Multivariate Logistic Regression Approach for Landslide Susceptibility Assessment of Antipolo, Rizal

Jaime Angelo S. Victor and Mark Albert H. Zarco

Institute of Civil Engineering, College of Engineering
University of the Philippines Diliman, Quezon City 1101

Abstract – *Slope instability associated with heavy rainfall or earthquake is a familiar geotechnical problem in the Philippines. This study aims to perform a detailed landslide susceptibility assessment of Antipolo City using a statistical approach. In this study, morphologic and non-morphologic factors contributing to landslide occurrence and their corresponding spatial relationships were considered. The multivariate logistic regression was performed in randomly selected datasets based on the landslide inventory. These were divided into training and test data sets based on K- cross fold validation scheme resulting to different models. The model selected for the final implementation has an overall accuracy of 91.66%, AUROC of 0.908, standard error of 0.002 and RMSE of 0.2478. Cross validation with deterministic approach using physically based slope stability models were performed, where there was no significant difference between the two approaches in identifying areas of highly and very highly susceptible to landslide occurrence. The study also shows that almost 40% of Antipolo City has been assessed to be potentially dangerous areas in terms of landslide occurrence. .*

Keywords—Landslide Susceptibility Assessment, Logistic Regression, Slope Stability

I. INTRODUCTION

Landslide is defined as the downward movement of slope materials, such as rock and soil, triggered by gravity due to other natural hazards such as earthquake and intense amount of rainfall. This phenomenon is generally affected by site morphology, geology and human activities, to name a few. Preparatory variables are site factors that cause marginal stability to the area. These factors include gradient, aspect, slope materials, drainage conditions are that may be attributed to failure without considering the landslide initiating factors. Natural triggering factors, such as excessive rainfall, seismic activities, volcanic activities, erosion and storm waves, cause the transition from marginally stable to unstable conditions.

There are diverse methods in landslide susceptibility mapping. Susceptibility is defined as the tendency of landslide occurrence in an area. Susceptibility is mathematically represented as the probability of spatial occurrence of documented landslides for a specific set of geo-environmental conditions [1].

In recent years, scientists all over the world have studied different GIS-based approaches in susceptibility assessment. Many have investigated particular areas using a single method; may be it heuristic, statistical, or deterministic. Others claim that a deterministic approach coupled with a defined hill slope hydrology is the best approach for spatial landslide prediction [2]. The use of GIS is an effective tool in landslide susceptibility mapping because it is capable of applying both quantitative and qualitative approaches in the analysis. One of the main benefits of using GIS is its capacity to do spatial data analysis, which is particularly useful in natural disaster assessment.

A validation based on a well-defined deterministic stability model would be a good verification of the reliability of the other models. This study aims to perform landslide susceptibility assessment of Antipolo, Rizal using logistic regression, with a supplement cross validation with deterministic approach.

Landslides are significant geomorphologic threats to lives and property. Annually, these phenomenon damages forests, agricultural land, residential and industrial areas. In Antipolo City, there is a current estimation that 40% of the total area are inhabited or developed. Despite the presence of steep slopes on mountainous regions, the development of the remaining area in the future, whether nearby or distant, is inevitable. The generation of a site-specific landslide susceptibility study will be beneficial to identify potential landslide event in particular locations. Planning control and effective zoning of the city may also be accomplished.

II. METHODOLOGICAL FRAMEWORK

2.1 *Digital Elevation Model*

The digital elevation model (DEM) used in this study is obtained from the DOST Disaster Risk and Exposure Assessment for Mitigation (DREAM) Project. This DEM, which was post-processed by the UP Training Center for Applied Geodesy and Photogrammetry (UP-TCAGP), was generated from Synthetic Aperture Radar (SAR). It has a 10-meter resolution and 10x10km tile size. The morphometric factors considered in this study are derived from the DEM.

