Application of a hybrid model of neural networks and genetic algorithms to evaluate landslide susceptibility

H. B. Wang[1, 2], J.W. Li[1], B. Zhou[1], Z. Q. Yuan[1], Y.P. Chen[3]

[1]School of Civil Engineering and Mechanics, Huazhong University of Science & Technology, Wuhan 430074, P.R. China

[2] Hubei Key Laboratory of Control Structure, Huazhong University of Science & Technology, Wuhan 430074, P.R. China

[3] Wenhua College, Huazhong University of Science & Technology, Wuhan 430074, P.R. China

Abstract

In the last few decades, the development of Geographical Information Systems (GIS) technology has provided a method for the evaluation of landslide susceptibility and hazard. Slope units were found to be appropriate for the fundamental morphological elements in landslide susceptibility evaluation. Following the DEM construction in a loess area susceptible to landslides, the direct-reverse DEM technology was employed to generate 216 slope units in the studied area. After a detailed investigation, the landslide inventory was mapped in which 39 landslides, including paleo-landslides, old landslides and recent landslides, were present. Of the 216 slope units, 123 involved landslides. To analyze the mechanism of these landslides, six environmental factors were selected to evaluate landslide occurrence: slope angle, aspect, the height and shape of the slope, distance to rivers and human activities. These factors were extracted in terms of the slope unit within the ArcGIS software. The spatial analysis demonstrates that most of the landslides are located on convex slopes at an elevation of 101-150 m with slope angles from 136°-225° and 41°-60°. Landslide occurrence

Corresponding author. *Email address*: huabin@mail.hust.edu.cn. Tel: 8627-87541954.

was then checked according to these environmental factors using an artificial neural network with back propagation, optimized by genetic algorithms. A dataset of 120 slope units was chosen for training the neural network model, i.e., 80 units with landslide presence and 40 units without landslide presence. The parameters of genetic algorithms and neural networks were then set: population size of 100, crossover probability of 0.65, mutation probability of 0.01, momentum factor of 0.60, learning rate of 0.7, max learning number of 10000, and target error of 0.000001. After training on the datasets, the susceptibility of landslides was mapped for the land-use plan and hazard mitigation. Comparing the susceptibility map with landslide inventory, it was noted that the prediction accuracy of landslide occurrence is 93.02%, whereas units without landslide occurrence are predicted with an accuracy of 81.13%. To sum up, the verification shows satisfactory agreement with an accuracy of 86.46% between the susceptibility map and the landslide locations. In the landslide susceptibility assessment, ten new slopes were predicted to show potential for failure, which can be confirmed by the engineering geological conditions of these slopes. It was also observed that some disadvantages could be overcome in the application of the neural networks with back propagation, for example, the low convergence rate and local minimum, after the network was optimized using genetic algorithms. To conclude, neural networks with back propagation that are optimized by genetic algorithms are an effective method to predict landslide susceptibility with high accuracy.

Keywords: landslide; Geographical Information Systems; genetic algorithms; back propagation neural networks; susceptibility evaluation

1. Introduction

Landslides are local phenomena occurring in different geomorphic contexts; they can be triggered by a variety of mechanisms, such as earthquakes or rainfall, and some of the causes are not yet well known. Landslides cannot be predicted accurately; however, the susceptibility of a given area to landslides can be determined and depicted using hazard zonation. Various methods have been proposed to partition landscape for the purpose of landslide hazard assessment and zonation mapping, including grid cells, terrain units, unique-condition units, slope units and topographic units (Carrara et al.,1991; van Westen, 1994; Guzzetti et al. 1999; Chung and Fabbri, 2003). With respect to the landscape partitioning methods mentioned above, slope units can be resized according to the prevailing failure type and dimension, thereby partitioning a river basin into nested subdivisions: coarser for larger landslides and finer for smaller failures. Because a clear physical relationship exists between landslides and the fundamental morphological elements of a hilly or mountain region, namely, drainage and divided lines, the slope-unit technique seems appropriate for landslide susceptibility assessment.

A careful review of the concepts, principles, techniques and methodologies for landslide susceptibility evaluation reveals that the most commonly used methods are geomorphological hazard mapping, analysis of landslide inventories, heuristic or index-based methods, functional, statistically based models and geotechnical or physically based models (Guzzetti et al. 1999; Wang et al. 2005; Fell et al., 2008). Recently, probabilistic models such as frequency ratio and logistic regression methods

have been applied to evaluate landslide susceptibility and have been integrated with Geographical Information Systems (GIS) (Ayalew and Yamagishi, 2005; Chung, 2006; Dahal et al., 2008; Nefeslioglu, et al., 2008; Yilmaz, 2009; Bai, et al., 2010; García-Rodríguez and Malpica, 2010; Hasekiogulları and Ercanoglu, 2012). Due to geological complexity of slopes and self-organized system, however, many variables are involved in slope stability evaluation, which display a highly nonlinear relationship with evaluation results. Under the consideration of the nonlinear characteristics of the sliding process, artificial neural networks (ANNs) have thus been introduced to produce landslide susceptibility and hazard maps (Ercanoglu and Gokceoglu, 2002; Neaupane and Achet, 2004; Catani et al., 2005; Gómez and Kavzoglu, 2005; Kanungo et al., 2006; Nefeslioglu, et al., 2008; Nefeslioglua et al., 2011).

The main characteristics of ANNs dealing with quantitative and qualitative indices include large-scale parallel distributed processing, continuously nonlinear dynamics, collective computation, high fault-tolerance, self-organization, self-learning and real-time treatment (Rumelhart and McClelland 1986). It is worth noting that a neural network system is a processing device, implemented as an algorithm or in hardware, whose design is inspired by the design and the function of mammalian brains; they react to training data input in such a way to alter their initial state, and they learn using unconventional algorithms. Neural networks integrated with GIS may be an effective approach when dealing with landslide hazard assessments where meaningful outcomes are difficult to achieve by means of standard mathematical models. Because

artificial neural network models are adaptive and capable of generalization, they can handle imperfect or incomplete data and can capture nonlinear and complex interactions among the several variables of a system (Ermini et al. 2005; Melchiorre et al. 2008; Lee and Pradhan. 2011).

However, it was found that the slow training speed and difficulty in achieving a local minimum cannot be resolved in practical applications for the most commonly used back propagation neural networks (BPNN). To solve this problem, several methods were proposed to improve the training speed of networks, such as improving error functions and adjusting the studying rates. For this kind of solutions, the BPNN is likely to converge to a local solution, which may not be the global solution, with the random selection of initial weights. Thus, a global search algorithm was then introduced, e.g., evolutionary programming, simulated annealing or genetic algorithms (GAs). Among them, GA has mainly been used to search for the optimal solution in BPNNs due to its excellent global search ability (Holland 1975; Sexton and Gupta, 2000; Kesign, 2004; Madaeni, et al., 2010; Chen and Zeng, 2011). There are two main aspects of applying GA to BPNNs for finding global optima in complex problems: one is to optimize the weights of the network, and the other is to optimize the topological structure of the network.

After an overview on landslide susceptibility using ANNs, it was noted that the weights were randomly selected and that the optimization cannot be carried out for global searching. This paper thus proposes a hybrid model of a GA and BPNN to evaluate landslide susceptibility for the optimization of weights. In this study,

landslide inventory was mapped after detailed field investigation and interpretation from high-resolution imagery. The outline of the watershed polygons was mapped as the ridge line using DEM, and reverse DEM data were used to detect the valley line. The combined DEM and reverse DEM, slope units were then mapped within a commercial software as ArcGIS (. After the definition of slope units, environmental factors were analyzed for the presence of landslide occurrence, and a hybrid of GA and BPNN was developed to evaluate landslide susceptibility. In this method, the GA was used to search for the optimal or approximately optimal connection weights and thresholds for the networks, and then, using back-propagation learning rules and training algorithms, the final weights could be adjusted.

2. Study area

2.1 General information

The study area is the Changshougou Valley, which is located northwest of Baoji city in Shaanxi Province (Fig. 1). The elevation is approximately 600-700 m a.s.l., and the relative relief exceeds 100-350 m over the Changshougou valley. In this area, three main geomorphological units may be distinguished: dissected loess plateau (yuan), loess hills (mao), and loess ridges (liang). The lithology of the strata in these units is mainly Neogene argillites and fluvial deposits consisting of clayey silts and gravels and Quaternary loess. Due to the erosion of the Pliocene lacustrine basin by stream systems and the reactivation of folds and faults, a very unstable base was formed for the extensive Quaternary loess deposits, which acts as the underlying cause of the loess instability. Loess is known as a “problem soil”; although it can sustain nearly

vertical slopes when dry, it is susceptible to catastrophic failure on reaching certain critical moisture contents.

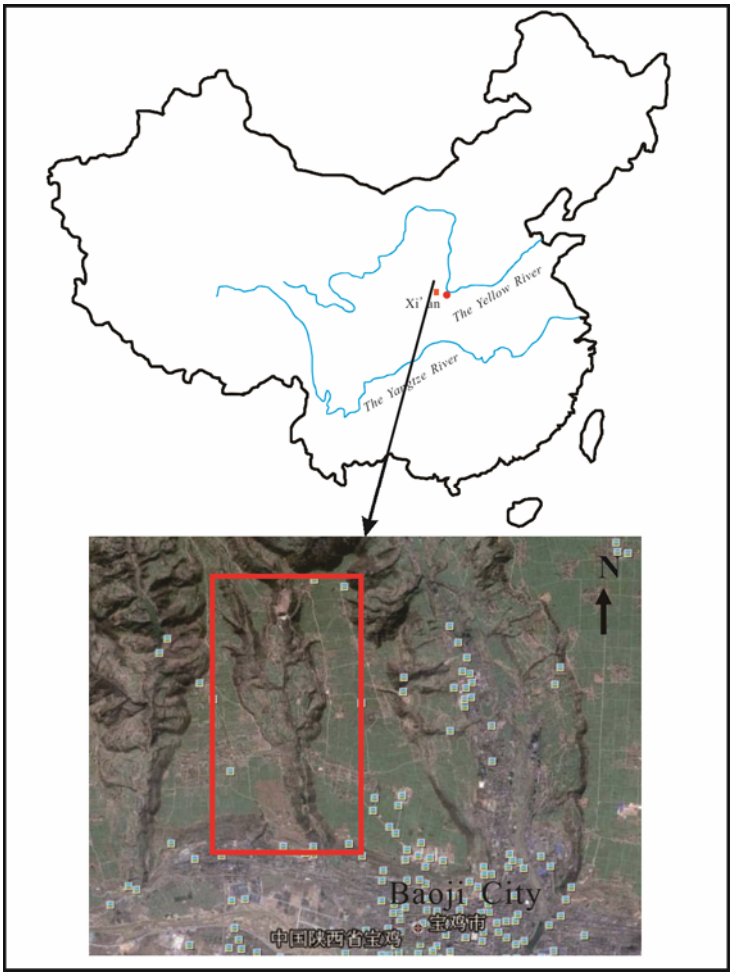


Fig. 1 Geographical location of the study area marked by the red rectangle （data from Google earth）

The climate of the study area is notable for Asiatic monsoons and marked seasonal shifts in dominant winds. The mean annual temperature is 12.9 °C, and the maximum is up to 41.6 °C. The mean annual precipitation reaches 679.1 mm, and almost 50% of the precipitation falls in the period from July to September.

