GIS-BASED LANDSLIDE SUSCEPTIBILITY ASSESSMENT ALONG THE NATIONAL ROAD 32 (VIETNAM) USING LOGISTIC REGRESSION AND INDEX OF ENTROPY MODELS

Dieu Tien Bui[1, 2]; Chung Tien Ho[3]; Inge Revhaug[1]

[1] Department of Mathematical Sciences and Technology
Norwegian University of Life Sciences, P.O. Box 5003IMT, N-1432, Aas, Norway
E-mail: Bui-Tien.Dieu@umb.no; Inge.Revhaug@umb.no
[2] Faculty of Surveying and Mapping, Hanoi University of Mining and Geology
Dong Ngac, Tu Liem, Hanoi, Vietnam
[3] Department of Tectonic and Geomorphology, Vietnam Institute of Geosciences and Mineral Resources
Thanh Xuan, Hanoi, Vietnam
E-mail: hotienchung@gmail.com

ABSTRACT

The main objective of this study is to evaluate and compare the results of applying the logistic regression and the index of entropy methods for landslide susceptibility assessment along the National Road 32 of Vietnam. First, a landslide inventory map with the total of 262 landslide locations that accounts for landslides that occurred during the last 20 years was constructed using data from various sources. Then, the landslide inventory was randomly partitioned into a training dataset 70% for building the models and the remaining 30% was used for models validation. Second, the factors that influence landslide occurrence are such as slope angle, slope aspect, relief amplitude, topographic wetness index (TWI), topographical shape, distance to roads, distance to rivers, distance to faults, and rainfall. Using these factors, landslide susceptibility indexes were calculated using logistic regression and index of entropy models. Finally, two landslide susceptibility maps were validated and compared using the validation dataset that was not used in the model building. Using the prediction-rate method, the prediction-rate curves and area under the curves (AUC) were calculated. The validation results showed that the AUC for the logistic regression model was 0.906 and for the index of entropy model was 0.890. It indicates that the prediction capability of the logistic regression is slightly better than those obtained by the index of entropy model.

1. INTRODUCTION

Landslides are considered among the most significant natural hazards that cause different types of damage affecting people, organizations, industries, and the environment (Glade, 1998). Global warming and its anticipated consequences for natural hazards are likely to incur increasing losses of lives, infrastructure, as well as soil and water resources (Korup *et al*., 2012). For the case of Vietnam, the country has been identified as a heavily vulnerable place to the worst manifestations of climate change such as sea level rise, flooding, and landslides (Tien Bui *et al*., 2012a), between 1980 and 2010, about 16,000 people were killed and almost 74 million people were affected as a result of natural disasters including landslides (UN-ISDR, 2012). Landslide disasters can be reduced by understanding the mechanism and developing appropriate tools for landslides prediction, assessment, risk management (Sassa and Canuti, 2008). However, only a few attempts have been carried out for landslide analysis in Vietnam such as (Lee and Dan, 2005; Tien Bui *et al*., 2011a; 2012a; 2011b; 2012b; 2012c). Therefore, study of landslides is an urgent task in Vietnam. The main objective of this study is to evaluate and compare the results of applying the logistic regression and the index of entropy methods for landslide susceptibility assessment along the National Road 32 of Vietnam.

2. STUDY AREA AND DATA

The National Road 32 is situated in the North West region of Vietnam. The total length is 250 km, from Deo Khe (Yen Bai province) to Binh Lu (Lai Chau province). Location and extent of the study area is between longitude $103^0 33'23''$E and $104^0 52'58''$E, and between the latitude $22^0 20'18''$N and $21^0 19'53''$N. The study area covers an area of about 3,164 km^2 (Figure 1). The elevation ranges between 120 and 3,140 m with average value of 1,078 m and standard deviation of 555.9 m.

