STATISTICAL AND HEURISTIC APPROACHES FOR SPATIAL PREDICTION OF LANDSLIDE HAZARDS IN LAOCAI, VIETNAM

QuocPhi Nguyen[1], Phuong Nguyen[1], KimLong Nguyen[2]

[1]Department of Environmental Sciences, Hanoi University of Mining and Geology, Vietnam
[2]Department of Petroleum Geology, Hanoi University of Mining and Geology, Hanoi, Vietnam

*Corresponding author: nguyenquocphi@humg.edu.vn

ABSTRACT

Steep terrain and high a frequency of tropical rainstorms make landslide a common phenomenon in Vietnam. This paper reports the use of spatial prediction models: Logistic Regression (LR), artificial neural network (ANN) and Analytical Hierarchy Process (AHP) approaches for landslide susceptibility mapping at Laocai city and its adjacent municipalities in Vietnam. A landslide database from study area was used to produce a susceptibility map along two main roads where small settlements and transportation routes are historically threatened by landslide phenomena. By means of statistical and heuristic methods, the tendency to landslide occurrences was assessed by relating a landslide inventory (dependent variable) to a series of causal factors (independent variables). The integration of spatial, temporal and size probabilities of landslide events helps convert the susceptibility maps into hazard maps of different scenarios.

1. INTRODUCTION

Landslide phenomena are one of the highest risk factors for people, environment and economic activities. In Laocai, landslide is widespread and recurrent phenomena due to its particular geological and geomorphological patterns. Historical investigations have revealed that, in the 10-year period (2000-2010), at least 36 single and multiple landslides have caused 78 deaths and injured people (Nguyen, 2011). The stability of natural or manmade slopes was governed by the interaction of several factors, such as: lithology, weathering profile, geological engineering and hydrogeological conditions, drainage network, slope angle, landcover/landuse, etc., and hence, there has been a growing interest in questioning relationship between landslide hazard and related variables.

To obtain a quantitative estimate for the spatial probability of landslide events in study area, multivariate analysis of landslide occurrence and environmental information (explanatory variables), including logistic regression, neural networks and Analytic Hierarchy Process methods were used. The spatial prediction models were tested to evaluate the degree of model fit and the model prediction capability. The degree of model fit was obtained by overall accuracy, Kappa index and ROC curve. Then, the model prediction capability was determined using independent landslide information not used to construct the models from database.

2. METHODOLOGY

2.1 Logistic regression model

Basically, Logistic regression analysis relates the probability of landslide occurrence (having values from 0 to 1) to the "logit" Z. The probability of landslide occurrence is expressed by:

$$P = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \quad (1)$$

The logit Z is assumed to contain the independent variables on which landslide occurrence may depend. The LR analysis assumes the term Z to be a combination of the independent set of landslide-related variables X_i (i =1,2,...,n) acting as potential causal factors of landslide phenomena. The term Z is expressed by the linear form:

$$Z = B_0 + B_1 X_1 + B_2 X_2 + ... + B_n X_n \quad (2)$$

where coefficients B_i (i =1,2,...,n) are representative of the contribution of single independent variables X_i to the logit Z and B_0 is the intercept of the regression function.

2.2 Neural networks

The Artificial Neural Network is made up of a large number of independent inter-connected units, which are neurons and synapses. Upon receiving sufficiently intense stimulus (input) from the preceding units, the unit is activated and sends signal to the connecting units.

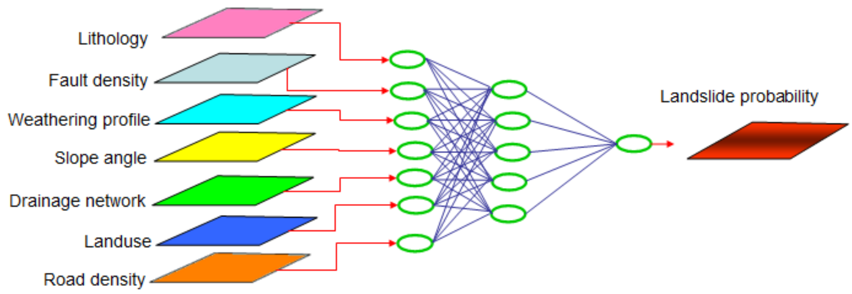


Figure 1. Landslide analysis in ANN

The transformation is completed in two phases (Lee et al., 2007): Firstly, each input signal is multiplied by the weight of the connection and the results of the single products are added to obtain an amount called total input. Secondly, the unit applies a transfer function which transforms the sum of the input signals into output signals. The behavior of an ANN depends on the architecture of the network and on both the weights assigned to the connections and the transfer function.