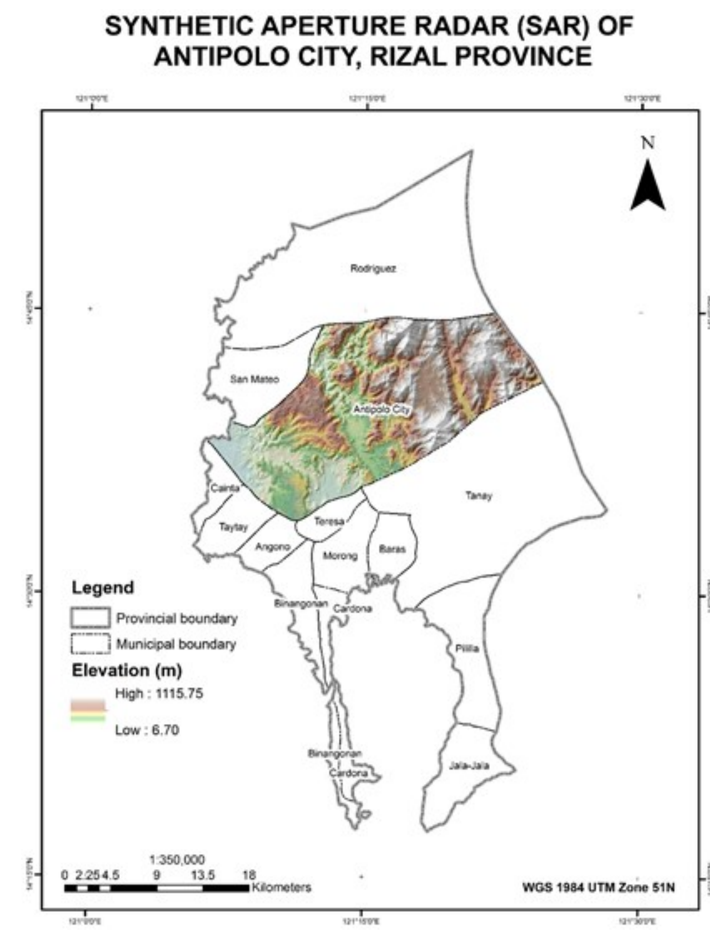

Figure 1. Digital Elevation Model of Antipolo City, Rizal

2.2 *Landslide Causal Factors*

This study considers the following landslide causal factors related to Antipolo City, Philippines. Parameter maps for 2.2.1 to 2.2.8 are derived from the digital elevation model through raster analysis in ArcGIS [3].

2.2.1 *Slope Angle*

Slope gradient is commonly viewed as a major contributing factor in landslide formation. It is the most important factor that needs to be taken into account as the principal factor in landslide susceptibility assessment [4].

2.2.2 *Elevation*

The relationship of elevation and landside occurrence have been discussed and cited in several studies [5]. It is because elevation is attributed to other factors such as slope, lithology, weathering, precipitation, ground motion, soil thickness, and land use.

2.2.3 *Curvature*

Landslide occurrences have also shown strong relationship with surface curvature. Curvature of an area may be categorized as planform curvature and profile curvature. It affects the convergence and divergence of flow across the surface. [6]

2.2.4 *Slope Aspect*

The exposure of slope, which may be described with respect to the cardinal directions, may affect the distribution of landslide occurrence. The facing of the slope affects hydrologic processes thus affecting weathering processes and vegetation and root development, especially drier environments [7].

2.2.5 *Topographic Wetness Index*

The effect of topographical characteristics of the area and run off generation may be described by the topographic wetness index (TWI). The use of TWI provides a mean of quantitative simulation the soil moisture conditions in a watershed [8]. High TWI values represent drainage depressions (steep, convex areas will shed water). Low values represent crests and ridges (concave, low gradients areas will gather water).

2.2.6 *Stream Power Index*

The stream power index (SPI) measures the potential erosive power of overland flow and it is considered as one of the factors affecting landslide occurrence [9]. High SPI values indicate a high likelihood of erosion in that area. Data above the 85th percentile is considered high and can identify areas of overland erosion.

2.2.7 *Road Network and Distance to Roads*

Roads on slopes may be treated as discontinuities and may constitute a barrier or a corridor for water flow and may induce instability. There is a significant correlation between the extent and frequency of landslide, and the distance to regional road system [10]. A constructed road beside slopes will decrease the load on both the topography and on the toe of the slope hence increasing the stress on the back of the slope. The decrease of load on slope toe may also result to the development of tension cracks.