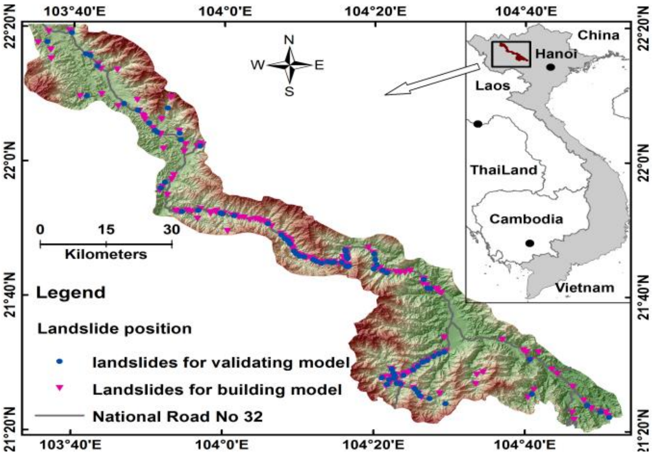

Figure 1. **Landslide inventory of the study area.**

Assuming that landslides will occur in the future based on the same conditions that influence them in the past (Guzzetti *et al.*, 2006), therefore the landslide inventory map was constructed first using various sources. The landslide inventory was used to derive the quantitative relationships between the landslide occurrence and landslide conditioning factors.

A total of 262 landslides that occurred during the last 20 years were registered in the landslide inventory map. The size of the smallest landslide is about 476 m2. The largest landslide covers an area of 37,326 m2. For building the landslide models, the landslide inventory map was converted into a raster format with a spatial resolution of 20x20 m. The landslide inventory map was then randomly split into a training dataset of 70% (182 landslides with 2,781 pixels) for building the landslide models and a validation dataset 30% (80 landslides with 1,011 pixels) for model validation (Figure1).

A total of ten landslide conditioning factors were selected and used for this study such as: slope angle, slope aspect, relief amplitude, topographic wetness index (TWI), and topographical shape, which were derived from a digital elevation model (DEM) with a spatial resolution of 20 x 20 m. The DEM was generated from national topographic maps at a scale of 1:50,000. Distance to roads and distance to rivers were computed based on river and road networks from the national topographic maps. Lithology and distance to faults were extracted and calculated from Geological and Mineral Resources maps at the scale of 200,000. In addition, a rainfall map was included in the analysis. The detail classes for the ten landslide conditioning factors are shown in Table 1.

Table 1. Landslide conditioning factors and their classes for this study.

Data layers	**Class**	**Class pixel**	**Landslide pixel**	**Weight (IoE)**	**Coefficients (LR)**	**P-*value* (LR)**
Slope (degree)	0-8	1050626	0	0.000	0.061	0.000
	8-15	707774	152	0.221		
	15-25	1949706	1071	0.664		
	25-35	2352056	936	0.553		
	35-45	1379361	508	0.443		
	> 45	431672	114	0.332		
Aspect	Flat	370810	0	0.000	0.054	0.000
	North	880893	176	0.247	-0.724	
	Northeast	954851	174	0.165	-1.222	
	East	887194	385	0.576	-0.337	
	Southeast	943832	460	0.659	-0.408	
	South	1016869	612	0.741	-0.285	
	Southwest	1061249	448	0.494	-0.178	
	West	893222	193	0.329	-1.312	
	Northwest	862275	333	0.412	0.054	
Relief amplitude	0–50	494988	85	0.453	22.786	0.000
	50–200	3797449	1994	0.754	21.688	
	200-350	3106823	665	0.603	20.747	
	350-500	449532	37	0.302	19.686	
	>500	27992	0	0.000	22.786	
TWI	<5	744	0	0.000	34.515	0.000
	5-10	6246498	2599	0.920	17.784	
	10-15	1328468	177	0.736	17.226	
	15-20	233435	5	0.552	15.210	
	>20	20661	0	0.000	34.515	
Toposhape	Ridge	1437448	544	1.147	18.827	0.031
	Saddle	113672	0	0.000	-1.255	
	Flat	374668	0	0.000	-1.128	
	Ravine	1399148	429	0.717	18.775	
	Convex hillside	1030755	352	0.861	18.796	
	Saddle hillside	2408283	1019	1.291	19.096	
	Slope hillside	16366	0	0.000	-0.244	
	Concave hillside	945577	415	1.434	19.205	
	Inflection hillside	60540	22	1.004	19.480	
	Unknown hillside	90484	0	0.000	18.827	
Lithology	Aluvium	239956	73	1.361	-0.584	0.000
	Conglomerate	789689	138	0.756	-0.900	
	Dyke	27674	19	1.512	0.159	
	Intermediate	42216	0	0.000	-20.256	
	K-Pluton	87073	0	0.000	-20.975	
	K-Volcanic	4030918	1218	1.209	-0.452	
	Limestone	240162	39	0.605	-0.412	
	P-Volcanic	5770	0	0.000	-20.496	
	Sandstone	237588	184	1.814	1.371	
	Schist	274344	78	0.907	-0.412	