2.3 Analytical Hierarchy Process (AHP)

Generally, in an AHP model, the elements of each level with its respective element are compared as pair-wise and therefore the local priority is calculated. Then, with assimilating the local priorities, overall priority was calculated. As expressed, all comparisons were pair-wise and based on oral judgments that expressed by preference values (Ayalew and Yamagishi, 2005a, 2005b).

Then, the relative weight of each variable must be determined by the mathematical mean method. The pair-wise comparison of variables, the summation weights matrix, and the average weights matrix of each variable are alternately calculated (Saaty, 1980).

3. LANDSLIDE DATABASE

A database for landslide susceptibility in study area was developed based on:

- Field surveys in late July-early August and in October of 2011. The surveys were carried on with high accuracy GPS on topographic maps of 1:50,000 (partly 1:10,000). Historical data in study area were also collected from previous studies and reports.

- The geological maps of 1:200,000 (BacQuang and LaoCai-KimBinh sheets), and recent 1:50,000 scale maps cover the study area were used.

- Topographic maps of 1:50,000, produced in 2004.

- Satellite images of Landsat 7 (ETM+), which include scenes of L71128044 and L72128045, taken in 1999, 2001, 2004 and 2010 with cloud cover <10% were used.

The landslide-related parameters were then prepared as shown in table 1.

Components	**Variables**	**Data sources**	**Map scales**
Geological conditions	Lithology	Geological maps	1:200,000 1:50,000
	Geological structures	Geological maps, RS images	1:200,000 1:50,000
	Weathering profile	Geological maps	1:50,000
	Geological engineering conditions	Geological engineering maps	1:50,000
	Hydrogeological conditions	Hydrogeological maps	1:50,000
Natural conditions	Elevation	Topographic maps	1:50,000
	Slope angle	Topographic maps	1:50,000
	Slope exposure	Topographic maps	1:50,000
	Drainage networks	Topographic maps, RS images	1:50,000
	Land cover	Topographic maps, RS images	1:50,000
Human-induced conditions	Landuse	Topographic maps, RS images	1:50,000
	Road density	Topographic maps, RS images	1:50,000
	House density	Topographic maps, RS images	1:50,000

Table 1. Derived variables utilized for landslide analysis.

Identification of the landslide locations and delimitations was carried out by fieldwork, supported by analysis of the satellite images and historical data. The geo-database, collects information related to 82 landslide bodies along two main roads from the main city of Laocai (Laocai City) toward its adjacent county (Baothang).