2.2 Types of landslides

There are a variety of landslides present in the Changshougou valley. The

148 examination of the Weihe River terraces and their overlying loess units has shown that
149 landslides fall into three broad temporal categories, i.e., palaeo-landslides, old
150 landslides, and recent landslides. Palaeo-landslides occurred before or during the
151 Pleistocene. The slide masses consist of materials of Tertiary to Middle Pleistocene
152 age, in which the late Pleistocene loess (Malan) sometimes occurs as a drape.
153 Landslides of Late Holocene age involve materials of Late Pleistocene loess, which
154 were developed entirely within loess. This kind of landslides was common in the
155 study area. As a result of human activities, recent landslides were triggered or
156 reactivated from other kinds of landslides. The principal mass movement types
157 recognized in the study area are summarized in Table 1（Wang, et al., 2011).

158 **Table 1 Classification of landslide types in the study area**

Type of materials	
	Mixed
	Loess only
Geometry and movement	
	Flows and complex mass movements
	Large rotational mass movements
	Planar slides
	Debris flows and mudflows
Types based on situation of the failure plane	
	Bedrock-contact landslides
	Palaeosol-contact landslides
	Mixed landslides
	Landslides entirely within loess
Age	

Palaeo-slides
Old slides
Recent slides

As can be seen in Fig. 3, there are three basic types of landslides that occur in a variety of materials based on the situation of the failure plane.

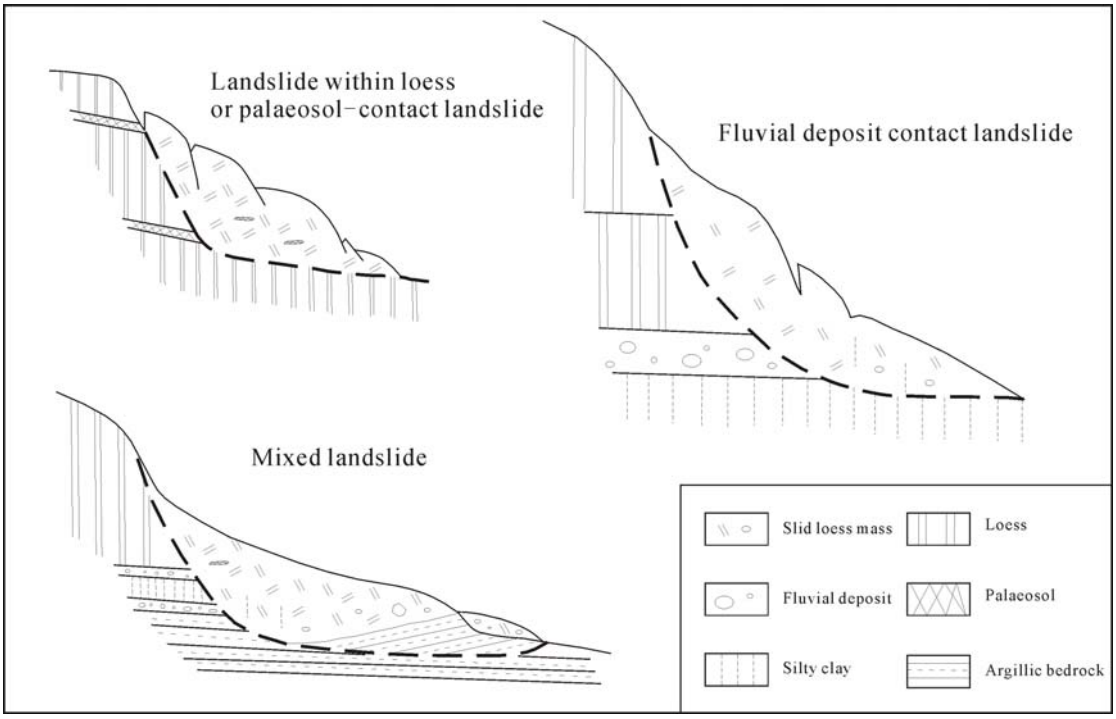


Fig. 2 Representative types of slope failures in the study area

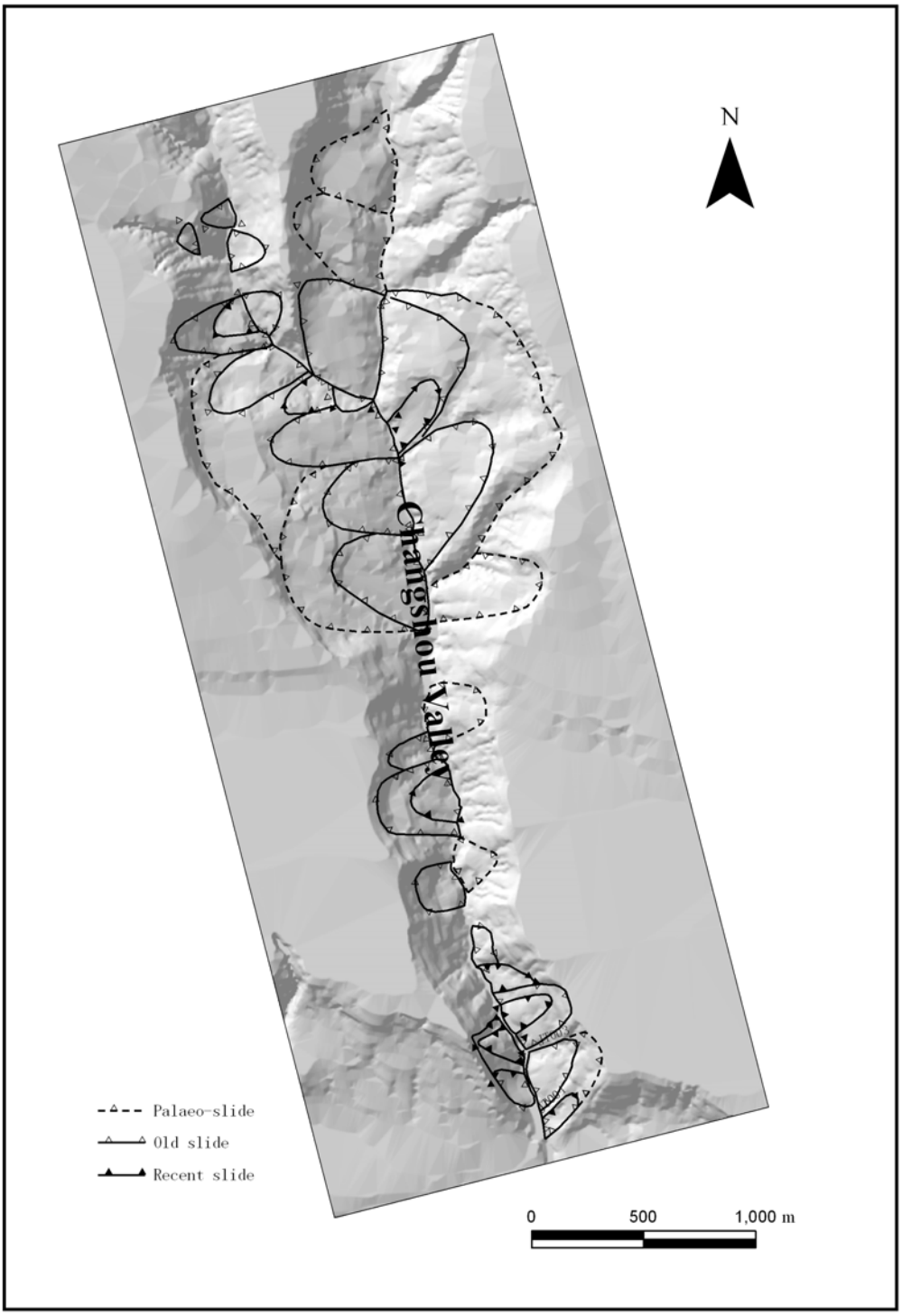
3. Data preparation

3.1 Landslide inventory

A landslide-inventory map was prepared primarily by the interpretation of high-resolution imagery and secondarily by site investigations. As shown in Fig. 3, 39 landslides in the study area are classified into three types: paleo-landslides, old landslides and recent landslides (Meng et al. 2000). Fifty percent of landslides are old as dormant-mature, and 25% have been reactivated recently. For all of the landslides, the slope after failure averages 22.6^{o}, with a maximum of 45^{o}. The slope

171 gradients of large-scale landslides tend to be gentle, at approximately 18°, which

172 implies that most of the paleo-landslides in the study area are stable.



173

174 **Fig. 3 Landslide inventory mapping in the Changshougou valley**

It was also observed that the landslides are concentrated at the confluence of two streams. As described in the previous section, these landslides are attributed to the undercutting of the slopes associated with gullying. Major factors affecting the initiation of slope failure and subsequent modes of movement appear to be dependent upon the morphology, the nature and degree of weathering of the underlying bedrock, and the moisture status of the loess deposits. Numerous ancillary factors include bedrock-loess interface, slope steepness, vegetation cover, and land utilization. As already noted, the upper surface of the Neogene bedrock is often weathered and contains variable amounts of smectites. In this zone, wetting-drying results in a progressive decrease in strength that may ultimately lead to slope failure. Undercutting of the slopes along the gullies is frequent, which increases the risk of slope failure. As a result of the high relative relief, the steep slopes and the relatively uniform geological and geomorphological conditions, the landslides tend to be very large and to occur in groups.

In terms of geological periods, landslides in the studied area fall into three categories: palaeo-slides, old slides, and recent slides. The most ancient landslides are not traceable in the historical records, and some recent landslides keep moving. Typically, a swarm of landslides is located at Zhuyuan village (Fig. 4), approximately 10 km north of Baoji. The landslide mass is 1000 m wide and 850 m long, with an estimated volume of 5.1×10^{7} m^{3}. Steeper slopes are generally over 40^{o} around the crest, with forested landslide terraces. Remote-sensing interpretations and field investigations indicate a maximum vertical displacement of 150 m along the slip surface of the

landslides. The shape and slope angle of the landslide scars suggest a single, major concave surface of rupture. For these studied landslides, the failure planes occurred either along the contacts between the fluvial deposits and the Neogene argillites or partially within the bedrock. With regard to the Neogene argillite in the loess plateau, the clay minerals are dominated by illite, chlorite, smectite and kaolinite, similar to the overlying Quaternary loess (Peng and Guo 2007). This type of argillite is thus subjected to long-term softening due to saturation in the contact zone. After a point is reached, at which they are no longer able to support the overlying loess, a progressive failure occurs. The occurrence of planar slides in the study area depends on the shear-strength conditions of the failure surface and the cohesion of the materials involved, usually presented in the Malan loess and in reworked loess slope deposits.