	Shale	679581	204	1.058	-0.627	
	Tuff	1204892	828	1.663	-0.584	
Distance to faults (m)	0–200	2419662	1166	0.119	0.432	0.000
	200–400	2027888	703	0.060	0.547	
	400–600	1442500	508	0.089	0.893	
	>600	1986891	404	0.030	0.432	
Distance to roads (m)	0–40	273124	378	1.784	3.486	0.000
	40–80	292995	634	3.568	2.908	
	80–120	288433	872	2.676	2.397	
	>120	7022389	897	0.892	3.486	
Distance to rivers (m)	0–40	541068	375	0.233	0.939	0.000
	40–80	581557	386	0.155	0.960	
	80–120	576604	432	0.311	0.947	
	>120	6177712	1588	0.078	0.939	
Rainfall (mm)	< 1500	892649	145	0.118	-0.361	0.000
	1500-1700	1397443	589	0.235	-0.202	
	1700-1900	2272315	1183	0.294	0.166	
	1900-2200	2060447	669	0.176	0.391	
	>2200	1254071	195	0.059	-0.361	

3. LANDSLIDE SUSCEPTIBILITY MAPPING USING LOGISTIC REGRESSION AND INDEX OF ENTROPY

3.1 Logistic regression (LR)

Logistic regression is a mathematical modeling approach that can be used for predicting the presence or absence of outcome based on the values of a set of predictor variables. The mathematical form of the logistic regression is as follows:

$$f(z)=\frac{1}{1+e^{-z}};\ z=\alpha+\beta_1x_1+\beta_2x_2+...+\beta_nx_n \quad (1)$$

where *f*(z) is the logistic function; Z is a linear sum of a constant and products of independent variables and their respective coefficients; α is the constant; $\beta_i\,(i=0,1,2,...,n)$ are the coefficients; $x_i\,(i=1,2,...,n)$ are the independent variables.

For building landslide model, 2781 landslide pixels (in the training dataset) and 2,781 points (from the non-landslide area) were used, and values of ten independent variables were extracted to build a training database. The result of the LR is shown in Table 1. The *P-value* of each of the ten landslide conditioning factors in the logistic regression is less than 0.05. This indicates statistical relationship between variables at the 95% confidence level (Tien Bui *et al*., 2011a). Using the logistic regression coefficients, landslide susceptibility index was calculated and the landslide susceptibility map was obtained (Figure 2).

