Research article

Comparison of Semi-Quantitative and Statistical Regression Models in Assessing Landslide Prone Areas in Wahig-Inabanga Watershed, Bohol, Philippines

TOMAS D. REYES, JR.*
Bohol Island State University, Bohol, Philippines
Email: tomseyer@yahoo.com / tomseyer@gmail.com

Received 3 March 2015 Accepted 10 November 2015 (*Corresponding Author)

Abstract The study attempted to compare and evaluate the two landslide hazard assessment models, semi-quantitative (index-based) and statistical regression (bivariate statistical analysis and logit regression) in predicting landslide prone areas in Wahig-Inabanga Watershed, Bohol, Philippines. This was performed by comparing the predictive power of each model based on the frequency distribution of past landslide events. Findings revealed that the combined bivariate statistical analysis and logit regression model outdone index-based method in predicting landslide occurrences. Results indicated high prediction accuracy on statistical model greater than the 75% threshold level set for evaluation on both pooled moderate to very high hazard zone and the combined high and very high hazard zone with accuracy values of about 83.82% and 76.72%, respectively. Conversely, the semi-quantitative model failed to meet the accuracy threshold. The study showed that statistical regression model, though relatively difficult to implement, can be a better substitute to the most commonly used semi-quantitative method as a decision-support tool for watershed management and land use planning in relation to landslide risk mitigation, reduction, adaptation, and management.

Keywords bivariate statistical analysis, landslide hazard, logit regression, semi-quantitative method

INTRODUCTION

Several computer-based tools are found useful in landslide prediction, hazard assessment, and mapping especially when these tools are made use in tandem with geographic information system (GIS). GIS serves as an indispensable tool for mapping areas prone to unpredictable hazard events, particularly landslides. One of the best advantages of using this technology is the possibility of improving hazard occurrence models by evaluating results and adjusting the input variables (Lanuza, 2008).

There are several methods used in landslide hazard assessment. Ayalew et al., (2005) and Reyes (2014) briefly discussed each method and grouped them into three major categories: semi-quantitative, quantitative, and hybrid. According to Ayalew et al., (2005) and also cited by Reyes (2014), some of the methods are simple, especially those which rely on subjective assessments. Others, however, depend on complex mathematical concepts and are difficult to understand. Some old approaches have long disappeared, others underwent a sort of refinement, and new methods are always coming. Many of the latest methods are not yet available in known commercial GIS packages either as built-in functions or additional modules. Data, then, are usually transformed to external software products for core analyses.

The semi-quantitative method is the collective process of index and overlay analysis, thus termed “index-based method”. It is also called as expert-driven (Zhu and Huang, 2006) in which expert opinions make great difference and become the basis during assessing of the type and degree of any natural hazard. In the Philippines, it is commonly used in provincial and municipal local government units for disaster risk reduction and management and even recommended by the Department of Environment and Natural Resources (DENR) to be a decision-support tool in forest management and conservation planning. A vulnerability assessment manual adopting this method has been prepared by the Ecosystems Research and Development Bureau (ERDB) and made available for public use since 2011.

On the other hand, the statistical regression model, uses logit regression to develop a functional relationship between a process and factors inherent in them. The applications of this model in the field of slope instability have evolved as an important tool, with specific reference to landslide hazard mapping. In landslide hazard mapping, an area is classified according to relative classes of instability on the basis of the degree of occurrence of landslide and mass movements (Jade and Sarkar, 1993).

In this study, both methods were applied for landslide hazard assessment in Wahig-Inabanga Watershed, Bohol, Philippines to determine which between the two models is more appropriate in predicting future landslide events based on the frequency distribution of past landslide occurrences.

METHODOLOGY

Landslide Hazard Mapping using Semi-quantitative (index-based) Method

The landslide hazard map prepared using the semi-quantitative method, particularly the index-based, was completed following the procedures suggested by the ERDB-DENR in its vulnerability manual published in 2011. It involved division of pre-defined landslide-related instability factors such as slope, soil type, rainfall, lithology, and land use into 5 classes using a set of criteria that influence vulnerability of the study area to landslide. These criteria were also used in assigning class ratings. The most influential class trait was given the highest rating of 1, while the least influential was rated 0.

This was followed by overlaying of instability factors based on desired factor weights. Weights used were 0.35 for slope, 0.20 for both rainfall and geology, 0.15 for land use, and 0.10 for soil type.