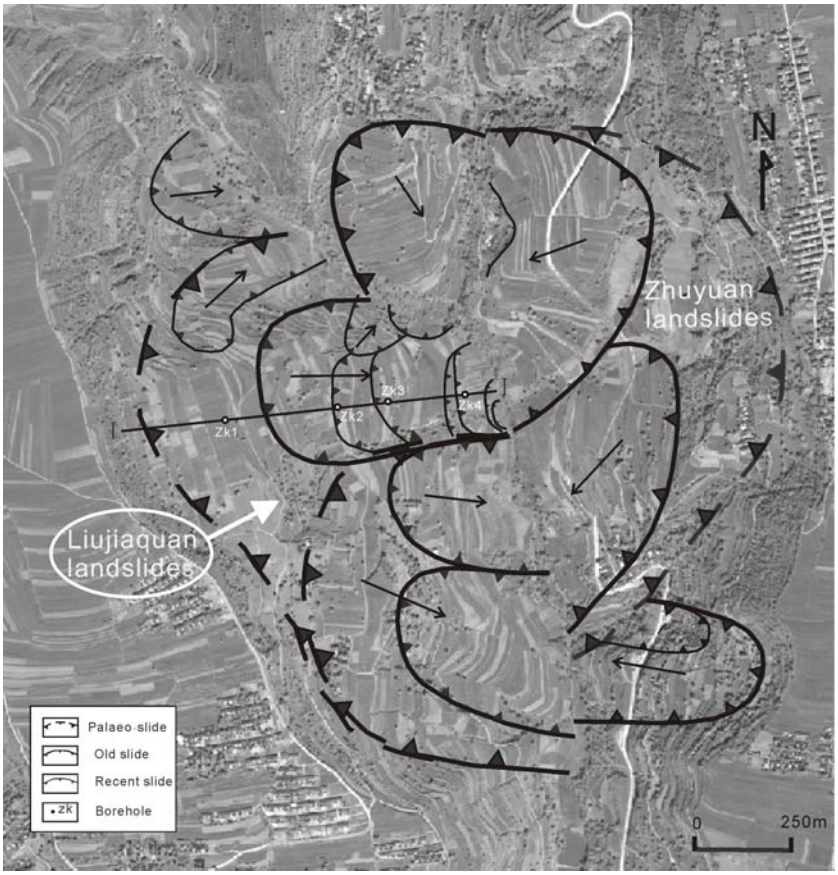
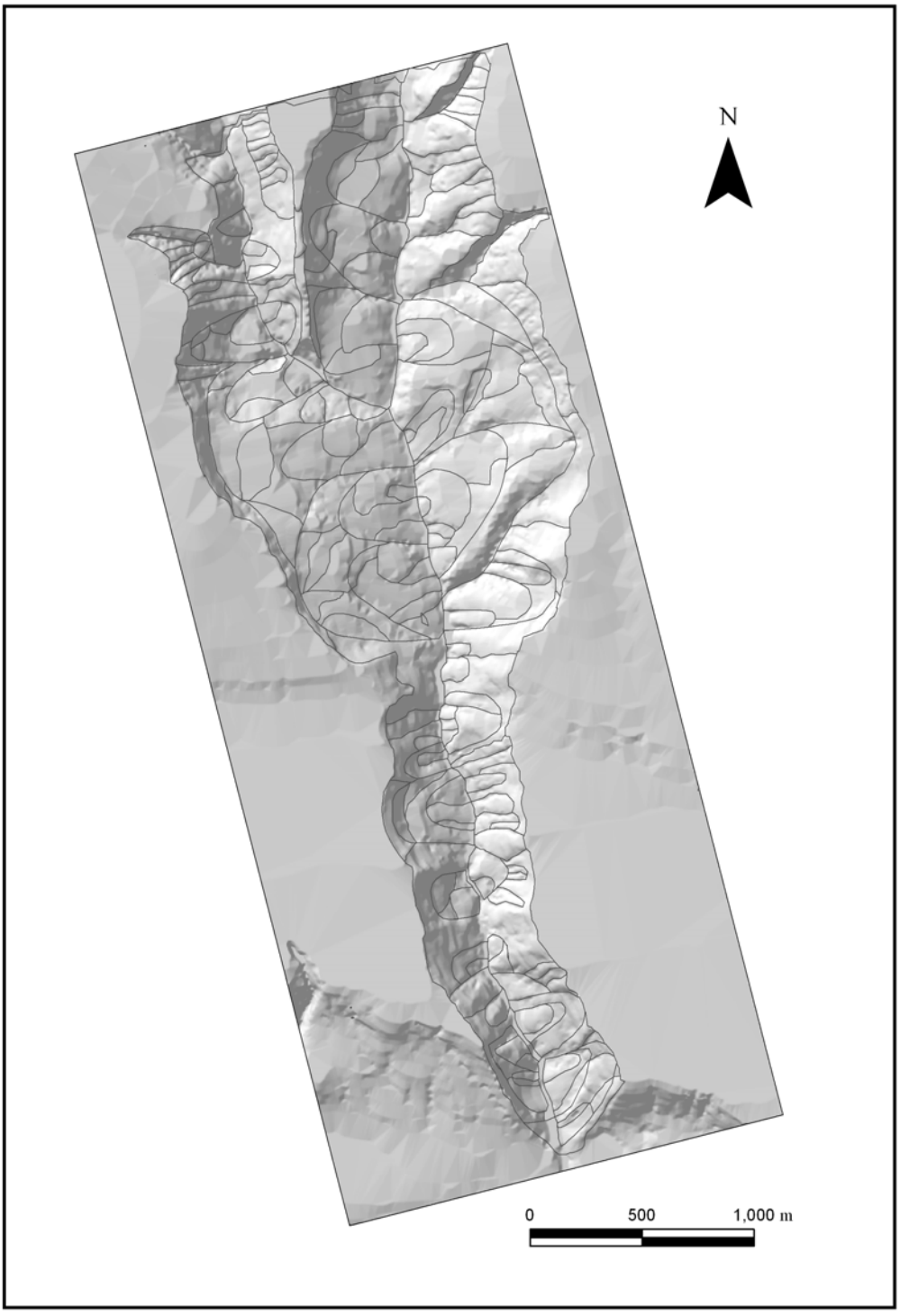


Fig. 4 View of the Zhuyuan and Lijiaquan landslides from a Quickbird image

3.2 Defining slope units

In this study, a GIS-based hydrologic analysis and modeling tool, Arc Hydro, is employed to draw the dividing lines for forming slope units automatically. Arc Hydro is an ArcGIS based software geared to support water resources applications (Maidment 2002). The software provides a method for the delineation of watersheds and stream networks using digital elevation models (DEM) of land-surface terrain. In the present study, the topographic maps were used to produce DEM with a contour interval of 10 m at a scale of 10,000. Using the DEM, the outlines of the watershed polygons are topographically mapped as the ridge lines, and the reverse DEM data can be used to detect the valley lines (Xie, et al., 2003). Using the DEM grid analysis, high DEM values can be turned into low values and low DEM values to high. After these values change, the original valley line can be turned into a ridge line. Meanwhile, the valley line can also be obtained by watershed analysis of the reverse DEM data. The combined DEM and reverse DEM analyses map the slope units within the of ArcGIS in the studied area (Fig. 5).



225

226 **Fig. 5 Slope units from combination of DEM and reverse DEM**

227 ## 3.3 Selection of environmental factors

228 In these slope units, a variety of environmental factors such as slope angle, aspect,

height and shape of slope were selected to evaluate landslide susceptibility. Two more environmental factors were selected, namely, distance to rivers and human activities. As for the distance to rives, the buffering zone was used to analyze the relationship between the distance to river and landsliding. It was shown that landslide frequency generally decreases with the increasing of the distance from drainage line increases, due to the gully erosion. In addition, the influence of human activity was analyzed under the consideration of land using on the slope area. After the detailed investigation, it is valued as 1 if the slope areas were used for vegetation, while the value is considered as 0 without any vegetation.

(1)Slope angle

Slope angle has a great influence on the susceptibility of a slope for landslide occurrence. As the slope angle increases, more of the load force is directed down the slope. The steeper slope, the more potential to landsliding. To quantify the relative frequency of landslides on different slope gradients, it is necessary to consider the distribution of the slope gradient categories using the available DEM dataset. Using the function of Mean in the ArcGIS Spatial Analyst model, an average slope angle was defined for each slope unit. In the study area, slope angles were categorized into six classes: 0-20°, 21°-30°, 31°-40°,41°-50°, 51°-60° and larger than 60° (Fig. 6).

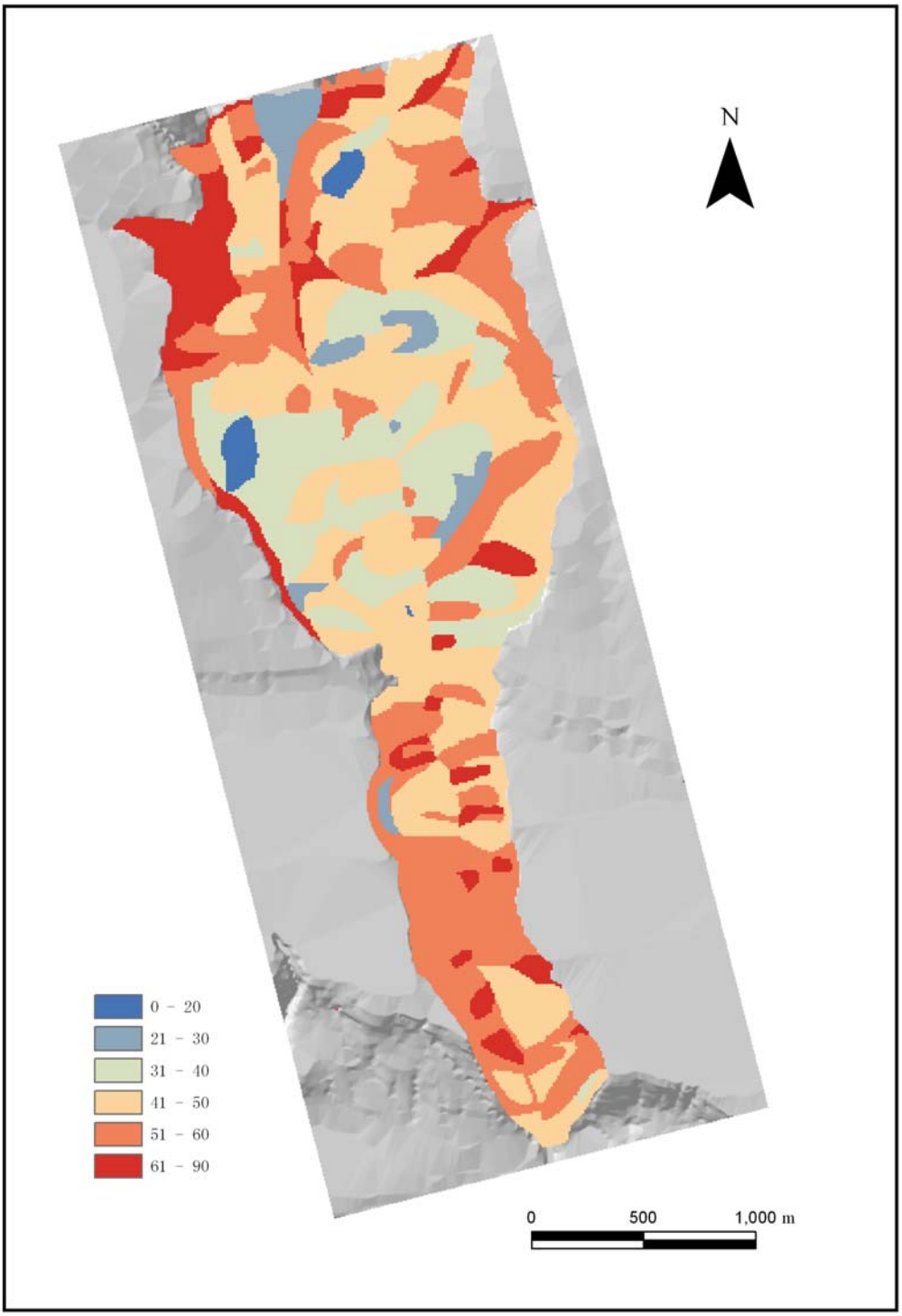


Fig. 6 Thematic map of the slope angle

(2) Slope height

Slope height plays an important role in landslide occurrence, especially in loess areas

(Derbyshire et al., 2000). Other conditions being equal, higher slopes can increase stress values in different areas of slope, making the slopes unstable. The slope height was classified into five classes: 0-50, 51-100 m, 101-150 m, 151-200 m and 201-300 m. Using Min and Max functions in the model of ArcGIS Spatial Analyst, the slope height was obtained from the difference of the minimum and maximum elevation over all the slope units.