2.2.8 *Distance from Water Body*

Previous studies [11] stated that there is a close spatial relation between landslides and the presence of watercourse or dense drainage lines. The degree of saturation of the materials on the bottom of the slope is one of the controlling factors of the stability of the area.

2.2.9 *Soil and Geology*

Soil characteristics are major factors contributing to slope stability. Surface soils factors are treated as independent causal factors that lead to landslide occurrence. The various structures of earth materials tend to lead to a variation in the stability, strength, and texture of rocks and soil [12]. Moreover, the underlying bedrock present in a site shows direct association with the site's stability [7]. Unstable bedding sequence is often a point of concern in landslide hazard analysis.

Soil and geological data were obtained from the Bureau of Soil and Water Management.

2.3 *Landslide Inventory*

Except for highly broadcasted and fatal landslides that occurred in the city (i.e. Cherry Hills landslide tragedy), very little data were available from the local disaster risk, reduction and management (DRRM) office of Antipolo with regard to past landslide incidence in the city. Most available records were those that resulted to fatalities and/or property damage. To properly understand landslide occurrence, landslides that occurred on uninhabited places such as forests and mountainous regions are also considered.

The landslide inventory data used in this study were obtained from two sources. The first source is the available landslide information from the Project Nationwide Operational Assessment of Hazards (NOAH) under the Department of Science and Technology (DOST). The second is from a landslide inventory data sourcing from independent interpretation of aerial photography (Google Earth [13] uses the GeoEye-1), satellite image evaluation and digital elevation analysis [14]. The union of the two sets were used and crosscheck was also performed to investigate possible overlaps. There were 1580 landslide events considered in the analysis. However, only spatial distribution is considered in this study. Temporal distribution is not considered. The mapping by grids resulted to 10,781 landslide pixels distributed to the study area. The total number of pixels for the whole study area is 3,527,765.

2.4 *Statistical Approach – Logistic Regression Model*

In multivariate logistic regression, the normal distribution of the dependent variable is not required. It assumes a linear relationship between the outcome and the logit of the independent variables. Also, there are no assumptions required with regards to the homogeneity of the variance and normally distributed error terms [14].

Various input for the independent variable considered in the analysis are allowed. It may be discrete, dichotomous, continuous or any combination of these, while the dependent variable is dichotomous [15]. The two possible values of the dependent variable may be regarded as presence/ absence, success/failure, or an event occurring/not occurring. The output prediction value is then represented as a probability between 0 and 1.

The logistic function f(z) [16] is defined as:

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

where z is defined as the linear sum of the product of the independent variables and their respective coefficients, and a constant.

$$z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where β_i(i=0,1, 2,…,n) are the coefficients, x_i (i=0,1,2,…,n) are the values of the independent variables, and α is a constant. The value of f(z) varies from 0 to 1 since z varies from $-\infty$ to $+\infty$.

The independent variables are the physical controlling factors that affect the instability of an area and the dependent variable is the presence or absence of landslides. Because of the binary response, two alternative groups are established; mapping units free of landslides, and mapping units having landslides.

The dependent variable, which is the observed landslide occurrence (1580 landslide events), is coded as 1 and 0 for presence or absence of landslide respectively. The statistical analysis was performed using randomly selected training data sets with equal number of presence and absence of landslide value. A total of 17320 cases were chosen as subsamples for determining the relationship of the dependent to the independent variables.

Multicollinearity tests were also done on the set of independent variables by computing tolerance and variance inflation factor (VIF) values. An evidence of presence of multicollinearity is tolerance value of less than 0.2, and consequently, a variance inflation factor of greater than 5 [17]. The VIF reflects how multicollinearity inflates the variance of the coefficient estimates. As an example, the standard error would be three times as large if the VIF for a variable is 9.

Before performing the regression analysis of the datasets, dummy variables for the categorical variables were created. Since some of the variables will be insignificant in the model, it is much easier to remove it by means of the generation of dummy variables. Dummy variables were also used as binary representation of categorical data (e.g. soil type, lithology) based on the presence (1) and absence (0). These were implemented in ArcGIS [3].

