Landslide susceptibility zonation - A comparison of natural and cut slopes using statistical models

QuocPhi Nguyen[a*], TruongThanh Phi[b], AnhNguyet Nguyen[b], VanAnh Lam[c]

[a] *Hanoi University of Mining and Geology, Vietnam*
[b] *Institute of Marine Geology and Geophysics, Vietnam*
[c] *Ban Ve Hydropower Joint Stock Company, Vietnam*
** corresponding author: nguyenquocphi@humg.edu.vn*

Abstract

The study aims to apply statistical approaches for landslide susceptibility mapping for natural and cut slopes at two study areas in Vietnam. A landslide database from two study areas with 837 natural and 82 cut slopes was used to produce susceptibility maps to predict the landslide hazard in the future. The distribution of landslides was identified from field surveys, research reports and remote sensing images. By means of Likelihood Ratio (*LR*), Weight of Evidence (*WoE*) and Certainty Factor (*CF*) approaches, the tendency to landslide occurrences was assessed by relating landslide inventory (dependent variable) to a series of causal factors (independent variables) which were managed in the GIS environment. The developed models produced reliable susceptibility maps of study areas and the probability level of landslide can be divided by four different classes (low, medium, high and very high). The overall performance achieved by the *LR*, *WoE* and *CF* analyses was assessed on validation datasets in both two study areas with Kappa statistic (*KIA*) > 0.7, area under curve (*AUC*) > 0.85 could be considered very satisfactory for landslide susceptibility zonation. All three models give over 80% of accuracy, in which *WoE* give best results. Landslide zonation map shows high and very high classes account for only 25% of total area of Quang Ngai province, but they can explain for nearly 90% of existing landslide locations. The weights of statistical approaches can also provide the important level of causal factors, relatively. In which, 3 most affluent factors for natural slopes are geological engineering conditions, landuse and the rock type (lithology) of the slopes, for man-made slopes are the angle of cut slope, weathering depth and the strength of slope materials.

Keywords: landslide hazard, predictive model, spatial analysis, cut slope.

1. Introduction

Landslides are more widespread than any other geological event and many factors can cause a slope to fail, such as natural occurrences or man-made activities. Landslide hazard mapping was defined as the quantitative prediction of the spatial distribution of slopes which are likely to be failure (Guzzetti et al., 1999). The causal factors that have been used for landslide hazard analysis can usually be grouped into geomorphology (topographic conditions), geology (rock types, structures, strength of slope materials), land use/land cover and hydrogeology (drainage state and ground water). However, the contributing factors are behaved differently for natural and cut slopes. A natural slope is different from a cut slope (road cuts, excavations, open-pit mining, etc.) in that the effects of rock types, fracture networks inside the rocks, the strength and weathering property of slope materials, the contribution of water on the surface or in the ground may undergone the test of time that will reveal tendencies of slumping, cracking and finally collapsing. On cut slopes, the slope angle, weathering depth, slope cover type, the property of upslope terrain, slope reinforcement, etc. may play an important role in their stability. A cut slope may expose soils that respond poorly to weathering elements, especially when the soil profile of slopes is not uniform and homogenous.

Vietnam is a mountainous country (mountains and hills account for 70% of the territory), and landslide is widespread and recurrent phenomena due to its particular geological, geomorphological patterns and especially, the unfavorable weather conditions associated with tropical climate and rainfalls. The stability of natural or manmade slopes was governed by the interaction of several causal factors. Due to this, there has been a growing interest in questioning relationship between landslide hazard and related variables. The focus is, therefore, on the recognition of landslide prone areas, which can be achieved by mapping susceptibility. The analysis is used to identify the factors that are related to landslides, estimate the relative contribution of factors causing slope failures, establish a relation between the factors and landslides, and to predict the landslide hazard in the future based on such a relationship. However, the causal factors for landsliding of natural and cut slopes are relatively different although under similar geo-environmental conditions. Statistical approaches were undertaken to assess these differences and evaluate their applicability to modeling susceptibility in the study areas.