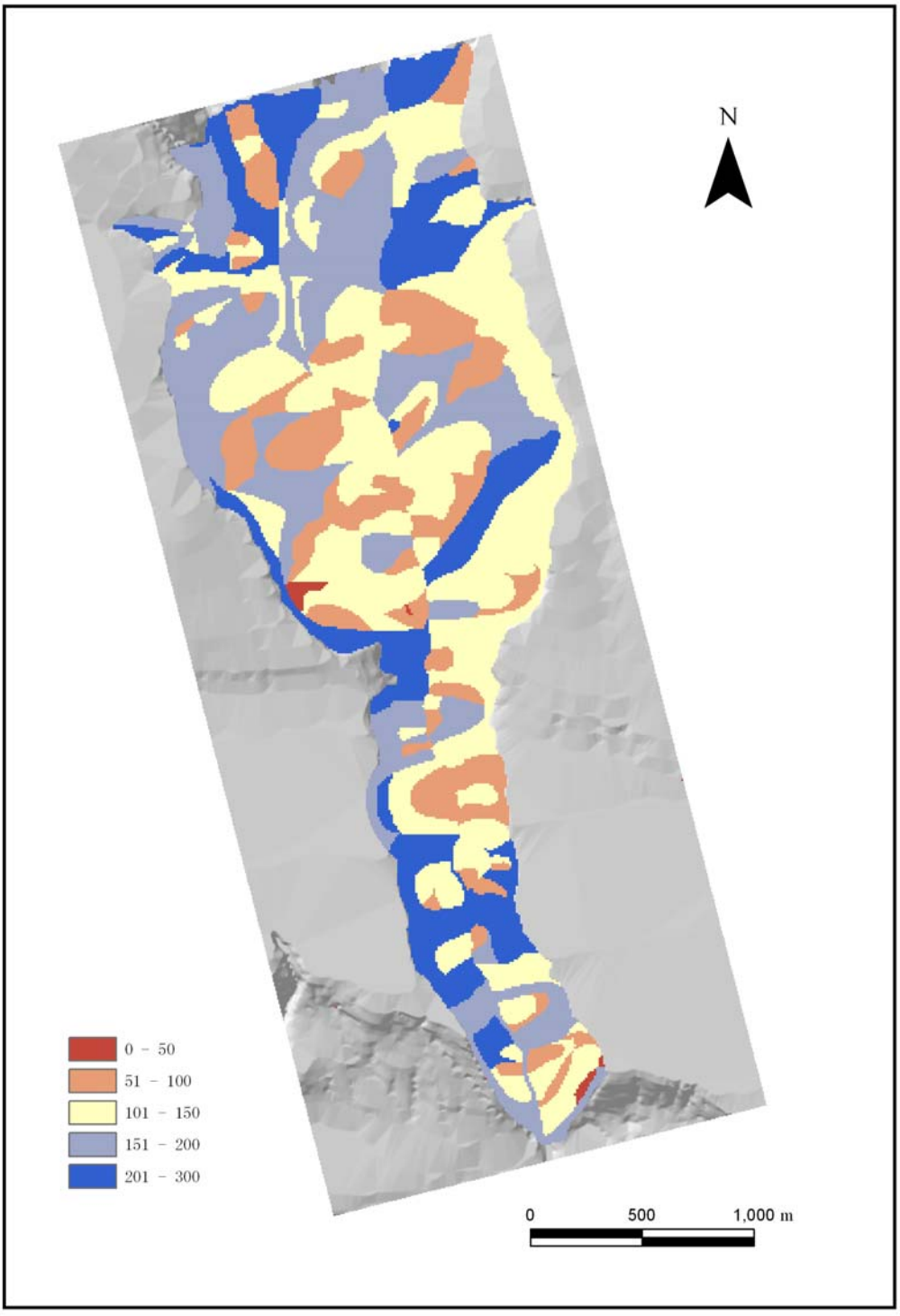


Fig. 7 Thematic map of slope height

(3) Slope aspect

Moisture retention and vegetation is mainly reflected by slope aspect, and then

influence landslide initiation, to some extent. To understand the relationship between slope aspect and landsliding, the slope aspect information was used from the Majority function within the ArcGIS. In the present study, it was categorized as 0-45, 46-90, 91-135, 136-180, 181-225, 226-270, 271-325 and 326-360 (Fig. 8).

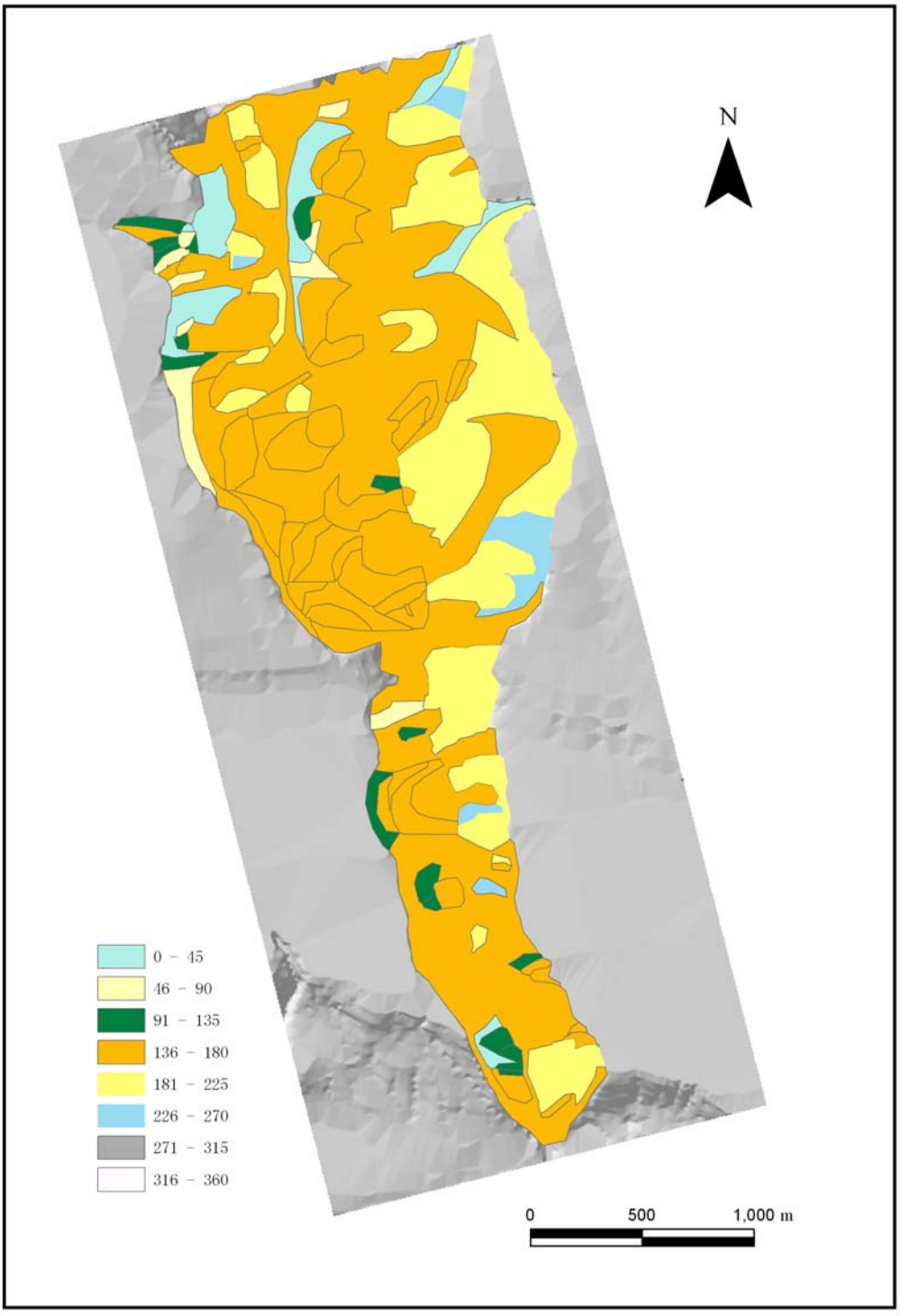


Fig. 8 Thematic map of slope aspect

(4) Slope morphology

It is known that water concentrates in concave topographic positions and makes the

slope susceptible to (sub) surface flow as the main hydrological triggering mechanism. According to the analysis of the relationships between the landslide occurrence and slope morphology, however, most of landslides are located in the convex topographic positions (Fig. 9). The slope morphology can be classified as plain, concave and convex in the study area.

