THE APPLICATION OF GIS-BASED LOGISTIC REGRESSION FOR LANDSLIDE SUSCEPTIBILITY MAPPING IN THE SHIHMEN RESERVOIR WATERSHED

Hung-Pin Huang[1*], Bo-wei Lin[2]

ABSTRACT

Since 921 Earthquake happened in 1999, number and scale of landslide are increasing on the hillside of Shihmen reservoir watershed, even to the stage of terminating water supply. In order to decrease the sediment amount entraining reservoir area, it is necessary to make the landslide susceptibility mapping first.

With Logistic regression, the significant factors of landslide in Shihmen reservoir watershed are seismic strength, geology, slope direction, fault, slope, rainfall and elevation. Comparing with the analytic result and construction allocation for landslide by Soil and Water Conservation Bureau in Shihmen reservoir watershed, there is 94.9% of construction sites are located on medium to high susceptible zones and 5.1% located on low susceptible zone.

Key Words: Landslide susceptibility mapping, Logistic regression model, Shihmen reservoir watershed

INTRODUCTION

The Shihmen Reservoir is designed for multiple functions, which is equipped with those functions as irrigation, water supply, flood mitigation, hydraulic power and so on. It has brought great advantages in agricultural and industrial development in Northern Taiwan. Since 921 Earthquake happened in 1999, number and scale of landslide are increasing on reservoir watershed, even to the stage of terminating water supply in 2005. For example, high sediment concentration of 242,000 NTU on Typhoon Aere in 2004, 80,400 NTU on Typhoon Matsa in 2005 and 99,000 NTU on Typhoon Talim in 2005. The cumulative sediment volume is increasing with years from the construction of reservoir in 1964. However, the available storage is 123 million cubic meter and available reservoir volume ratio is 56.10% in March 2009.

To analyze the feasibility of the results, this study uses multivariable regression to set up the landslide susceptible model, associated with considering the safety of building, farmland and roadway, to derive the risk evaluation model. The result could compare the feasibility of results with the landslide treatment of government (Soil and Water Conservation Bureau).

1 Professor, Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C. (*Corresponding Author; Tel.:886-2-33663456; Fax:886-2-23639568; Email: benhuang@ntu.edu.tw)

2 Graduate student, Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C.

LITERATURE REVIEW

There is much literature studying the landslide susceptible mapping on the world. The principle factors could be separated into three categories of geography, hydrology and man-made perturbation. The analyzing measures for landslide susceptibility were used initially by qualititative evaluation without the assistance of computer treatment [Carrara and Merenda, 1974, Varnes, 1984, Lee *et al.*, 1990]. In which, the scholar or specialist of civil engineering, geology, geography, hydrology or soil and water conservation could identify the landslide susceptibility with some academic or empirical indexes. Recently, scholars tries to use computer to set up the landslide model by the information of geography, hydrology and man-made perturbation to calculate the occurrence probability of landslide qualitatively [Chung and Fabbri, 1993, Hearn, 1995, Lee, 2001, 2002, Chang *et al.*, 2007].

Recently, some literature made landslide susceptible mapping with geographic information system by indirect mapping technique. This quantitative measure uses statistical model or specialist decision model associated with the relationship of geomorphologic factors mutually to evaluate the occurrence probability of landslide [Duman and Can, 2006]. The susceptible evaluation and calculation could be accomplished by different mathematical calculations of statistical analyses [Ayalew and Yamagishi 2005, Can *et al.* 2005], index maps [Ayenew and Barbieri, 2005], frequency ratio model [Lee and Pradhan, 2006, Lee and Sambath, 2006], artificial neural networks [Lee *et al.*, 2004, Gomez and Kavzoglu, 2005], fuzzy approaches [Muthu *et al.*, 2008].