The data were then exported to SPSS™ [18] for logistic regression analysis using the binary logistic tool. The analysis was repeatedly performed as the insignificant variables were removed. Insignificant variables were determined as variables exceeding the threshold value of 0.05 (level of significance) [19]. Variables exceeding the tolerance value accept the null hypothesis thus affecting the model insignificantly. The coefficients of the variables in the model were identified after all the variables portray significant influence in the occurrence of the event.

2.5 *Modelling and Validation Scheme*

There was a total of 10,781 landslide pixels covering the study area. To eliminate bias, uniformly sampled non-landslide points are paired with equal number of landslide points as test data set [5]. From this total, the test data set that was used in the modelling consists of 8600 randomly selected landslide pixels (80% of the total landslide pixels) and an equal number of randomly selected non-landslide pixels. The non-landslide pixels were sampled from the study area considering a 50-meter buffer from identified landslides.

To assess the predictive performance of a generated model, K-fold cross validation scheme was used. Considering the bias of the technique, a usual choice of the number of folds, k, is 5 or 10 [20]. The set of 17,200 selected cells was divided into 5 subsets (k=5), wherein each set has equal number of landslide (LS) and non-landslide (NLS) cells (LS:NLS=1). This scheme resulted to five models estimating landslide susceptibility of Antipolo City. Model 1 for landslide susceptibility was generated by using the first four subsets as the training data and the remaining subset as independent validation data. Model 2 was generated using the 1st, 3rd, 4th and 5th subset as training data and the 2nd subset as validation data. The process was continued to generate Models 3 to 5.

In order to determine the variation of regression results based on the input data, a sensitivity analysis was done by using input training data sets with a landslide to non-landslide event relationship of 2, 3 and 4. Random sampling was done while maintaining the total number of calibration set and data set as that of LS:NLS of 1. The same process for estimation of the coefficients was implemented.

2.6 *Measures of Model Performance*

The performance of each model was measured using the corresponding test data set by creating a confusion matrix. The coefficients of the model were used to manually compute the predicted probability of each case of the test data set. A cut-off value of 0.5 was used for the predicted probability. The predicted probability of each case is then compared to the observed outcome of the landslide occurrence. The correctly predicted and wrongly predicted cases of occurrence and non-occurrence of landslide were tabulated to compute for the overall correct percentage of each model. Moreover, the usefulness of a logistic regression model is characterized by a 25% increase in the accuracy compared with proportional by chance accuracy [21].

The measure of how well each model classifies each landslide used in model training is called success rate. To assess the predictive capability of the model, receiver operating characteristic (ROC) curves were plotted. The curve is a plot of the proportion of objects correctly classified (true positives) vs. the proportion of the objects wrongly classified (false positives) [22]

In addition to evaluating AUROC (area under ROC), standard error and root-mean-square error have been used as standard statistical metrics to measure model performance [23].

2.7 *Comparison with Deterministic Approach*

The deterministic approach of evaluating landslide susceptibility is governed by using slope stability models in determining a factor of safety of the area. These models provide valuable quantitative data in landslide risk analysis. The main disadvantage of performing deterministic analysis is the amount of input data required, mostly coming from actual field measurements and laboratory tests. Because of this requirement, this approach is generally applied on site specific, large-scale analysis covering a relatively small area [24]. Stability Index Mapping was performed in parallel to the statistical model as mean of cross validation. This approach uses a GIS-based methodology implementing the slope stability model coupled with a hydrologic model for rainfall induced landslide susceptibility assessment. Infinite slope stability model was implemented with the following input: recharge and transmissivity, soil density, angle of internal friction, and cohesion. To perform the comparison, the six standard SINMAP predicted states are reduced to four susceptibility classes based on quantiles (similar to the logistic regression approach) [25] [26]; low, moderate, high and very high.