3.2 Index of Entropy (IoE)

According to Bednarik *et al*.,(2010), entropy is an index to the degree of disorderliness between decisions and attributes. It has been used to assess the correlation between classes of landslide conditioning factor parameters and landslide occurrence. Using index of entropy (IoE), the weight (W_j) for classes of landslide conditioning factors is as follows:

$$W_{\text{j-class}} = I_j P_{ij} Reclass\,;\; I_j = (H_{jmax} - H_j)/H_{jmax}\,;\; H_{jmax} = Log_2 S_j\,;\; H_{jmax} = -\sum_{i=1}^{S_j} (p_{ij}) log_2 p_{ij} \qquad (2)$$

$$(p_{ij}) = p_{ij} / \sum_{i=1}^{S_j} p_{ij}\,;\; p_{ij} = \text{Landslide pixel}(\%) / \text{Class pixel}(\%) \qquad (3)$$

where H_j and H_{jmax} represent entropy values; I_j is the information coefficient; S_j is the number of classes of landslide conditioning factors. $W_{\text{j-class}}$ is the weight class's value.

Using Eqs (2), (3), weight values were calculated for all the classes of landslide conditioning factors (Table 1). Finally, the landslide susceptibility index was calculated.

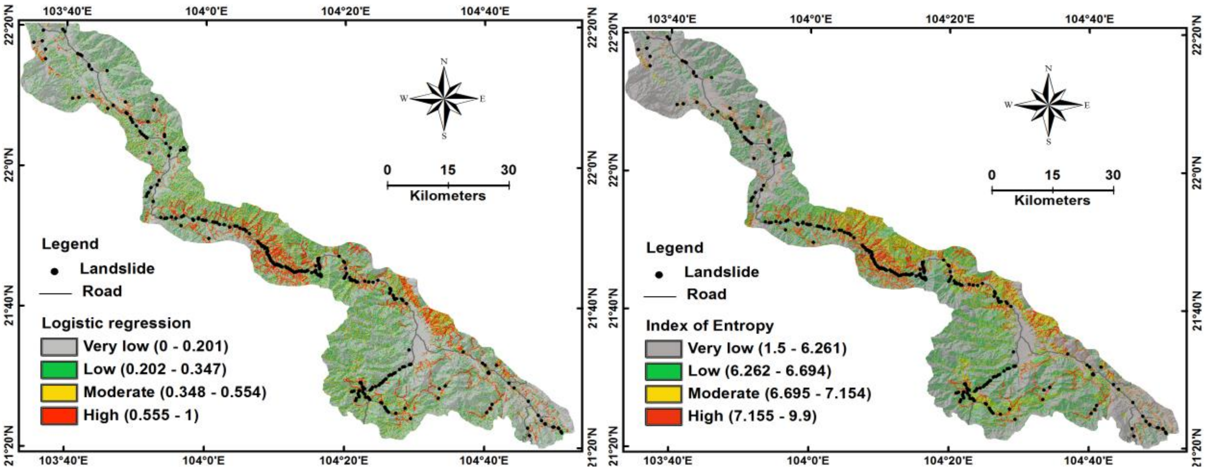

Figure 2. Landslide susceptibility maps using the LR and the IoE models.

4. VALIDATION AND COMPARISON OF LANDSLIDE SUSCEPTIBILITY MAPS

According to Chung and Fabri (2003), the most important and the absolutely essential component in landslide modeling is to carry out a validation of the predicted results. And without validation, the landslide model will have no scientific significance. The success rate was obtained by comparing the landslide pixels in the training dataset with the two landslide susceptibility maps.

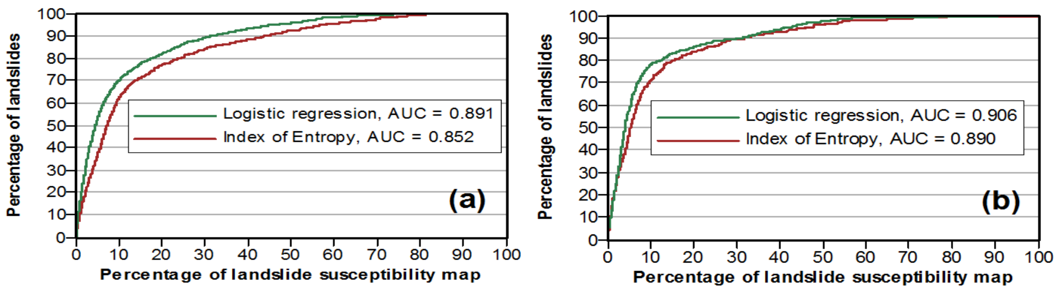

Figure 3. Success rate (a) and prediction rate (b) for the LR and IoE models.

