and landsliding. Water Resour Res 38(12):1274
Moreiras SM (2005) Landslide susceptibility zonation in the Rio Mendoza Valley, Argentina. Geomorphology 66(1):345–357
Nadim F, Kjekstad O, Peduzzi P, Herold C, Jaedicke C (2006) Global landslide and avalanche hotspots. Landslides 3(2):159–173
Nefeslioglu HA, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Eng Geol 97(3):171–191
Park NW (2008) Geostatistical integration of different sources of elevation and its effect on landslide hazard mapping. Korean J Remote Sens 24(5):453–462
Parry S (2011) Geomorphological mapping: methods and applications. Elsevier, Amsterdam
Peruccacci S, Brunetti MT, Luciani S, Vennari C, Guzzetti F (2011) Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy. Geomorphology 139–140:79–90
Pradhan B (2011) Use of GIS-based fuzzy logic relations and its cross application to produce landslide susceptibility maps in three test areas in Malaysia. Environ Earth Sci 63(2):329–349
Pradhan B, Singh R, Buchroithner M (2006) Estimation of stress and its use in evaluation of landslide prone regions using remote sensing data. Adv Space Res 37(4):698–709
Qin S, Jiao JJ, Wang S (2002) A nonlinear dynamical model of landslide evolution. Geomorphology 43(1):77–85
Refice A, Capolongo D (2002) Probabilistic modeling of uncertainties in earthquake-induced landslide hazard assessment. Comput Geosci-UK 28(6):735–749
Sidle RC, Ochiai H (2006) Landslides: processes, prediction, and land use. American Geophysical Union, Washington, DC
Wan S (2009) A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map. Eng Geol 108(3):237–251
Wang Z, Fukao Y, Pei S (2009) Structural control of rupturing of the Mw7.9 Wenchuan Earthquake, China. Earth Planet Sci Left 279(1–2):131–138
Wang WD, Guo J, Fang LG, Chang XS (2012) A subjective and objective integrated weighting method for landslides susceptibility mapping based on GIS. Environ Earth Sci 65(6):1705–1714
Wibowo A, Kafle B, Kermani AM, Lam NTK, Wilson JL, Gad EF Damage in the 2008 china earthquake. In: Proceedings of Australian earthquake engineering society conference
Wu W, Sidle RC (1995) A distributed slope stability model for steep forested watersheds. Water Resour Res 31(8):2097–2110

Yalcin A (2008) GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. Catena 72(1):1–12
Zhang GP, Xu J, Bi BG (2009) Relations of landslide and debris flow hazards to environmental factors. Chin J Appl Ecol 20(3):653–658