The occurrence of landslide is kind of the binary item which is ‘zero’ as happening while as ‘one’ for no occurrence. Hence, the Logistic regression is used frequently to carry out the multivariable analysis [Dai and Lee, 2001, Lee, 2001, Ohlmacher and Davis, 2003, Beguería, 2006]. There are many applications of Logistic regression. For example, The applications were carried out in Northeast Kansas of U.S. [Ohlmacher and Davis, 2003], Kakuda-Yahiko mountainous area in Middle Japan [Ayalew and Yamagishi, 2005], Hendekarea in Turkey [Yesilnacar and Topal, 2005], Flemish Ardennes in Belguim [Van Den Eeckhaut *et al.*, 2006] and Ho-she-creek watershed in Taiwan [Chang *et al.*, 2007]. Evaluation of landslide susceptibility by Logistic regression has got a good result even for different geomorphologic and hydrologic conditions.

The Logistic regression could be integrated geographic information system very well to get landslide susceptibility mapping. And, literature review showed that Logistic regression has a better result than frequency ratio model [Lee et al, 2005]. Moreover, Logistic regression is frequently used to evaluate the risk analysis and get landslide susceptibility mapping.

STUDY INFORMATION

Soeters and van Westen [1996] as well as Lee [2001] suggested, in the analysis of multivariable statistics, the calculation of regression equation with data matrix is more suitable to select medium-scale (1:25,000~1:50,000) of maps. The regional scale, 1:250,000 of map is suitable for analyses of types or density of landslide where as large-scale of map (1:10,000) is for safety analysis with slope stability model. Hence, this study selected medium-scale and large-scale of maps for analysis as shown in Table 1.

This study followed the procedure of evaluation of landslide probability presented by Einstein in 1988. With impact factors, the landslide susceptibility that is analyzed from landslide region and landslide-related factors could be used to get landslide probability. The risk of landslide could be completed by adding safety targets.

Table 1 Scales of maps this study selected

Item	Year of made	Data structure	Scale	Note
Landslide	2005	Image	1:5,000	Aerial photograph
Rainfall	1987～2005	Grid	40m×40m	Maximum cumulative 72-hours rainfall
DTM	2006	Grid	5m×5m	Derive from DTM
Geologic information	1986	.shp(polygon)	1:50,000	Digitized from map
Earthquake	1991~2004	.shp(polygon)	40m×40m	Cumulative PGA value of 16 events of earthquake

The geomorphologic factors includes in elevation, slope, aspect, geology, fault and earthquake. In which, both of slope and aspect factors are transformed by 5m x 5m DTM. With octant and flat zone (smaller than 5%), aspect could be separated into nine zones. Geologic zones are digitized from the geologic map (1:50,000) by Central Geological Survey. And, the influence zone of fault is defined as both sides of 100 meter long of fault by Building Code.

In order to explore the relationship between earthquake, rainfall and landslide, this study collected sixteen events of disaster-induced earthquake occurred from 1991 to 2004. Because there is only one earthquake monitoring station located at Sankuang elementary school in Shihmen reservoir watershed. This study used Kriging method to plot the PGA contour by other monitoring stations neighboring watershed.

In Shihmen reservoir watershed, there are fourteen raingauges of Shihmen, Changhsing, Fuhsing, Hsiayun, Kaoy, Sankuang, Kalaho, Yufeng, Hsiuluan, Anpu, Chenghsipao, Palieng, Paishih and Hsichiushishan. This study collected average annual rainfall data of these raingauges from 1987 to 2005 for analysis.

LOGISTIC REGRESSION MODEL

Logistic regression model is the most popular multivariable regression method for analyzing landslide susceptibility around the world. The Logistic probability function can be shown as the following:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_i x_i)}{1 + \exp(\beta_0 + \beta_i x_i)} \tag{1}$$

The $\pi(x)$ can be carried out by Logistic transformation to general linear regression function:

$$g(x) = \text{logit}[\pi(x)] = \log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_i x_i \tag{2}$$

DISCUSSIONS

Moreover, after Logistic regression being implemented, Likelihood Ratio Test would be carried out for testing each coefficient of its own parameter of regression equation. In which, the hypothesis is setup that the coefficients of regression equation are not equal to zero, respectively. This study checks out each coefficient of disaster-induced susceptible model to understand its own uncertainty and feasibility. The most common test way of Logistic regression is Relative Operating Characteristic (ROC) curve which has shown its high accuracy on probability forecast system [Swets, 1988]. Recently, much literature is presented by using this method to evaluate landslide susceptibility [Ayalew and Yamagishi, 2005, Yesilnacar and Topal, 2005, Van Den Eeckhaut *et al.*, 2006, Beguería, 2006].