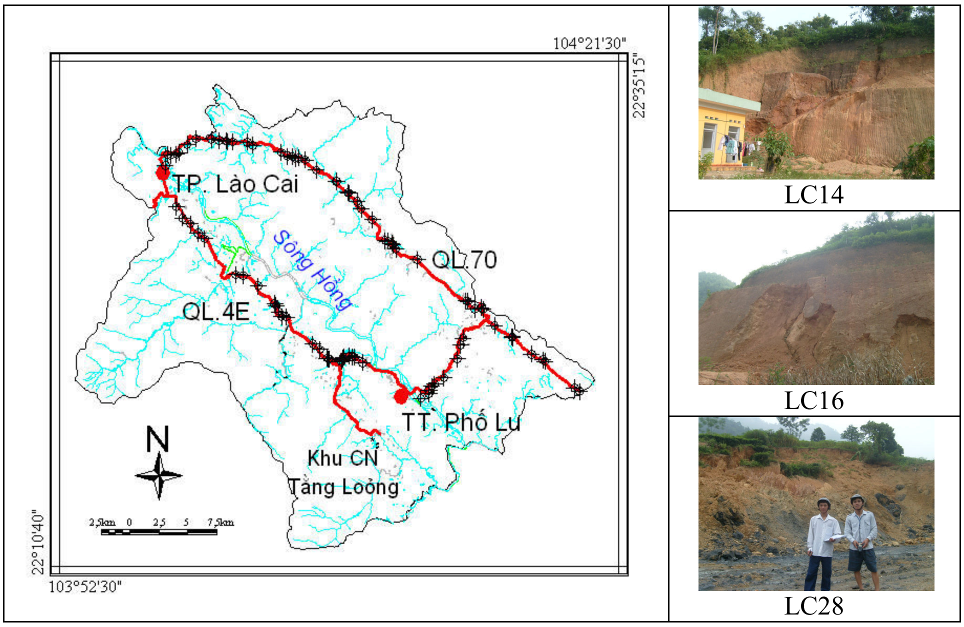


Figure 2. Landslide location map.

Factors are grouped based on geological, natural and human-induced conditions (table 1). Hence, in addition to making a landslides inventory in the investigated area, the 13 variables described in table 1 were used as causal factors. The study was carried out using a 30m grid size DTM.

4. RESULTS

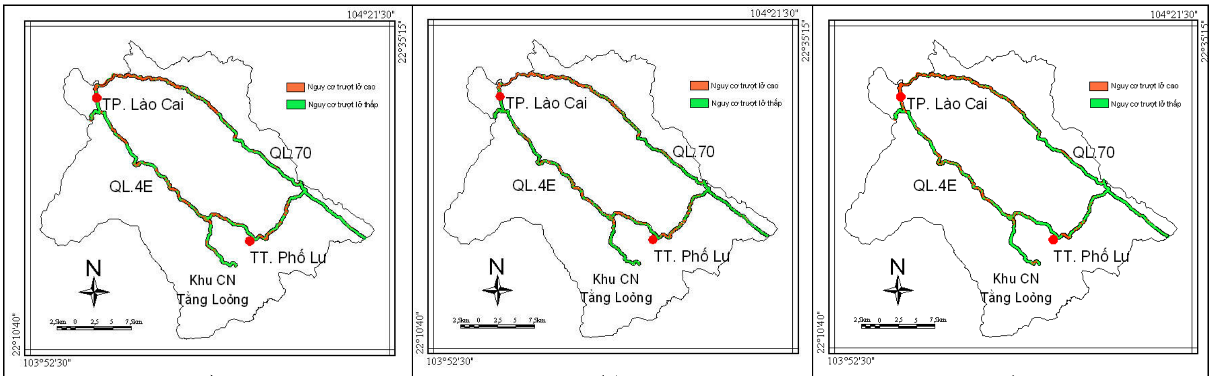
4.1 Landslide susceptibility models

For the logistic regression calculation, at each step, variables are evaluated for removal one by one if they do not contribute sufficiently to the regression equation. The variables included in the model were lithology, geological structures, geological engineering conditions, weathering profile, elevation, slope exposure (aspect), land cover (NDVI), landuse, drainage networks, and road and house densities as shown in equation below.

Hazard = 1.41 + [DEM] * 1.49 + [Aspect] * 0.26 + [NDVI] * 0.92 + [Landuse=Forest-Dense] * 0.5 + [Landuse=Forest-Spare] * -0.89 + [Landuse=Agriculture] * -0.79 + [Landuse=Barelands] * 0.27 + [Landuse=Urban] * -0.19 + [FaultDensity] * 0.21 + [StreamDensity] * -0.13 + [EngGeology=DaPhien] * -0.48 + [EngGeology=Cuoiket] * 0.32 + [EngGeology=TramTichTre] * 2.63 + [Geology=SinhQuyen] * 0.16 + [Geology=NgoiChi] * 0.6 + [Geology=NuiConVoi] * -0.51 + [Geology=HaGiang] * -0.51 + [HouseDensity] * 1.35 + [RoadDensity] * -0.31 + [WeatheringCrust=BCThachAnh] * 0.8 + [WeatheringCrust=TTichTre] * -0.43 + [WeatheringCrust=Carbonat] * 0.53

For the neural network, the work is divided in two steps: the training phase, in which one third (33%) of the total landslides were selected as training set and the weights were calculated, and the validation procedure, in which the obtained susceptibility map was cross-validated with the inventory database. The network was 13x7x1, with 13 input layers (13 landslide-related variables), 7 hidden layers and an output layer showing the presence/absence of landslides.