2. Methodology

2.1 Likelihood Ratio (LR)

It is common to assume that land slide occurrence is determined by landslide-related factors, and that future landslides will occur under the same conditions as past landslides. Using this assumption, the relationship between landslides occurring in an area and the landslide-related factors can be distinguished from the relationship between landslides not occurring in an area and the landslide-related factors. Likelihood ratio method can be expressed as a frequency ratio that represents the quantitative relationship between landslide occurrences and different causative parameters. The likelihood ratio is defined as follow:

$$w_{ij} = \frac{f_{ij}^*}{\bar{f}_{ij}^*} = \frac{A_{ij}^*}{A^*} \times \frac{A - A^*}{A_{ij} - A_{ij}^*} \tag{1}$$

Where: w_{ij} - likelihood ratio of class i of parameter j;

f_{ij}^* - frequency of observed landslides in class i of parameter j;

$\bar{f}_{ij}^*$ - frequency of non-observed landslides in class i of parameter j.

Therefore, the greater the ratio above unity, the stronger the relationship between landslide occurrence and the given factor's attribute, and the lower the ratio below unity, the lesser the relationship between landslide occurrence and the given factor's attribute.

2.2 Weight of Evidence (WoE)

The weight of evidence (*WoE*) modeling method is a quantitative "data-driven" method used to combine datasets. The method was first applied in medicine (Spiegelhater and Kill-Jones, 1984). Afterwards, Bonham-Carter et al. (1990) used this approach to identify the gold mineralization in Nova Scotia, Canada. The applications in geology using a *WoE* model to estimate the relative importance of evidence by statistical means was developed by Bonham-Carter (1994).

$$P\{E \mid B\} = \frac{P\{E\} \times P\{B \mid E\}}{P\{B\}} \tag{2}$$

Prior probabilities and posterior probabilities are the most important concepts in the Bayesian approach. The probability P is usually determined with knowledge about the occurrence of an event E in the past under equal conditions, and as the prior probability $P\{E\}$. When these evidences are integrated into the calculation of the probability, a conditional or posterior probability $P\{E|B\}$ is obtained. Bayes theorem gives the relationship between prior and posterior probabilities:

$$P_0 = f = \frac{A^*}{A} \tag{3}$$

Evidence layers $j=1,2,\ldots n$ are added one after the other yielding posterior probabilities P_j are given by:

$$P_j = \frac{f_{ij}}{f} P_{j-1} = \left(\frac{A_{ij}^*}{A_{ij}} \times \frac{A}{A^*} \right) P_{j-1} \tag{4}$$

By overlaying landslide locations with each evidence (causative factor), the statistical relationship can be measured and assessed as to whether and how significant the evidence is responsible for the occurrence of past landslides. Hence, starting from a prior probability P_0, which is the density of the observed landslides in the study area.

In practice, in *WoE* positive and negative weights (W^+ and W^-) are calculated, the magnitude of which depends on the measured association between the response variable (the landslides) and the predictor variables (causative factors):

$$W_{ij}^{+} = \ln \frac{P\{B \mid E\}}{P\{B \mid \bar{E}\}} = \ln\left(\frac{f_{ij}^{*}}{\bar{f}_{ij}^{*}}\right) \tag{5}$$

$$W_{ij}^{-} = \ln \frac{P\{\bar{B} \mid E\}}{P\{\bar{B} \mid \bar{E}\}} = \ln\left(\frac{1-f_{ij}^{*}}{1-\bar{f}_{ij}^{*}}\right) \tag{6}$$

In above expressions, the bar above the symbols means the opposite, i.e. *E* means not *E*.

Finally, the contrast C_{ij}, measures and reflects the spatial association between the evidence feature and the landslide occurrence, and is given by:

$$C_{ij} = W_{ij}^{+} - W_{ij}^{-} \tag{7}$$

The contrast is positive for a positive spatial association, and negative for a negative spatial association. Hence, the contrast is the rating of each class of each factor that influences landslide occurrence: $w_{ij} = C_{ij}$.