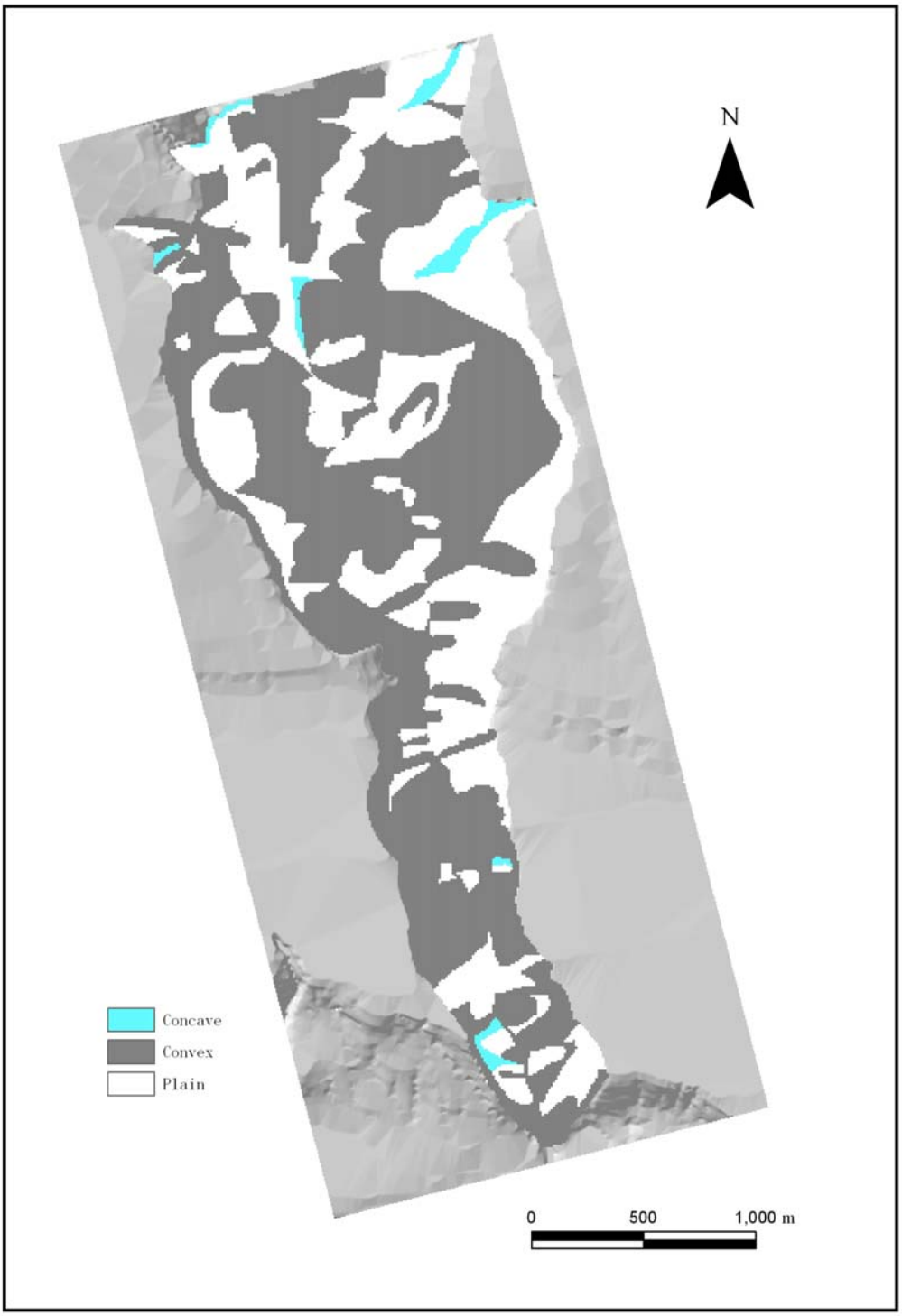


Fig. 9 Thematic map of the slope shape

4. Methods

4.1 Normalization of data

Input-output data for the GA-BPNN training and testing are from a database, including slope, aspect, elevation, shape of slope, distance to river and human activity. These sets are utilized to test network modeling. As the dimension and magnitude of the original sample data are different, the input and test data should be normalized before training, i.e.,

$$T = 2 \cdot \frac{X - X_{\min}}{X_{\max} - X_{\min}} - 1 \qquad \text{(Eq. 1)}$$

where X represents the original data, $X_{\max}$ and $X_{\min}$ are the maximum and minimum of original data, respectively. T is the target data after normalization.

4.2 GA-based BPNN

The BPNN is trained by repeatedly presenting a series of input/output pattern sets to the networks. The networks gradually learn the input/output relationships of interest by adjusting the network weights to minimize the error between the actual and predicted output patterns of the training sets (Fig. 9). After the learning process is completed, the network weight coefficients cannot be changed. In this model, the usage of networks with only forward calculations is needed in pattern recognition and prediction, and the calculation can be executed very quickly.

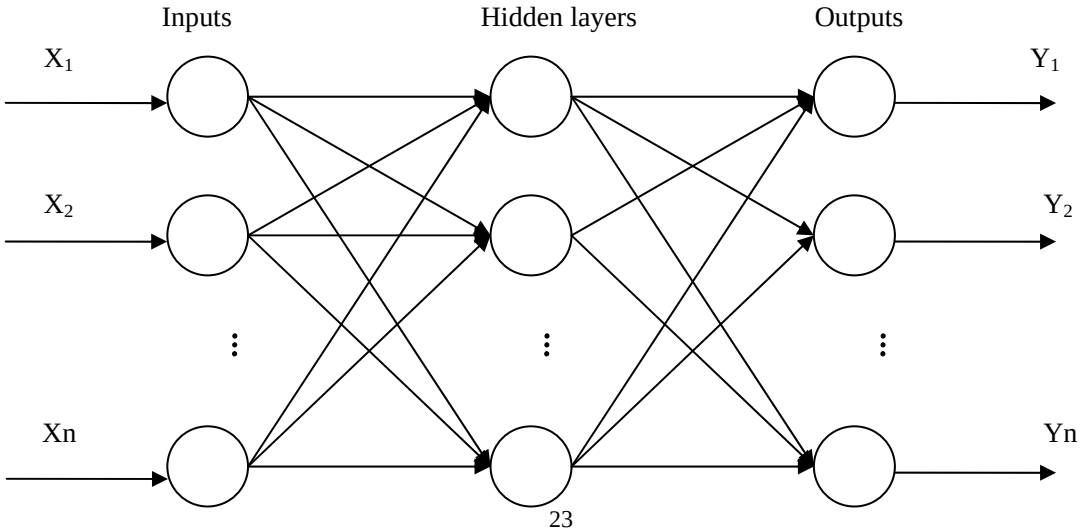


Fig. 10 Architecture of three-layer BPNN

The GA-based BPNN learning process consists of two stages (Fig. 11): employing GAs to search for the optimal or approximate optimal connection weights and thresholds for the networks and then using the back-propagation learning rules and training algorithms to adjust the final weights. The implementation procedure of the network training is programmed within Matlab using the GAs and Neural Networks Tool Boxes.

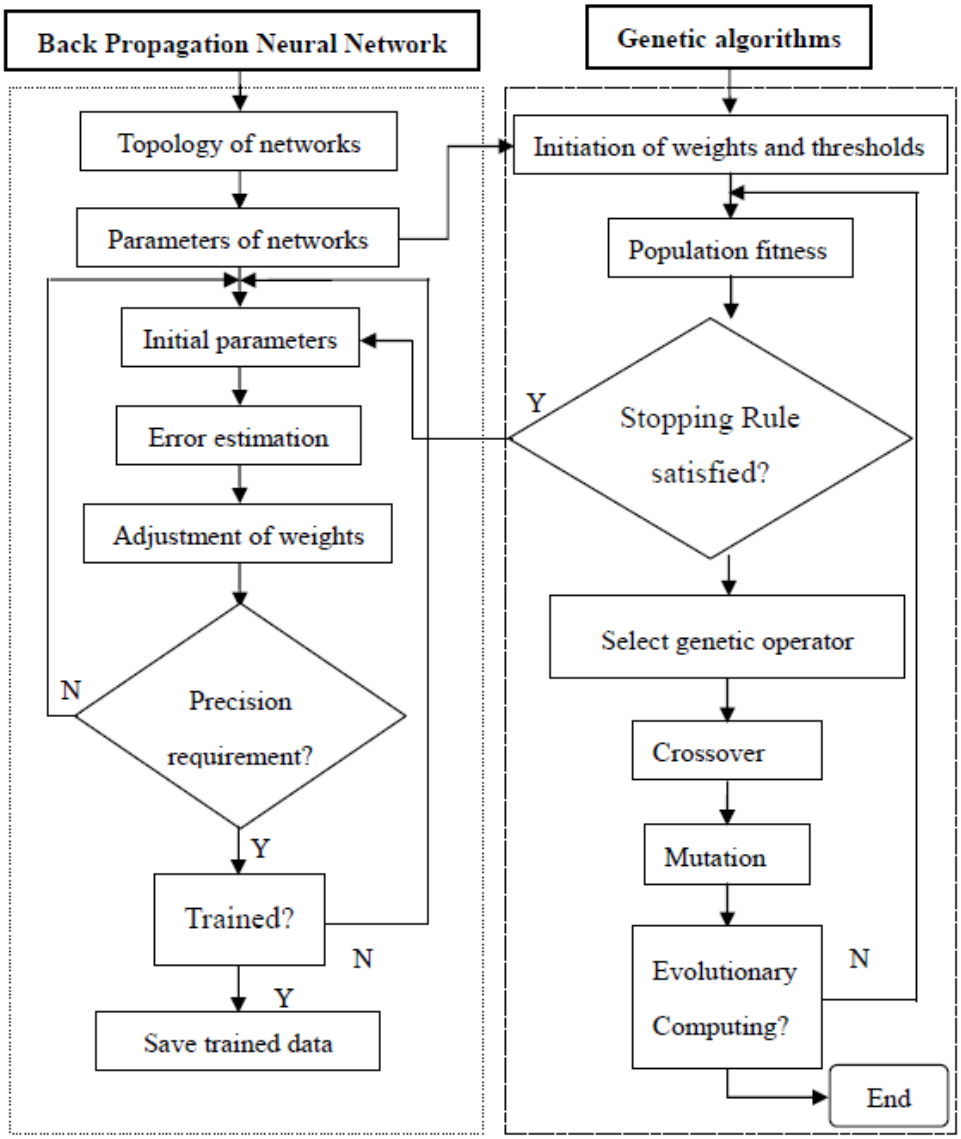


Fig. 11 Framework of GA-based BPNN

In this processing, the BPNN weights and thresholds are represented as genes in a chromosome, and the global optimum is then searched for using the selection, crossover and mutation operators of the genetic algorithm. This procedure is completed by applying a BP algorithm on the GA-established initial connection weights and thresholds. If the BP network's total mean square error is larger than the expected error, the weights and thresholds will be updated; otherwise, they are saved as initial value of the BP network training. To train the BPNN, the learning rate was adjusted to follow Eq. 2 using the momentum and the self-adapting methods, which was programmed using Matlab.

$$lr(k+1)=\begin{cases}1.05\cdot lr(k); mse(k+1)<mse(k)\\ 0.7\cdot lr(k); mse(k+1)>1.04mse(k)\\ lr(k); else\end{cases} \quad \text{(Eq. 2)}$$

where *lr* means adaptive learning rates, *mse* represents mean square errors for BPNN, and *k* is training time.

After the initiation of weights and thresholds was formed from the BPNN training, an initial population *S(0)* was randomly generate, because it is not known a priori where the globally optimal strings. In GAs, each member of *S(0)* is a string of length that corresponds to the problem coding. From this initial population, subsequent populations, *S(1)*, ..., *S(t)*, ..., will be computed by employing three genetic operators of selection (reproduction), crossover and mutation. Then, the fitness score of each individual string of the current population *S(t)* is computed, and the strings in the current population are copied and placed in the intermediate population, which is proportional to their fitness relative to other individuals in the population. Crossover

and mutation are applied to the intermediate population for the creation of the next population. The two new strings, called as offspring, are formed by the juxtaposition of the first part of one parent and the last part of the other parent. It can continue with the calculation of population fitness until some stopping criterion applies to find final population. Subsequently, they were further adjusted under the BP learning rule to the best result, by which the landslide susceptibility can be accurately predicted. The detailed framework is shown in Fig. 12 for landslide susceptibility using GA-based BPNN.