By means of MINITAB Program, this study used Logistic regression to get landslide susceptible evaluation model as shown in Equation (3). Each parameter is under the 5 percent of confidence interval.

$$Z = \text{Constant} + \beta_1\times(\text{Elevation}) + \beta_2\times(\text{Slope}) + \beta_3\times(\text{Aspect}) + \beta_4\times(\text{Geology}) + \beta_5\times(\text{Fault}) + \beta_6\times(\text{Ave. annual precipitation}) + \beta_7\times(\text{Earthquake})$$

$$P=\exp(Z)/(1+\exp(Z)) \tag{3}$$

This study used seven tests of Log-Likelihood, Pearson Goodness-of-Fit Tests, Deviance Goodness-of-Fit Tests, Hosmer-Lemeshow, Pairs, Somers' D and Goodman-Kruskal Gamma to explore the hypothesis of Logistic regression. And, the p-value of Log-liklihood located inside the range of confidence interval, where as those of Pearson and Deviance Goodness-of-Fit Test are located outside the confidence interval. Pairs test is used to check hypothesis that two population means are equal. The homogeneity of means is defined if the probability of landslide is higher than the critical value and landslide occurrence exists. Because the homogeneity occupied 75.8%, the accuracy of model is acceptable. Moreover, Somers' D and Goodman-Kruskal Gamma showed the similar positive results.

From Equation (4), the most significant factor of Logistic regression is earthquake; the factors of rainfall, geology, aspect, fault, slope and elevation are less in sequence. And, the landslide susceptible mapping by this model is plotted as Figure 1.

$$\begin{aligned} Z= \ & (-0.001402)\times(\text{Elevation}) \\ & + (0.014163)\times(\text{Slope}) \\ & + (\text{Aspect}) + (\text{Geology}) \\ & + (-0.54385) \times (\text{Fault}) \\ & + (0.026519) \times (\text{Max. 72 hours cumulative rainfall}) \\ & + (2.58404) \times (\text{Cumulative PGA}) \end{aligned} \tag{4}$$

This study defined "high" landslide probability level is from 0.219 to 0.01 got from Equation (3), where as "medium" level is from 0.01 to 0.002 and "low" level is from 0.002 to 0. Following with the procedure of Yesilnacar and Topal [2005], the ratio of each landslide susceptible probability to total watershed area (PTA), and the ratio of each landslide susceptible probability to existed landslide area (PLB) and PTA/PLB were calculated. That is, "high" landslide probability should have high PLB. These results are shown in Table 2.

From Table 2, PLB value is high for "high" landslide probability, where as very low for "low" landslide probability. Fortunately, the ratio of total landslide area to watershed area is around 0.01, this insignificant result could be neglected.