The prediction capability of the two landslide models was assessed by comparing the landslide pixels in the validation dataset with the two landslide susceptibility maps. Furthermore, areas under the curves (AUC) were calculated (Figure 3). The result shows that, the logistic regression (AUC=0.906) is slightly better than the index of entropy model (AUC=0.890). The landslide susceptibility indexes were reclassified into 4 classed based on the percentage of area (Tien Bui *et al*., 2012c): high (10%), moderate (10%), low (20%), very low (60%) (Figure 2).

5. DISCUSSION AND CONCLUSION

In this paper, we have evaluated and compared the results of applying the logistic regression and the index of entropy for landslide susceptibility assessment along the National Road 32 of Vietnam. The two maps present a predicted spatial distribution of landslides. They do not present "when" and "how frequently" a landslide will occur.

The landslide inventory with 282 landslide locations and ten landslide conditioning factors (slope angle, slope aspect, relief amplitude, topographic wetness index (TWI), topographical shape, distance to roads, distance to rivers, distance to faults, rainfall) were used for building the models. Using the success-rate and the prediction-rate methods, the landslide susceptibility maps were validated and compared with landslide locations. Area under the prediction–rate curve is 0.906 and 0.890 for the LR and the IoE models respectively. Although the two models have a good prediction capability, the LR performs better than the IoE model in both success rate and prediction rate.

Acknowledgements

In this paper, we used the original data from the project: "Status and Forecasting of Geo-hazard (landslide and flash flood) along the National Road No.32"- Ministry of Natural Resources and Environment of Vietnam (MONRE). This research was funded by the Norwegian Quota Scholarship program (Norway).

REFERENCES

Bednarik, M., Magulova, B., Matys, M., Marschalko, M., 2010. *Landslide susceptibility assessment of the Kralovany-Liptovsky Mikulas railway case study*. Physics and Chemistry of the Earth, 35, 162-171.

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

Glade, T., 1998. *Establishing the frequency and magnitude of landslide-triggering rainstorm events in New Zealand*. Environmental Geology, 35, 160-174.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006. *Estimating the quality of landslide susceptibility models*. Geomorphology, 81, 166-184.

Korup, O., Gorum, T., Hayakawa, Y., 2012. *Without power? Landslide inventories in the face of climate change*. Earth Surface Processes and Landforms, 37, 92-99.

Lee, S., Dan, N.T., 2005. *Probabilistic landslide susceptibility mapping on the Lai Chau province of Vietnam: focus on the relationship between tectonic fractures and landslides*. Environmental Geology, 48, 778-787.

Sassa, K., Canuti, P., 2008. *Landslides-Disaster Risk Reduction*. Springer.

Tien Bui, D., Lofman, O., Revhaug, I., Dick, O., 2011a. *Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression*. Natural Hazards, 59, 1413–1444.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012a. *Landslide susceptibility assessment in Vietnam using Support vector machines, Decision tree and Naïve Bayes models*. Mathematical Problems in Engineering. Doi:10.1155/2012/974638, 2012.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2011b. *Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS*. Computers & Geosciences. Doi 10.1016/j.cageo.2011.10.031, 45, 199-211.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012b. *Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg-Marquardt and Bayesian regularized neural networks*. Geomorphology.Doi:10.1016/j.geomorph.2012.04.023

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B., 2012c. *Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models*. CATENA, 96, 28-40.

UN-ISDR, 2012. *Viet Nam - Disaster Statistics*..http://www.preventionweb.net/english/countries/statistics/?cid=190