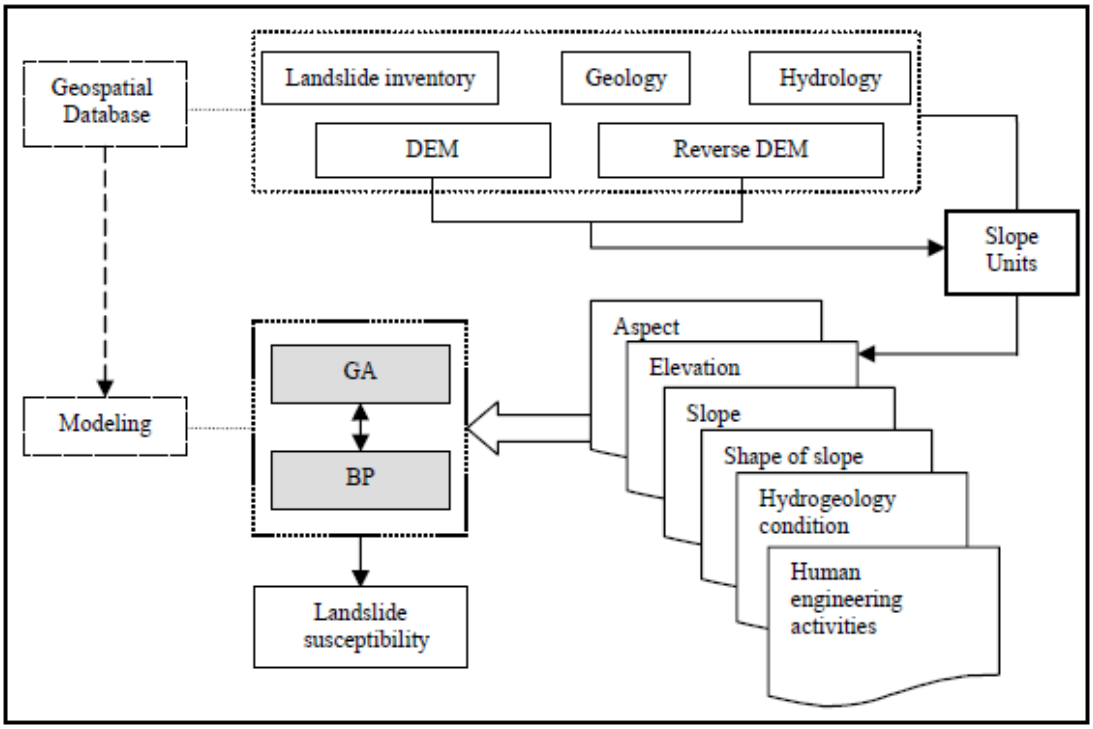


Fig. 12 Framework of landslide susceptibility using GA-based BPNN

Using BPNN modeling optimized by genetic algorithms, the parameters of GAs and neural networks are set in the present study. The population size is 100; crossover probability is 0.65; mutation probability is 0.01; momentum factor is 0.60; learning rate is 0.7; max learning number is 10000; and target error is 0.000001. From a

database of 216 landslides in units of slope, 120 landslides were randomly used for training the neural network models, and 96 landslides were used for the validation of landslide susceptibility. In the analysis of susceptibility, the slope unit affected by landsliding was valued as 1, while the unit without landslides was assigned as 0.

For the GA-improved BPNN, the network was trained for 488 times, and the error sum of the squares vs. the generation and fitness is shown in Figs. 13 and 14, respectively. In the figures, the blue line is the best solution trend, while the red one is the mean of the whole population. The final error is 9.96914e-007 from the relationship of the epochs and errors in Fig. 15. It can be demonstrated that the requirement was met for the BPNN training for landslide susceptibility assessment.

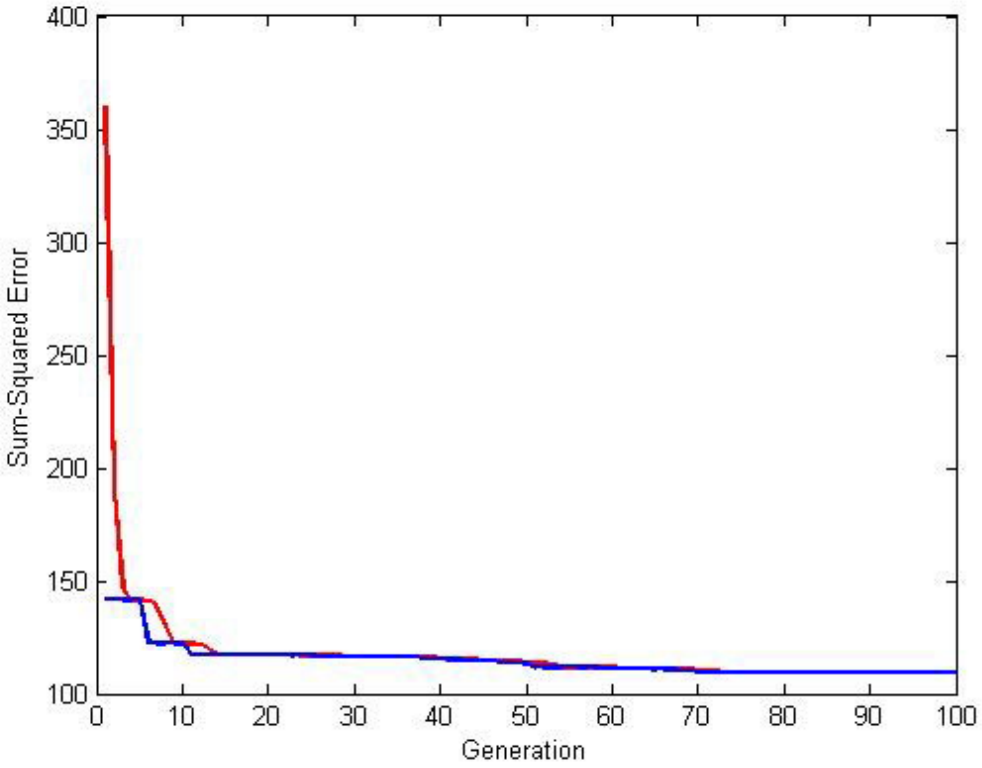


Fig. 13 The error sum of squares vs. generation

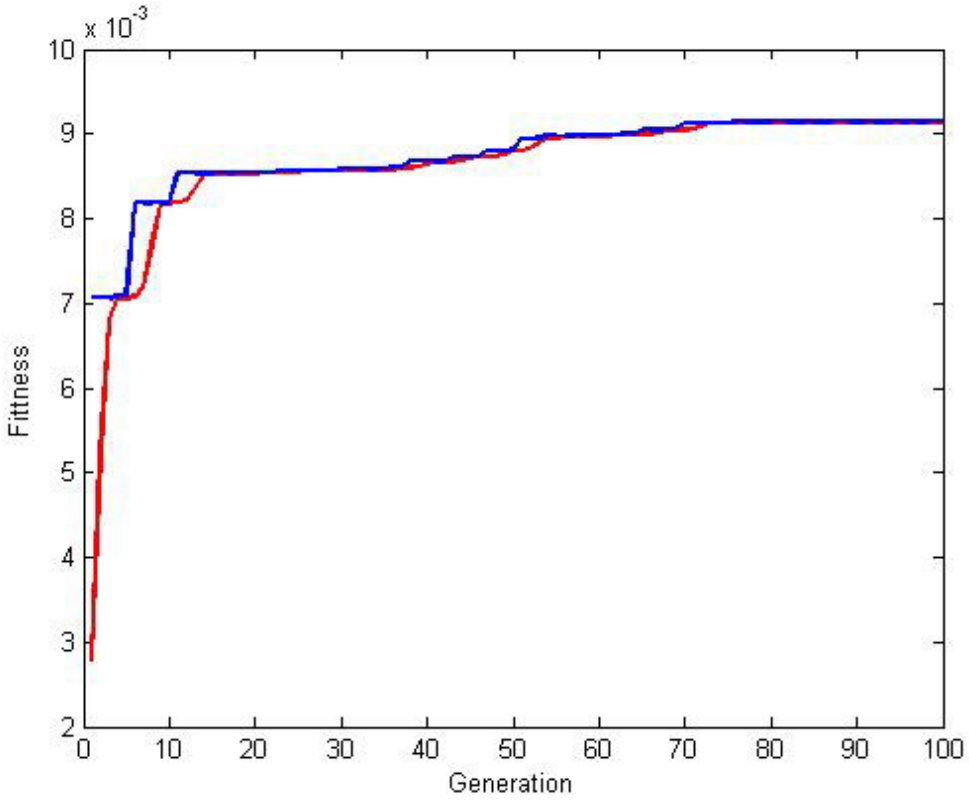


Fig. 14 Generation vs. fitness

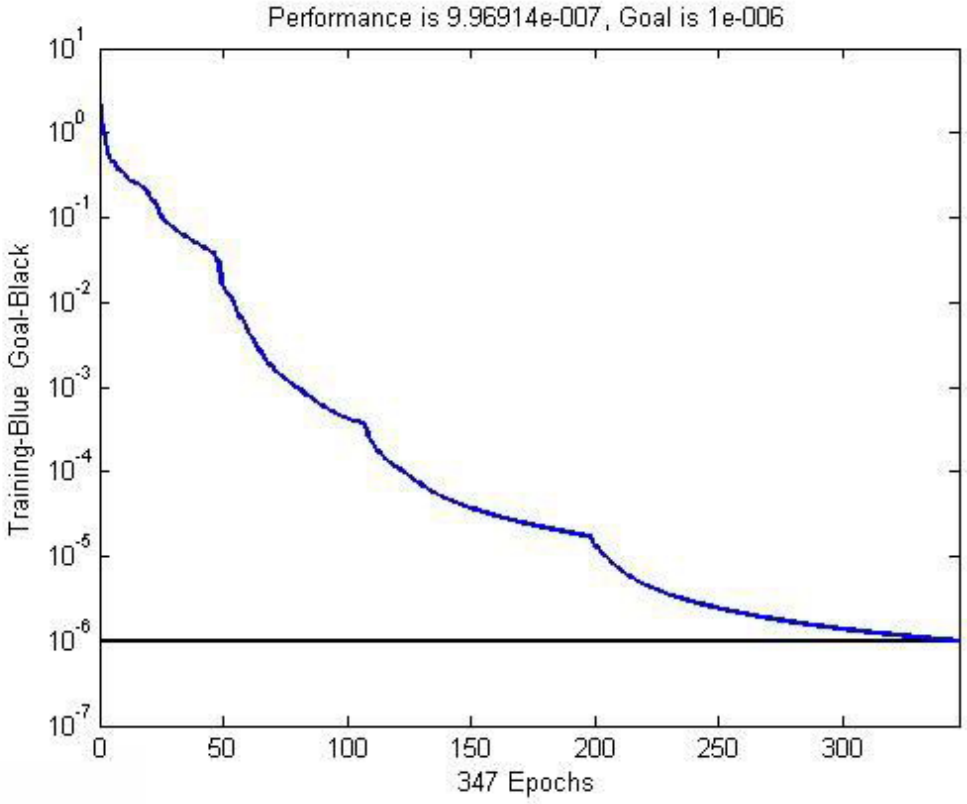


Fig. 15 The relationship between epochs and errors

5. Results

Following the process of mapping the slope units, the environmental factors can be categorized into slope, aspect, slope height, shape of slope, distance to rivers and human activities. These factors were statistically analyzed by slope units within the scope of GIS, after the construction of the geo-database. As mentioned above, the geological condition is almost the same in the study area, in which the strata in these units are mainly Neogene argillites, and the fluvial deposits consist of clayey silts and gravels and Quaternary loess. The environmental factors are analyzed by focusing on the slope parameters, such as slope angle, aspect, height and the shape of the slope.