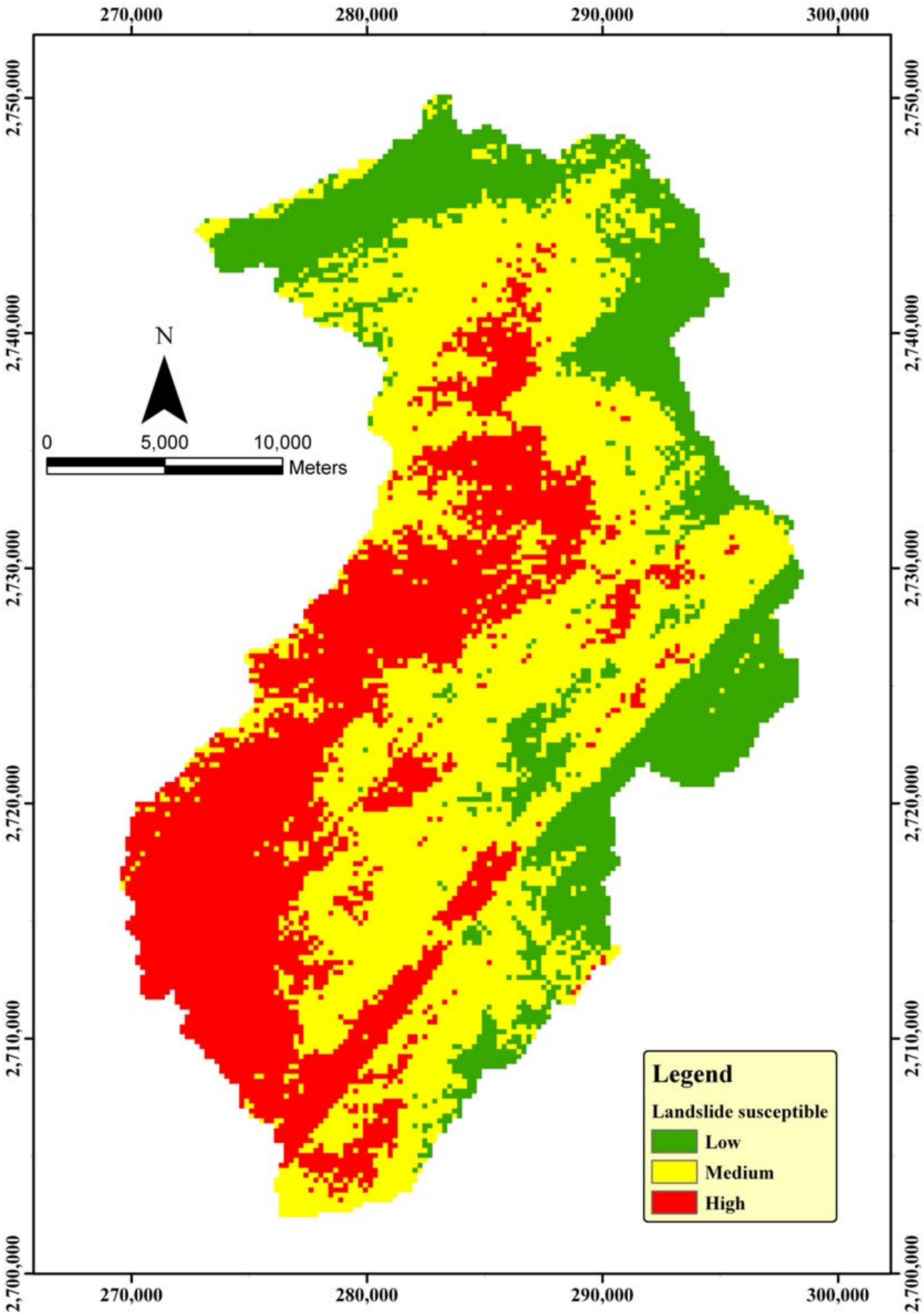

Fig. 1 Landslide susceptible mapping

Table 2 Results of different probability levels of landslide

Probability levels	PTA & PLB		
	PTA	PLB	PTA/PLB
High	31.48%	73.54%	0.43
Medium	46.16%	28.23%	1.64
Low	22.36%	0.66%	33.64

The area under ROC curve, called AUC, could identify the feasibility of model. In which, the range of AUC is from 0.5 to 1. The larger the AUC value is, the more accuracy is. In this study, that the AUC value is 0.791 has shown the predictive result of this model could be acceptable.

Table 3 AUC and its confidence interval

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.791	0.003	0.000	0.785	0.796

To present the feasibility of allocation of landslide treatments by Soil and Water Conservation Bureau (SWCB), the landslide susceptible mapping is overlapped by construction sites. And, most of the construction sites are located inside the "high" and "medium" landslide susceptible ranges. Hence, the allocation of landslide treatments by SWCB is feasible in Shihmen reservoir watershed.

Table 4 showed that there is 94.9% of construction sites located inside "medium" and "high" susceptible zones where as 5.1% inside "low" zone.

Table 4 Result of landslide susceptibility overlapping construction sites

Landslide susceptibility	First-stage construction sites located in susceptible zone	Second-stage construction sites located in susceptible zone	Total construction sites	Ratio
High	130	62	192	49.1%
Medium	122	57	179	45.8%
Low	16	4	20	5.1%

CONCLUSIONS

With Logistic regression, the significant factors of landslide in Shihmen reservoir watershed are seismic strength, geology, slope direction, fault, slope, rainfall and elevation.