Landslide Hazard Mapping using Logit Regression (Combined Bivariate Statistical Analysis and Logit Regression) Model

The statistical regression model, same with the index-based method, also necessitated factor and class weighing. Bivariate statistical analysis was used to determine class weights, while logit regression allowed the computation of factor weights. However, the logit regression, unlike semi-quantitative method, required the utilization of landslide inventory or landslide occurrence map (van Westen, 1994 as cited by Wahono, 2010) to implement factor and class weighing. This means that factor and class weights are dependent on the landslide inventory and not on pre-defined vulnerability or susceptibility criteria. To do this, the landslide inventory map was overlaid with nine significant landslide-related instability parameters like elevation, slope, aspect, lithology, distance from fault line, distance from rivers, distance from roads, rainfall, and land use. Landslide pixels laid on each class of instability factors were computed as landslide frequencies. These frequencies served as class weights and were used as class numerical values in logit regression. Important outputs of logit regression in SPSS included regression coefficients of all parameter considered as factor weight and the model prediction probability.

Details on how these two maps were generated are discussed in the DENR Vulnerability Assessment Manual (ERDB, 2011) for the index-based method and the works of Reyes (2014), Ayalew et al., (2005) and Ayalew and Yamagishi (2005) for the logit regression with bivariate statistical analysis referred in their reports as quantitative method.

Comparison of the Two Models

Model comparison was performed to determine which of the two approaches was more reliable in landslide hazard prediction. Comparison was based on the frequency distribution of past landslide events [=pixels] rested on the pooled upper moderate to very high hazard zones [$P(Y=1) \geq 0.5$ logit regression default cut-off value] and the combined zone rated as high and very highly [$P(Y=1) \geq 0.6$] prone to landslide occurrences using the 75% model prediction accuracy threshold. This was done by applying the overlay and extract by sample function in spatial analyst tool of ArcGIS.

RESULTS AND DISCUSSION

Landslide Hazard Assessment

Semi-quantitative method: Table 1 presents the summary results of landslide hazard assessment using the semi-quantitative method. Based on Table 1 and depicted in Fig. 1, the biggest part of the watershed, about 71.50% or 44,540 ha, was predicted moderately prone to landslides. Considerable areas had estimates of low (10,400 ha) and high (7,338) hazard ratings, while very small areas of the watershed were estimated very low (4 ha) and very high (13 ha). From these results, it appears that the semi-quantitative method overestimated the moderate landslide hazard zones and underestimated the very low and very high landslide hazard areas. As shown in Fig. 1, most of the relatively flat areas in the watershed fell within the moderate landslide hazard zone.

Table 1 Landslide hazard class ratings, area of coverage (ha) and percent distribution generated using semi-quantitative method

Class Range	Rating	Area (ha)	Percent (%)
< 0.2	Very low	4	0.01
0.2-0.4	Low	10,400	16.69
0.4-0.6	Moderate	44,540	71.50
0.6-0.8	High	7,338	11.78
> 0.8	Very high	13	0.02
Total		62,295	100

Table 2 Landslide hazard class ratings, area of coverage (ha) and percent distribution generated using statistical regression model

Class Range	Rating	Area (ha)	Percent (%)
< 0.2	Very low	38,180	61.29
0.2-0.4	Low	10,360	16.63
0.4-0.6	Moderate	6,692	10.74
0.6-0.8	High	4,101	6.58
> 0.8	Very high	2,962	4.75
Total		62,295	100

Statistical regression model: Table 2 shows the landslide hazard class ratings generated using the statistical regression model, and their corresponding area (ha) and percent distribution. Results indicate that more than 60% of the total area of the watershed (about 38,180 ha) was identified to have very low probability of landslide occurrence. About 16.63% or 10,360 ha had low landslide hazard, while roughly 6,692 ha or 10.74% was estimated to fall under the moderate landslide class. Conversely, high and very high landslide ratings were predicted for areas mostly situated on the upper elevations of the watershed (Fig. 2) having 4,101 ha and 2,962 ha, respectively.

The results show a decreasing area distribution against the increasing vulnerability of the area to landslide.