According to the statistical analysis, 36% of the slope units affected by landslides occurred on slopes with angles between 41° and 50°, and 30% were occurred on slopes with angles from 51°-60° (Fig. 16). From the analysis of the relationships between the slope height and landslide occurrence (Fig. 17), almost 45% of the landslides occurred at elevations between 101 and 150 m (Fig. 18). It was also found that 38% and 56% of landslides were toward the southeast and southwest, respectively. However, other five variables made little contribution to the landslide occurrence. In addition, it was also demonstrated that the concave terrain was more stable, after large-scale landslides were observed to have occurred in the loess areas (Fig. 19).

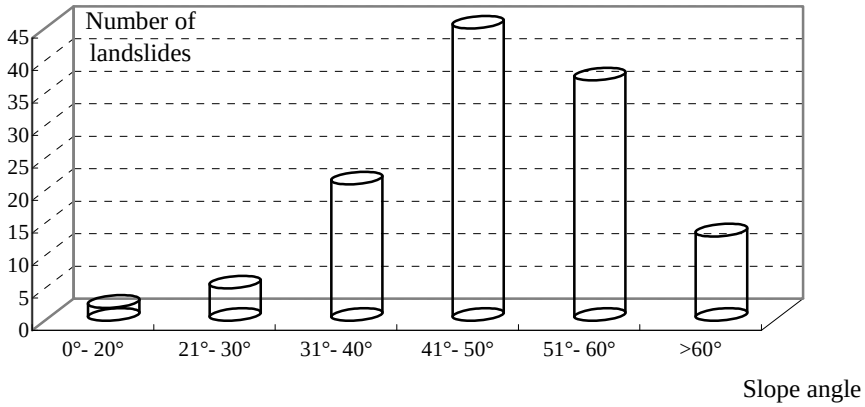


Fig.16 relationships between slope angle and landslide occurrence

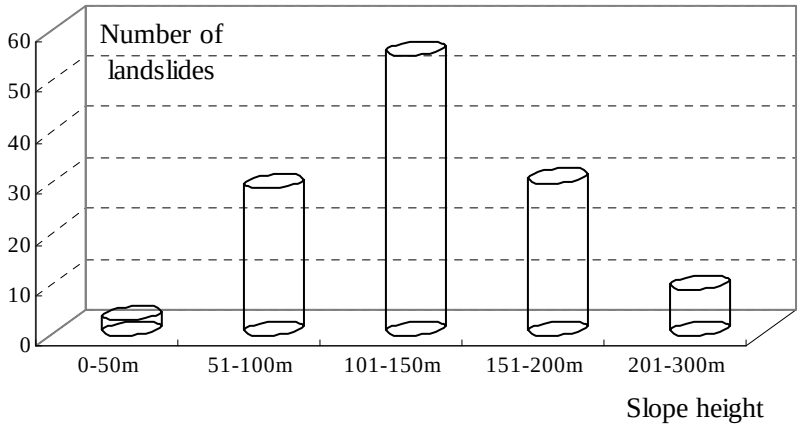


Fig.17 relationships between slope height and landslide occurrence

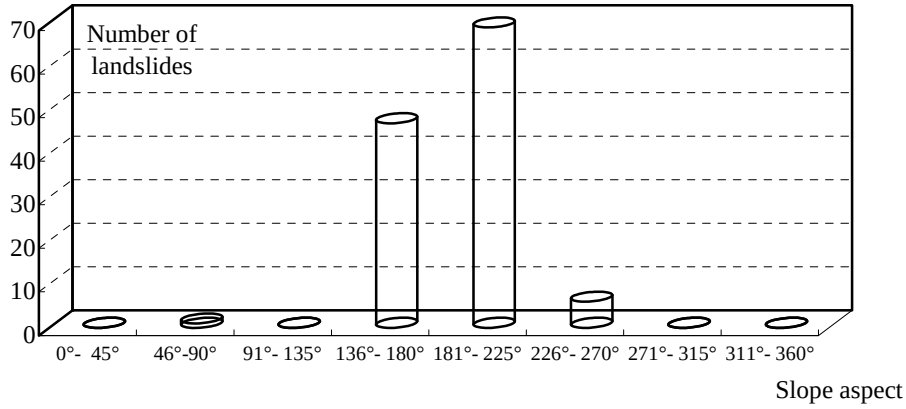


Fig.18 relationships between slope aspect and landslide occurrence

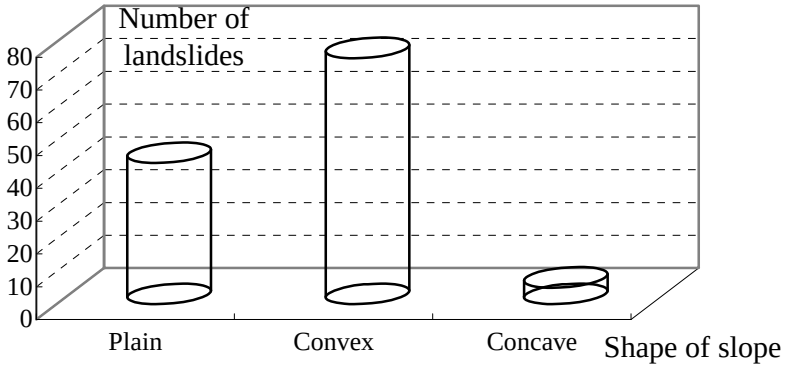
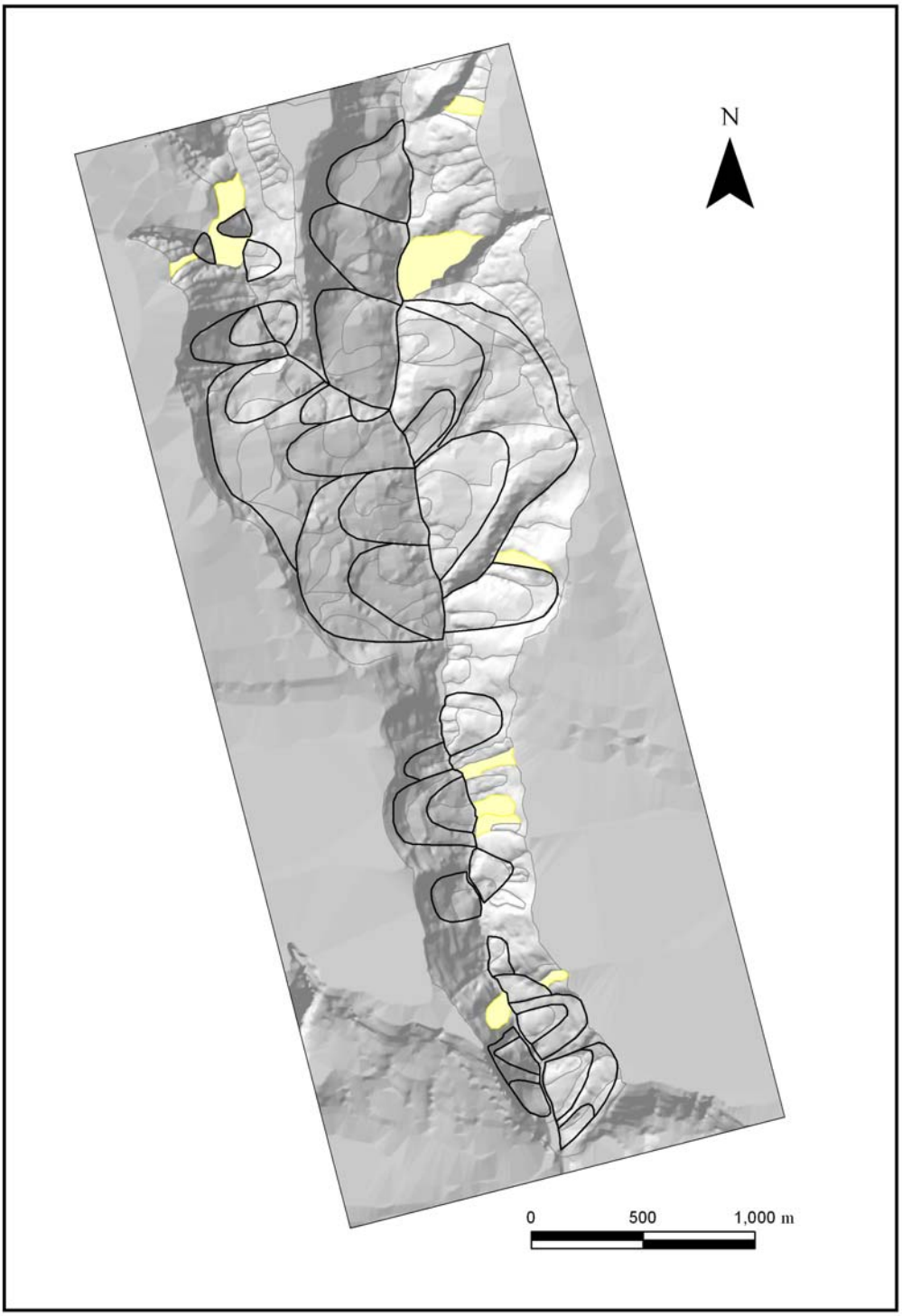


Fig.19 relationships between the shape of slope and landslide occurrence

After the GA-improved BPNN training, the susceptibility of landslides was predicted, as shown in Fig. 20. Comparing landslide occurrence with the susceptibility map, it was noted that the prediction accuracy of landslide occurrence is 93.02%, whereas the units without landslide occurrence is predicted with an accuracy of 81.13%. To sum up, the verification demonstrates satisfactory agreement with the accuracy of 86.46% obtained between the susceptibility map and landslide locations.



394

395 **Fig. 20 Map with predicted landslide susceptibility (slopes potential to landslides**

396 **in yellow)**

397 Ten slopes (in yellow in Fig. 20) were predicted to be prone to landslides. In view of

the environmental factors, most of these slopes were toward the southeast or southeast, were higher than 100 m and had slope angles greater than 43°. Furthermore, seven of the slopes were in convex topographic conditions, whereas the other slopes were located in concave topographic conditions.

6. Discussions and Concluding remarks

As a useful tool, which addresses a nonlinear system and is capable of response to inputs and adaptation to the environment, the most widely used BPNN is capable of evaluating landslide susceptibility at both the regional and site-specific scales (Lee et al. 2003; Neaupane and Achet 2004). The BPNN can be applied better over a wide area using non-parametric variables with large extensions. However, the BPNN is prone to falling into local extremes, and their convergence is slow. To overcome these drawbacks, a GA-based BPNN was proposed to optimize the neural network weights for landslide susceptibility assessment, and the topology was subjectively kept in three layers. In the optimization of the GA-based BPNN, the number of neurons can be decided in the hidden layer by the Kolmogorov theorem. For the training of the BPNN, the weights and thresholds were represented as genes of a chromosome, and the global optimum was then searched for using the selection, crossover and mutation operators of the genetic algorithm.