Comparing with the analytic result and construction allocation for landslide by Soil and Water Conservation Bureau in Shihmen reservoir watershed, there is 94.9% of construction sites are located on medium to high susceptible zones and 5.1% located on low susceptible zone.

REFERENCES

Goan, L. H. and Chen, J. S. (1995), "The Third Forest Resources and Landuse Inventory in Taiwan," *Forest Bureau*, Taipei, Taiwan.

Beguería, S. (2006), "Changes in land cover and shallow landslide activity: A case study in the Spanish Pyrenees," *Geomorphology*, 74, 196-206.

Davis, T. J., Klinkenberg, B. and Keller, P. C. (2004), "Evaluating Restoration Success on Lyell Island, British Columbia Using Oblique Videogrammetry," *Restoration Ecology*, 12(3), 447–455.

Dai, F. C. and Lee, C. F. (2001), "Terrain-based mapping of landslide susceptibility using a geographical information system: a case study," *Can. Geotech. J.*, 38, 911–923

Francescato, V., Scotton, M., Zarin, D.J., Innes, J.C. and Bryant, D.M. (2001), "Fifty years of natural revegetation on a landslide in Franconia Notch, New Hampshire, U.S.A.," *Can. J. Bot*., 79, 1477–1485.

Gray, D. H. and Sotir, R. B. (1996), "Biotechnical and soil bioengineering slope stabilization: a practical guide for erosion control," *John Wiley & Sons,* New York, p.56.

Lee, S. and Min, K. (2001), "Statistical analysis of landslide susceptibility at Yongin, Korea," *Environmental Geology*, 40, 1095–1113.

Khazai, B. and Sitar, N. (2000), "Assessment of Seismic Slope Stability Using GIS Modeling," *Geographic Information Sciences*, 6(2):121-128, 2000.

Ohlmacher, G. C. and Davis, J. C. (2003), "Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA," *Engineering Geology*, 69, 331–343.

Su, H. J. (1984), "Studies on the climate and vegetation types of the natural forests in Taiwan (II) Altitudinal vegetation zones in relation to temperature gradient," *Quart. Journal Chin. For.,* 17(4): 57-73.

Varnes, D. J. (1978), *Slope movement types and processes, Landslide analysis and control.* In: Schuster R. L. Krizek R. J. (eds) Special Report 176. *Transportation Research Board*, Washington, DC, pp. 11-80.

Carrara, A. and Merenda, L. (1974), "Methodology for an inventory of slope instability events in Calabria (southern Italy)," *Geologia Applicatae Idrogeologica*, 9:237~255

Hearn, E. L. and Wilson, R. C. (1995), "Shaking Intensity Thresholds for Rock Falld and Slides: Evidence form 1987 Whittier Narrows and Superstition Hills Earthquake Storng-Motion Records," *Bull. Seis. Soc. Am.*, 85(6):1739~1757.

Lee, S. W., Cheng, J. D., Ho, C. W., Chiang, Y. C., Chen, H. H. and Tsai, K. T. (1990), "An Assessment of Landslide Problems and Watershed Management in Taiwan, Rearch Needs and Applications to Reduce Erosion and Sedimentation in Tropical Steep Lands," Proceeding of the Fiji Symposium, Jun 1990: *LAHS-AISH Publ.* No. 192.

Varnes, D. J. (1984), *Landslide Hazard Zonation: A Review of Principle and Practice*, UNESCO Press, Paris, pp.63.

Chung, C. J., and Fabbri, A. G. (1993), "The Representation of Geoscience Information for Data Integration," *Nonrenewable Resources*, 2(3):122~131.

Lee, S., Choi, J. and Min, K. (2002), "Landslide susceptibility analysis and verification using the Bayesian Probability model," *Environmemtal Geology (Berlin)*, 43:120~131

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

Yesilnacar, E. and Topal, T. (2005), "Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey)," *Engineering Geology*, 79, 251-266

Van Den Eeckhaut, M., Vanwallwghem, T., Poesen, J., Govers, G., Verstraeten, G. and Vandekerckhove, L. (2006), "Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium)," *Geomorphology*, 76, 392-410.