In this study, all w_{ij} layers using *LR* and *WoE* methods for the different causative factors were constructed in GIS software. Then, these were summed up by equation 8 to obtain a resultant landslide susceptibility index map:

$$LSI = \sum_{i=1}^{n} w_{ij} \tag{8}$$

Where: *LSI* - Landslide Susceptibility Index;

w_{ij} - weight of class *i* in parameter *j*;

n - number of parameters.

2.3 Certainty Factor (CF)

The basic principles of the certainty factor (CF) approach were first introduced in MYCIN, an expert system for the diagnosis and therapy of blood infections and meningitis (Shortliffe and Buchanan, 1975). The CF theory for landslide hazard was used by Chung and Fabbri (1993, 1998), Binaghi et al. (1998), and Lan et al. (2004). The certainty factors are calculated as follows:

$$CF_{ij} = \begin{cases} \dfrac{f_{ij} - f}{f_{ij}(1-f)}; & \text{if } f_{ij} \geq f \\[2ex] \dfrac{f_{ij} - f}{f(1-f_{ij})}; & \text{if } f_{ij} < f \end{cases} \tag{9}$$

Where: CF_{ij} - Certainty factor given to class *i* of parameter *j*.

f_{ij} - the landslide density within the class *i* of parameter *j*.

f - the landslide density of study area.

Statistically, f_{ij} is the conditional probability having a number of landslide events occurring in class *i* and *f* is the prior probability of total number of landslide event occurring in the study area. The certainty factor is a number to measure the certainty cause for an observed outcome. The range of the certainty factor is *[-1,+1]*. The minimum *-1* means definitely false and *+1* means definitely true. Positive values mean an increasing certainty in causality, while negative values correspond to the opposite, i.e. the presence of the factor tends to disfavor the occurrence. A value close to *0* means that it is difficult to give any indication about the causality.

Next, the *CF* values of the causative factor are pairwise combined using the *CF* combination rule.

A combination of two *CF* values, CF_x and CF_y from two different layers of information, is a CF_z value obtained as follows (Chung and Fabbri, 1993; Binaghi et al., 1998):

$$z = \begin{cases} x + y - xy & if \quad x, y \geq 0 \\ \dfrac{x + y}{1 - \min(|x|, |y|)} & if \quad xy < 0 \\ x + y + xy & if \quad x, y < 0 \end{cases} \tag{10}$$

3. Landslide susceptibility zonation

3.1 Landslide inventory

Two datasets of landslide on natural slopes (Quang Ngai province) and man-made slopes (Lao Cai province) have been prepared. Quang Ngai province is located in the central part of Vietnam and landslide phenomena are one of the highest risk factors for people, environment and economic activities. The rocks of study area are composed from Proterozoic metamorphic rocks to Neogene basaltic lavas and Quaternary unconsolidated materials. Some deltaic and coastal slopes are composed of recently formed semi-consolidated and unconsolidated gravel and sand deposits. Landslide inventory of Quang Ngai province composed information of location, classification, volume, activity and if possible, date of occurrence with 837 natural landslides, which are further divided into 18 falls/topples, 172 rotational slides, 202 flows and 445 complexes in term of landslide type (Nguyen, 2013).