During the processing of GA-BPNN, it is very important for the selection of training datasets, since the slope failure can be a combination of several environmental factors. In the present study, the training samples were randomly selected, and some of parameters were determined by experience, for example, learning rates, crossover

probability and mutation probability. To some extents, the accuracy of prediction was affected by the accurate selections of these parameters.

Meanwhile, the advantages of using genetic algorithms were based on the performance of neural networks on the testing datasets, instead of only on the minimal square error in the modeling datasets. In addition, the minimal improvement of the genetic algorithms in this study occurred in the ratios between the numbers of chromosomes of a generation.

In the present study, through the use of direct-reverse DEM technology, the Changshougou valley was divided into 216 slope units, of which 123 units were affected by landslides. According to the mechanism analyses of the landslides in the loess area, six environmental factors were selected to evaluate landslide occurrence, such as slope height, slope angle, aspect, shape of slope, distance to rivers, and human activities. After the spatial analysis of the environmental factors, a case study was presented for landslide susceptibility prediction using BPNN modeling optimized by genetic algorithms. From a database of 216 slopes, 120 units, including 80 with landslide presence and 40 without, were used for training the neural network models, and 96 slopes, i.e., 43 with landslide presence and 53 without landslide presence, were used for the validation of landslide susceptibility. Comparing landslide presence with a susceptibility map, it was noted that the prediction accuracy for landslide occurrence is 93.02%, whereas units without landslide occurrence could be predicted with an accuracy of 81.13%. It was also noted that 10 slopes were predicted to be prone to landslides. In view of the environmental factors, most of these slopes are toward

442 southeast or southeast, are at an elevation greater than 100 m, and have slope angles
443 greater than 43°. It was also noted that seven of the slopes are in convex topographic
444 conditions, whereas the other slopes are located in concave topographic conditions.
445 Furthermore, the prediction of 10 slopes can be used as a general planning tool but is
446 not intended for individual site-specific evaluations.
447 As for the landsliding in loess areas, the six environmental factors were selected for
448 the susceptibility of landslides using the slope units. Geological conditions is not
449 complicated, since there are no faults and folds in the study area. All the selected
450 factors were statistically analyzed using GIS, however, the classification was
451 objectively determined. In fact, the large-scale landslides are composed of slope units,
452 the deformation is thus completely different from a slope unit. Further detailed studies
453 should be conducted for a single large-scale landslide in hazard assessment at a
454 site-specific scale. However, the rainfall is the main trigger for the slope failure in this
455 kind of areas, especially in northwest China. Detailed information for the landslide
456 occurrence is needed for the evaluation of hazard, therefore, landslide mitigation can
457 be effectively carried out in the Changshougou Valley.
458 To sum up, the verification indicates a satisfactory agreement, with an accuracy of
459 86.46% between the susceptibility map and landslide locations. In this case study, it
460 was also found that some disadvantages can be overcome in the application of BPNN,
461 such as low convergence rates and susceptibility to local minimums, after the
462 optimization was carried out using GAs. To conclude, GA based BPNN are an
463 effective method to predict landslide susceptibility with high accuracy.

Acknowledgements

This research was supported by funding from the Key Program of Natural Science Foundation of Hubei (2009CDA007). In addition, partial support from the Doctoral Fund from the Ministry of Education of China (20100142110059) is acknowledged. A special note of appreciation is extended to the Ministry of National Science and Technology for their funding support (2012BAK10B00). Two anonymous reviewers were greatly appreciated for their comments and suggestions.

References

Ayalew, L., Yamagishi, H.: The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, central Japan, Geomorphology, 65,15-31,2005.

Bai, S., Lu, G.., Wang, J., Zhou, P., Ding, L.: GIS-based rare events logistic regression for landslide-susceptibility mapping of Lianyungang, China, Environmental Earth Sciences, 62, 139-149, 2011.

Battiti, R.: Accelerated back propagation learning: Two optimization methods, Complex Systems, 3, 331–342, 1989.

Carrara, A., Cardinali , M., Detti,R., Guzzetti, F., Pasqui,V., Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, Earth Surface Processes Landforms, 16,427–445, 1991.

Catani, F., Casagli, N., Ermini, L. , Righini, G and Menduni, G.: Landslide hazard and risk mapping at catchment scale in the Arno River basin, Landslides, 2, 329-342, 2005.

Chen, H.Q., Zeng, Z.G.: Deformation Prediction of Landslide Based on Genetic-Simulated Annealing Algorithm and BP Neural Network. Fourth International Workshop on Advanced Computational Intelligence, Wuhan, Hubei, China; October 19-21, 2011.

Chung, C.F., Fabbri, A.G.: Validation of spatial prediction models for landslide hazard mapping, Natural Hazards, 30,451-472, 2003.

Chung, C.J. : Using likelihood ratio functions for modeling the conditional probability of occurrence of future landslides for risk assessment. Computer and Geosciences, 32, 1052-1068, 2006.

Dahal, R.K., Hasegawa, S., Nonomura, A., Yamanaka, M., Dhakal, S., Paudyal, P.: Predictive modelling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of-evidence, Geomorphology, 102, 496-510, 2008.

Derbyshire E, Meng, X.M., Dijkstra, T.A. (eds).: Landslides in the thick loess terrain of north-west China. John Wiley, 2000.

Ercanoglu, M. and Gokceoglu, C.: Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach, Environ Geol 41:720-730, 2002.

Ermini, L., Catani, F., Casagli, N.: Artificial Neural Networks applied to landslide susceptibility assessment, Geomorphology, 66, 327-343, 2005.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z.: Guidelines for landslide susceptibility, hazard and risk zoning for land use planning, Engineering Geology, 102, 85-98, 2008.

García-Rodríguez, M. J. and Malpica, J. A.: Assessment of earthquake-triggered landslide susceptibility in El Salvador based on an Artificial Neural Network model, Nat. Hazards Earth Syst. Sci., 10, 1307-1315, 2010.

Gómez, H., Kavzoglu, T.: Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela, Engineering Geology, 78, 11-27, 2005.

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

Hagan, M.T., Demuth, H.B., Beale, M.: Neural Network Design, PWS publishing company, a division of Thomson learning, Boston, USA, 1996.

Hasekiogullar, G. D., Ercanoglu, M.: A new approach to use AHP in landslide susceptibility mapping: a case study at Yenice (Karabuk, NW Turkey), Natural Hazards, 63, 1157 - 1179, 2012.

Haykin, S.: Neural Networks: A Comprehensive Foundation (2nd edition), Publisher: Prentice Hall, 1999.

Kanungoa, D.P., Arorab, M.K., Sarkara, S., Guptac, R.P.A.: Comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas, Engineering Geology, 85, 347–366, 2006.

Kesign, U.: Genetic algorithm and artificial neural network for engine optimisation of efficiency and NOx emission. Fuel, 83, 885–895, 2004.

530 Lee, S., Ryu, J.H., Min, K., Won, J.S.: Landslide susceptibility analysis using GIS and
531 artificial neural network, Earth Surface Processes and Landforms, 23,
532 1361-1376, 2003.
533 Lee, S., Pradhan, D.: Regional landslide susceptibility analysis using
534 back-propagation neural network model at Cameron Highland, Malaysia,
535 Landslides, 7(1), 13-30, 2010.
536 Madaeni, S.S., Hasankiadeh, N.T., Kurdian, A.R., Rahimpour, A.: Modeling and
537 optimization of membrane fabrication using artificial neural network and genetic
538 algorithm. Separation of Purification Technology, 76, 33–43, 2010.
539 Maidment, D.: Arc Hydro: GIS for water resources. ESRI 380, New York Street,
540 Redland, California, 2002.
541 Melchiorre, C., Matteucci, M., Azzoni, A., Zanchi, A.: Artificial neural networks and
542 cluster analysis in landslide susceptibility zonation, Geomorphology, 94,
543 379-400, 2008.
544 Meng, X.M., Dijkstra, T, D., Derbyshire, E.: Loess slope instability. In: Derbyshire E,
545 Meng, X.M., Dijkstra, T.A. (eds). Landslides in the thick loess terrain of
546 north-west China. John Wiley, 175-181, 2000.
547 Mu, Y., Yu, H.Q.: Simulation of biological hydrogen production in a UASB reactor
548 using neural network and genetic algorithm, Int. J. Hydrogen Energy, 32,
549 3308–3314, 2007.
550 Neaupane, K.M., Achet, S.H.: Use of back propagation neural network for landslide
551 monitoring: a case study in the higher Himalaya, Engineering Geology, 74,

213–226, 2004.

Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H.: 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, 171–191, 2008.

Nefeslioglua, H.A., Gokceoglub, C., Sonmez, H., Gorum, T.: Medium-scale hazard mapping for shallow landslide initiation: the Buyukkoy catchment area (Cayeli, Rize, Turkey). Landslides, 8(4), 459-483, 2011.

Peng, S.Z., Guo, Z.T.: Clay mineral composition of the Tertiary red clay and the Quaternary loess-palaeosols as well as its environmental implication. Chinese J. Quaternary Science, 27, 277-285, 2007.

Rumelhart, D.E., McClelland, J.L.: Parallel Distributed processing: Exploration in the Microstructure of Cognition. MIT-Press, Cambridge, MA, 1986.

Saengrung, A., Abtahi, A., Zilouchian, A.: Neural network model for a commercial PEM fuel cell system, J. Power Sources, 172, 749–759, 2007.

Sexton, R.S., Gupta, J.N.D.: Comparative evaluation of genetic algorithm and back propagation for training neural networks, Inform. Sciences, 129, 45–59, 2000.

Van Westen, C.J.: GIS in landslide hazard zonation: a review, with examples from Andes of Colombia, In: Price, M. and Heywood, I. (eds.), Mountain Environments and Geographic Information Systems, Taylor & Francis, Basingstoke, U.K., pp135-165, 1994.

Wang, H.B., Liu, G.J., Xu, W.Y., Wang, G.H.: GIS-based landslide hazard assessment: An overview. Progress in Physical Geography, 29, 548-567, 2005.

574 Wang, H.B., Zhou, B., Wu, S.R., Shi, J.S.: Characteristic analysis of large-scale loess
575 landslides: a case study in Baoji City of Loess Plateau of Northwest China.
576 Natural Hazards and Earth Sciences, 11, 1829–1837, 2011.
577 Xie, M.W., Esaki, T., Zhou, G.Y.: GIS-based probabilistic mapping of landslide
578 hazard using a three-dimensional deterministic model, Natural Hazards, 33,
579 265-282, 2004.
580 Yilmaz, Y.: An Agent Simulation Study on Conflict, Community Climate, and
581 Innovation in Open Source Communities, International Journal of Open Source
582 Software and Processes, 1(4), 1-25, 2009.
583 Zhou, X.: Genetic algorithm based on new evaluation function and mutation model
584 for training of BPNN, Tsinghua Science and Technology, 7, 28-31, 2002.
585