In AHP analysis, weights are given by scores collected from 7 experts, among them, 3 were working intensively on slope stability problem in study area. The relative importance of independent variables can be expressed by the weight, highlighting the causal factors that are most strongly related to the occurrence of landslides. Among the variables, geological structures, slope angle, weathering profile, geological engineering, hydrogeological conditions and drainage networks receive much attention from experts and can be considered as triggering factors.



a) b) c)

Figure 3. Landslide susceptibility models: a) Logistic regression; b) ANN and c) AHP

4.2 Validation of susceptibility maps

The overall performance of a landslide susceptibility model is generally judged on the accuracy (number of correctly classified cells), Kappa index of agreement and Area Under the Curve (AUC) of the ROC (Relative Operating Characteristic) curves. Validation results of developed models are shown in table 2 and figure 4.

Model	Kappa index	Accuracy	AUC
Logistic regression	0.7264	86.3%	0.904
Neural network	0.9789	98.7%	0.973
AHP	0.8305	91.3%	0.935

Table 2. Validation results for landslide susceptibility models.

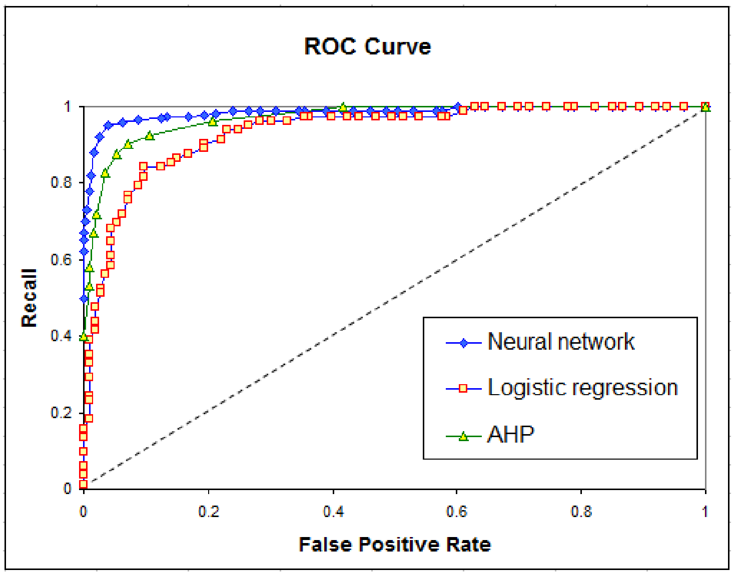


Figure 4. ROC curves of landslide susceptibility models.

Validation results show that all three developed models with Kappa index > 0.7, AUC > 0.9 and the overall accuracy > 85% could be considered very satisfactory for landslide susceptibility mapping, in which, ANN model gives the best result. Thus, neural network model can be selected as landslide susceptibility model for study area.

4.3 Landslide hazard analysis

The next step in the analysis was the conversion of the susceptibility maps into hazard maps. For this, three probabilities were calculated for pixels belonging to each class (Guzzetti et al., 2005; Glade et al., 2005; Fell et al., 2008):

(1) Magnitude probability, (P_M), defined as the probability that if a landslide occurs of a given size.