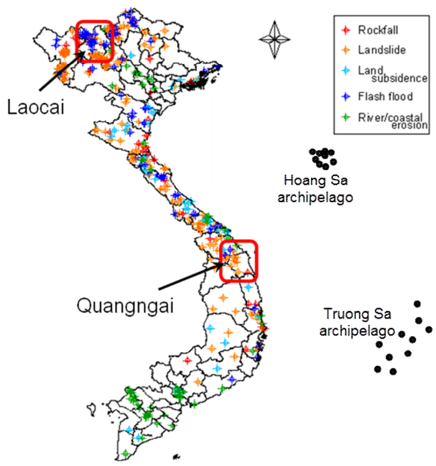
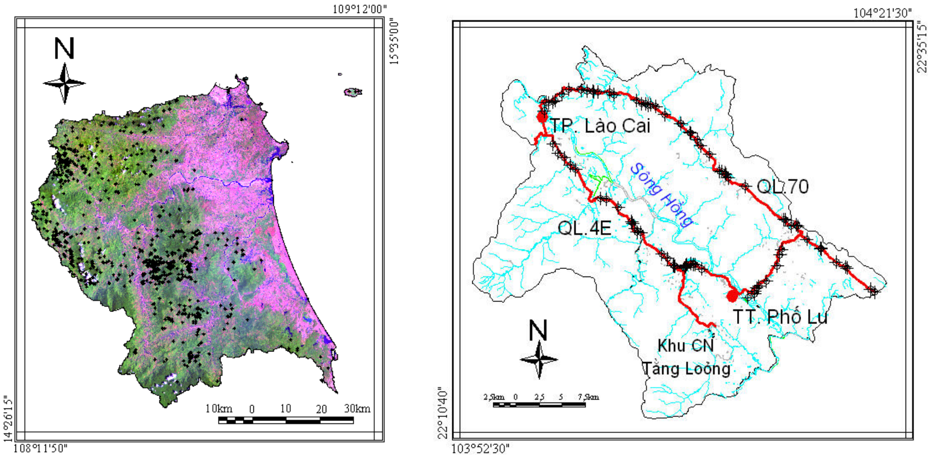


Figure 2. Location of study areas

Lao Cai is a mountainous province located on the highest mountain range in the northern part of Vietnam and heavily struck by storms and flash floods yearly. Historical investigations in Lao Cai 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 area from the main city of Lao Cai (Lao Cai city) toward its adjacent county (Baothang) forms a hub of highest economic activities in Lao Cai with infrastructure developments, industrial zones, construction sites. Landslide data of 82 cut slopes were collected along two major roads connect Lao Cai City and Baothang County. Most of the cut slopes are left unsupported and become prone to failure during rains.



(*a*) (*b*)

Figure 3. Landslide location in Quang Ngai (*a*) and Lao Cai (*b*) datasets

3.2 Landslide causal factors

Landslide susceptibility mapping relies on a rather complex knowledge of slope movements and their controlling factors (Dai et al., 2002; Pradhan and Lee, 2010). Collected data are then divided into different components based on geological, natural and human-induced conditions and the 13 causal factors related to landslides, namely lithology, geological structures, weathering profile, geological engineering conditions. hydrogeological conditions, elevation, natural slope angle, slope exposure, drainage networks, land cover, landuse, road density and population density were used for natural landslides in Quang Ngai dataset. For man-made landslides in Lao Cai dataset, different causal factors have been analysed, which are the rock and soil type, slope material strength, weathering depth, the presence or absence of groundwater, the angle of cut slope, the property of upslope terrain, slope cover type, slope reinforcement and the distance from the road.

4. Results

4.1 Landslide susceptibility models

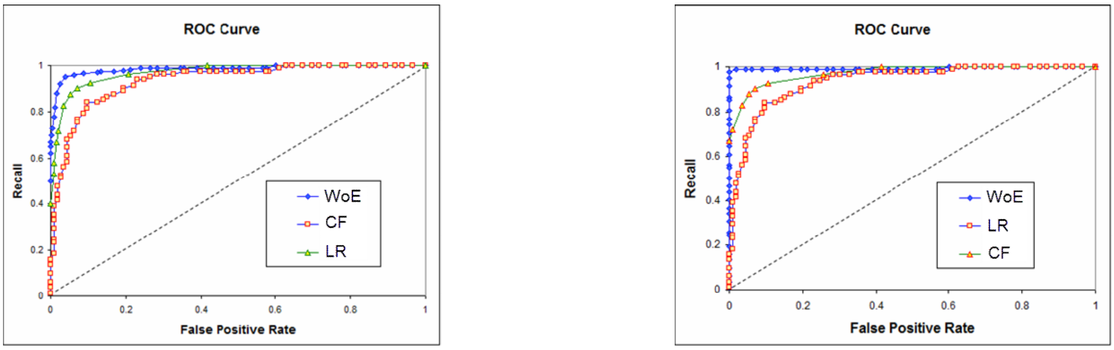
Landslide susceptibility models were constructed by Likelihood Ratio (*LR*), Weight of Evidence (*WoE*) and Certainty Factor (*CF*) approaches based on the physical parameters as defined above. To make the results easier to interpret, the landslide susceptibility maps were divided into four classes based on standard deviations of the probability histograms: low, medium, high and very high.