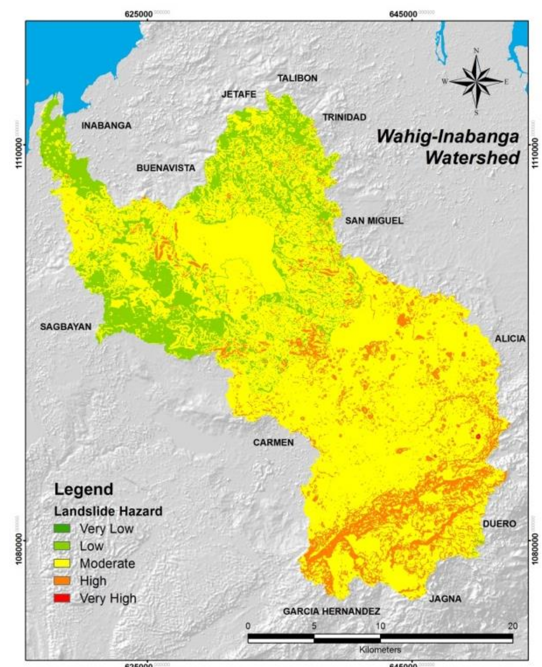

Fig. 1 Landslide hazard map generated using semi-quantitative method

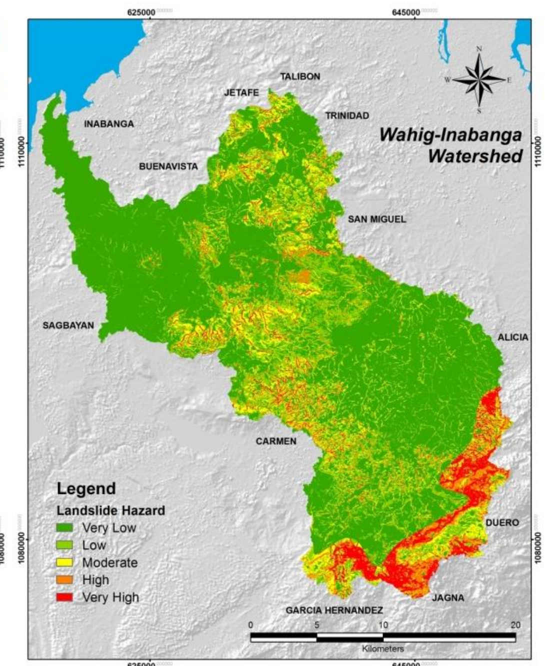

Fig. 2 Landslide hazard map generated using the statistical regression model

Table 3 Comparison of two models showing the frequency distribution of landslide pixels in different hazard classes

Hazard		Statistical regression		Semi-quantitative	
Class range	Rating	Frequency	Percent	Frequency	Percent
< 0.2	Very Low	83	4.39	0	0.00
0.2-0.4	Low	124	6.56	48	2.54
0.4-0.6	Moderate	232	12.27	1,005	53.15
0.6-0.8	High	364	19.25	772	40.82
> 0.8	Very High	1,088	57.54	66	3.49
Total		1,891	100	1,891	100

Method Comparison

Table 3 indicates the result of the landslide inventory layer and the landslide hazard maps overlay. It is noticeable that there was a direct agreement between the landslide frequency (= number of pixels lying on each hazard class) and the hazard zones for the statistical regression model, a characteristic of an ideal method (Fig. 3). The highest landslide frequency of 1,088 or 57.54% was obtained from the very high landslide hazard zone. This was followed by high and moderate hazard zones with 364 (19.25%) and 232 (12.27%), respectively.

Conversely, the distribution of landslide pixels was variable among the hazard zones of semi-quantitative model, thus no relationship was observed. The result on Table 3 clearly shows that most of the landslide pixels were found on moderate (1,005 or 53.15%) and high (772 or 40.82%) landslide hazard zones, while only 66 pixels (3.49%) fell on very high hazard zone.

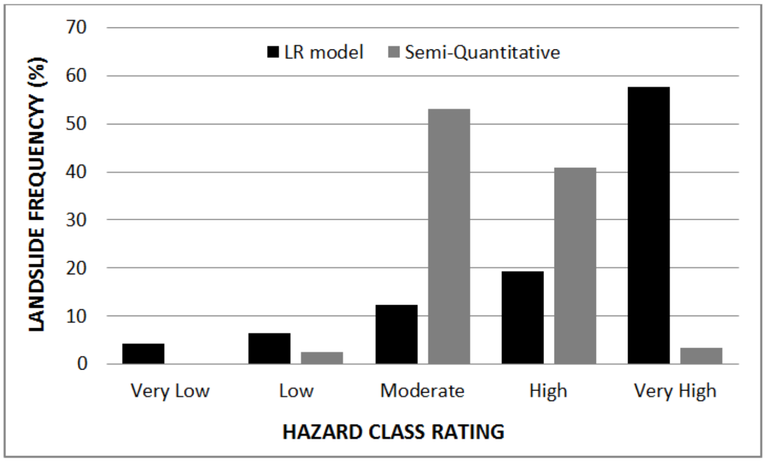

Fig. 3 Comparison of two models based on the percent distribution of landslide pixels in different hazard classes