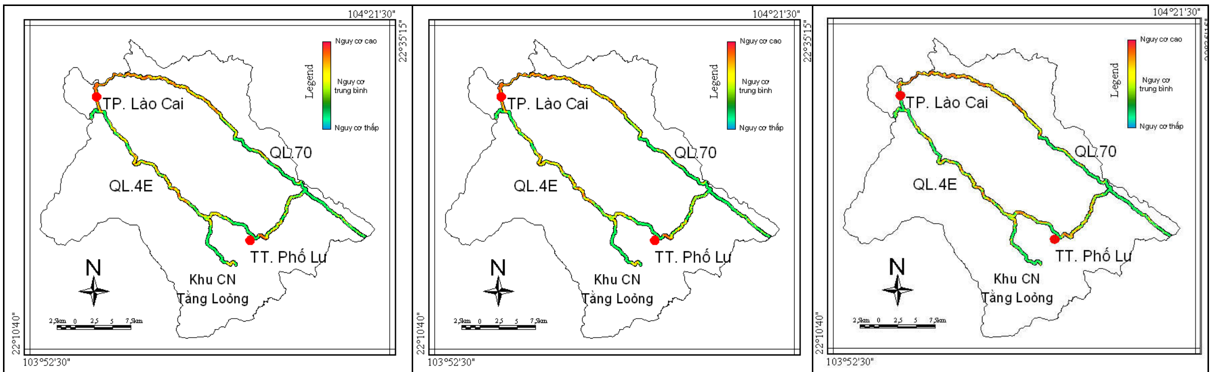
(2) Spatial probability, (P_S), defined as the probability that if a landslide occurs within a given susceptibility class.

(3) Temporal probability, (P_T), defined as the annual probability of occurrence of a particular landslide class.

Incorporation of the temporal probability into susceptibility maps is a difficult task, particularly when landslide information was not systematically collected. The probability values derived in this study is based on an inventory of a limited time period from 1989 to 2011 (Nguyen, 2011), and the data from 22-years period are used to estimate the frequency of landslides and the calculation of temporal probability. The joint probability for landslide hazard (H_L) is then calculated by:

$$H_L = P_M \times P_S \times P_T$$

The size of landslides is categorized follow Lomtadze's classification (Lomtadze, 1977) with 51 small, 26 medium and 5 large landslides. Temporal probability of return periods is calculated by using the Poisson distribution model (Crovelli, 2000) and number of landslides per unit length for different return periods is generated. Finally, landslide hazard maps are produced for different scenarios of 3, 5, 10 and 20 years.



a) b) c)

Figure 5. Examples of hazard maps of medium-sized landslides for 3 time periods: a) 3 years; b) 5 years and c) 10 years

5. DISCUSSION AND CONCLUSIONS

The susceptibility and hazard maps prepared in this study are a step forward in the management of landslide hazard in study area. The logistic regression, neural network and AHP models have demonstrated to be suitable tools to represent the relationships between landslides and causal factors. Among them, ANN model shows best result in term of successful predictive model for spatial pattern of landslides in study area. Such a result is achieved by the inspection of the coefficients (in LR) or weights (in ANN and AHP) that determine the role played by influencing factors on the landslide phenomenon.

This study aimed to quantify landslide hazard along road sections where landslides are small in size but occur frequently. Hazard assessment including small-sized landslides is important if these occur as even small slips can result in an accident. Landslide hazard maps are produced in terms of the number of landslides expected in a given location and its annual probability of occurrences, given the incompleteness of the historical landslide data. Although the limited availability of information of landslide occurrence, especially the time span of collected data and exact date of landslides, the calculation of hazard maps was still feasible. To bring the prediction into real life, the combination of landslide hazard maps with additional information, such as travel distance, vulnerability assessment… is currently conducting for subsequent risk analysis and estimates the probable losses due to future landslide events in study area.

6. REFERENCES

Ayalew, L. and Yamagishi, H., 2005a. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Engineering Geology* 81, 432–445.

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

Crovelli R. A., 2000. Probability models for estimation of number and costs of landslides. *USGS Open-file report* 00-249.

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

Glade T., Anderson M., Crozier M.J., (Editors) 2005. *Landslide Hazard and Risk*. John Wiley & Sons, London, pp. 802.

Guzzetti F., Reichenbach P., Cardinali M., Galli M., Ardizzonem F., 2005. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72, 272– 299.

Lee, S., Ryu, J. H. and Kim, I. S., 2007. Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression and artificial neural network models: case study of Youngin, Korea. *Landslides* 4, 327–338.

Lomtadze V. D., 1977. Engineering geodynamics. Nedra, Leningrad, 479pp.

Nguyen, Q.P., 2011. *Application of statistical methods for analyzing causal factors and landslide susceptibility mapping on major roads in Laocai province*. Scientific report, Hanoi University of Mining and Geology, Vietnam.

Saaty, T.L., 1980. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. NY, McGraw Hill.