Nearly 90% of the identified landslides actually fall within high and very high category for landslide dataset on natural slopes and similarly, nearly 85% for man-made slopes. The results clearly show existing ground conditions in these areas are very likely to create serious landslide problems in the future.

4.2 Validation

The Kappa Index of Agreement (*KIA*) and the accuracy (defined as the percentage of landslide pixels classified correctly) were used to evaluate the performance of the training and the testing phases. The *KIA* was used to compare the specific class differences between classifications. This index measures the association between two images on a category-by-category basis. The values produced vary from -1 to 1 for each category (group) in the image. If the input images are in perfect agreement then $KIA = 1$; if there is no agreement then $KIA = -1$ and if the difference is produced by chance then $KIA = 0$.

In addition, the Relative Operating Characteristic (*ROC*) was used as an alternative approach to the assessment classification of the predictive rule. In the *ROC* analysis, the susceptibility map is compared with a dataset reporting the presence/absence of occurrences in the same area. Values close to 1 indicate a very good fit (perfect classification) whereas a random fit of the model produces values of the Area Under the Curve (*AUC*) close to 0.5 in the *ROC* space. Validation results of developed models are shown in table 1.



(a) *(b)*

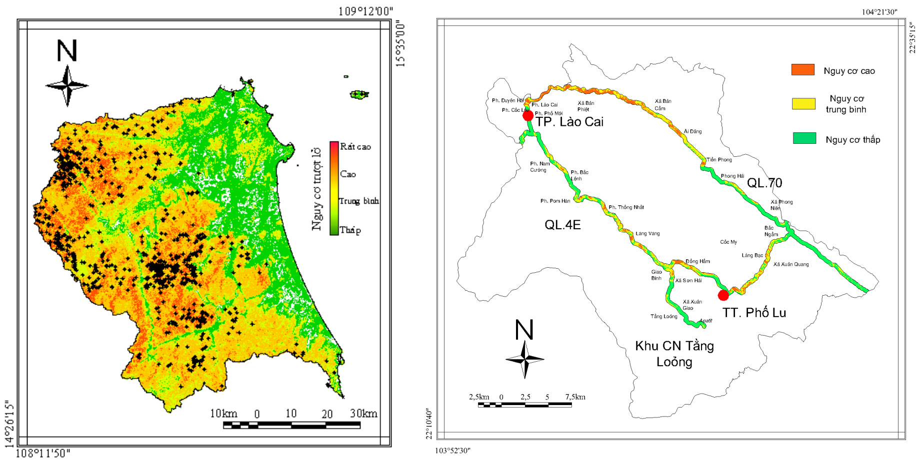
Figure 4. Curves of success rate from Quang Ngai (*a*) and Lao Cai (*b*) datasets

Validation results show that all three developed models with *KIA > 0.7*, *AUC > 0.85* and overall accuracy > 80% could be considered very satisfactory for landslide susceptibility mapping. Among 3 analysed models, *WoE* model gives the best results for both landslide on natural slopes (Quang Ngai dataset) and man-made slopes (Lao Cai dataset), *LR* and *CF* models come next and interchange their position for the natural and man-made slopes. Finally, *WoE* model was selected as landslide susceptibility model for both 2 study areas.

Table 1. Validation results for landslide susceptibility models.