A process of combined zonation (combining classes) was also used to clearly evaluate the predictive power of each model. The combined zone is referred to as unstable zone (=area) in the study of Dhakal et al., (2000) such as the pooled upper moderate to very high hazard, and the high and very high landslide hazard classes. Table 4 reveals the result of the model comparison based on the computed landslide frequency on these combined zones. Compared to the semi-quantitative method, statistical regression model, at par, had higher prediction accuracy values of 83.82% and 76.72% [both greater than the 75% threshold level] based on the frequency and percentage of landslide events that fall on moderate to very high [P(Y=1) ≥ 0.5] and high and very high [P(Y=1) ≥ 0.6] hazard zones, respectively.

Table 4 Comparison of two different models showing the frequency distribution of landslide pixels

Model	P(Y=1) ≥ 0.5*		P(Y=1) ≥ 0.6**	
	Frequency	Percentage	Frequency	Percentage
Semi-Quantitative	1,386	73.29	838	44.32
Statistical Regression	1,585	83.82	1,452	76.78

*Note: * = default logit regression cut-off value in SPSS; upper moderate to very high hazard zone*
*** = areas rated as high and very highly prone to landslide*

With lower computed prediction accuracy, the semi-quantitative method, unfortunately, failed to meet the threshold level set for acceptability which only means that this method is not suitable for landslide hazard assessment and mapping particularly in Wahig-Inabanga Watershed. The statistical regression model, then, becomes a better alternative method and substitute to semi-quantitative or index-based method.

CONCLUSION

Based on the findings of the study, it is concluded that the statistical regression method is a better option to use when assessing landslide hazards in Wahig-Inabanga Watershed, Bohol, Philippines. The advantage of applying a bivariate statistical analysis provided numerical values on instability factor classes which were used in determining factor weights through logit regression. The idea of factor weighing in logit regression is to find the best fitting function in defining the relationship between the presence or absence of landslides and a set of landslide-related instability parameters. The objectivity of logit regression method in determining the significance of instability parameters in landslide prediction is wanting in semi-quantitative or index-based method.

REFERENCES

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.

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

Dhakal, A.S., Amada, T. and Aniya, M. 2000. Landslide hazard mapping and its evaluation using GIS, An investigation of sampling schemes for a grid-cell based quantitative method. Photogrammetric Engineering and Remote Sensing, 66 (8), 981-989.

ERDB. 2011. Manual on vulnerability assessment of watersheds. Ecosystems Research and Development Bureau, Department of Environment and Natural Resources, College, Laguna.

Jade, S. and Sarkar, S. 1993. Statistical models for slope instability classification. Engineering Geology, 36, 91-98.

Lanuza, R.L. 2008. Soil erosion and landslide vulnerability of Mananga Watershed, Cebu, Philippines. SYLVATROP, 18, Nos 1-2, January-December.

Lopez, A.V.B., Estigoy, D.A., Tubal, R.S., Andrada, M.G., Baldo, H.S., Dano, A.M. and Ebora, J.B. 2008. Landslide and fire vulnerability assessment of Pudong Watershed, Benguet, Philippines. SYLVATROP, 18, Nos 1-2, January-December.

Reyes, T. Jr., D. 2014. Modeling of landslide and water-induced erosion in Wahig-Inabanga Watershed, Bohol. Unpublished Doctoral Dissertation. University of the Philippines Los Baños.

Van Westen, C.J. 1994. GIS in landslide hazard zonation, A review with examples from the Andes of Colombia. In Mountain Environment & Geographic Information System. Edited by Price, M.F. and Heywood, D.I.

Wahono, B.F.D. 2010. Applications of statistical and heuristic methods for landslide susceptibility assessments. Unpublished Technical Report. Gadjah Mada University and International Institute for Geo-Information Science and Earth Observation.

Zhu, L. and Huang, J.F. 2006. GIS-based logistic regression method for landslide susceptibility mapping in regional scale. Journal of Zhejiang University, SCIENCE A. ISSN 1009-3095 (Print), ISSN 1862-1775 (Online). www.zju.edu.cn/jzus; www.springerlink.com.