Model	Quang Ngai dataset			Lao Cai dataset		
	KIA	*Accuracy*	*AUC*	*KIA*	*Accuracy*	*AUC*
Likelihood Ratio (*LR*)	0.723	83.2%	0.882	0.831	91.3%	0.953
Weight of Evidence (*WoE*)	0.952	97.4%	0.957	0.978	98.7%	0.973
Certainty Factor (*CF*)	0.867	92.8%	0.935	0.780	86.3%	0.891

The weights from *WoE* model can also provide the important level of causal factor, relatively. In which, 3 most affluent factors for natural slopes are geological engineering conditions, landuse and the rock type (lithology) of the slopes, for man-made slopes are the angle of cut slope, weathering depth and the strength of slope materials.



(a) *(b)*

Figure 6. Final susceptibility map prepared by *WoE* approach for Quang Ngai (*a*) and Lao Cai (*b*) datasets

In final susceptibility table, high and very high classes reveal the geographical distribution of the areas most prone to landslide occurrences. The susceptibility map of Quang Ngai province shows the high and very high classes accumulate for only 25% of study area, but contribute for almost 90% (738/837) of all landslides (table 2).

Table 2. The distribution of landslide susceptibility classes in Quang Ngai province.

Susceptibility class	**Area (km^2)**	**Percentage (%)**	**Landslides**	**Percentage (%)**
Low	1545,02	29,98	8	0,96
Medium	2320,21	45,03	91	10,87
High	1047,49	20,33	350	41,82
Very High	240,28	4,66	388	46,36
Total	***5153***	***100***	***837***	***100***

5. Discussion and conclusions

The landslide susceptibility maps prepared in this study is a step forward in the management of landslide hazard for both natural slopes and cut slopes in 2 study areas. The *LR*, *WoE* and *CF* models have demonstrated to be suitable tools to represent the relationships between landslides and causal factors, in which, *WoE* model shows best results in term of successful predictive model. As the main outcome of this work, a landslide susceptibility map was finally produced and validated.

The use of statistical approaches in this study has demonstrated that they can be powerful tools for landslide evaluation with high accuracy in term of spatial predictive models. The landslide susceptibility maps obtained from the study can provide very useful information for decision making and policy planning in landslide areas.

References

Akgun A., Dag S. and Bulut F., 2008. Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environmental Geology* 54, 1127-1143.

Ayalew L., Yamagishia H., Maruib H. and Kannoc T., 2005. 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.

Bonham-Carter G.F. and Reddy R.K.T., 1990. Preliminary results using a forward-chaining inference net with a GIS to map base-metal potential: Application to Snow Lake Greenstone Belt, Manitoba, Canada. *Proceedings of International Workshop on Statistical Prediction of Mineral Resources*, Wuhan, China.

Bonham-Carter G.F., 1994. *Geographic information systems for geoscientists: Modelling with GIS*. Pergamon, 398pp.

Dai F. C., Lee C. F. and Ngai Y. Y.,2002. Landslide risk assessment and management: An overview. *Engineering Geology* 64, 65-87.

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

Lee S. and Sambath T., 2006. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology* 50, 847-855.

Nguyen Q. P., 2011. *Application of statistical methods for analyzing causal factors and landslide susceptibility mapping on major roads in Lao Cai province*. Unpublished report, Hanoi University of Mining and Geology, Vietnam, 55 pp.

Nguyen Q. P., Nguyen Q. L., Nguyen K. L., Nguyen P., Nguyen T. T. M., 2013. Geological hazard zonation using Bayesian statistics: A case study for natural landslides in Quang Ngai province. *Proceedings of National GIS Conference*, Hanoi, 310-319.

Pradhan B. and Lee S., 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling. *Environmental Modelling Software* 25, 747-759.

Shortliffe E.H. and Buchanan B.G., 1975. A mathematical model inexact reasoning in medicine. *Mathematical Bioscience* 23, 351-379.

Spiegelhater D. and Kill-Jones R.P., 1984. Statistical and knowledge approaches to clinical decision-support systems, with an application in gastroenterology. *Journal of the Royal Statistical Society* 147, 35-77.