III. RESULTS AND DISCUSSIONS

3.1 *Parameter Maps*

3.1.1 *Morphometric Preparatory Factors*

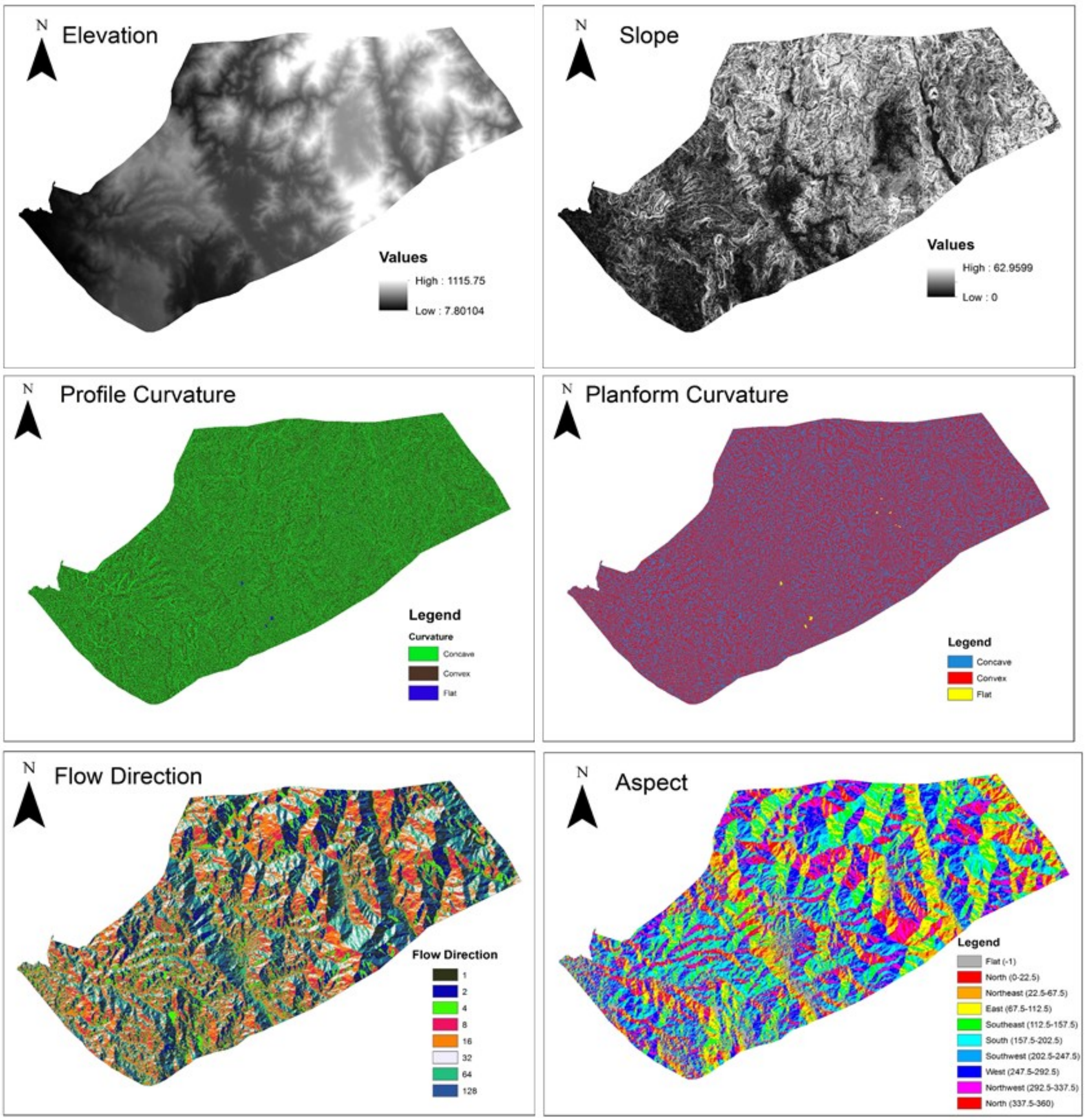

Figure 2. Thematic factor maps of morphometric variables related to landslide occurrence: elevation, slope, profile and planform curvature, flow direction and aspect raster maps of Antipolo City.

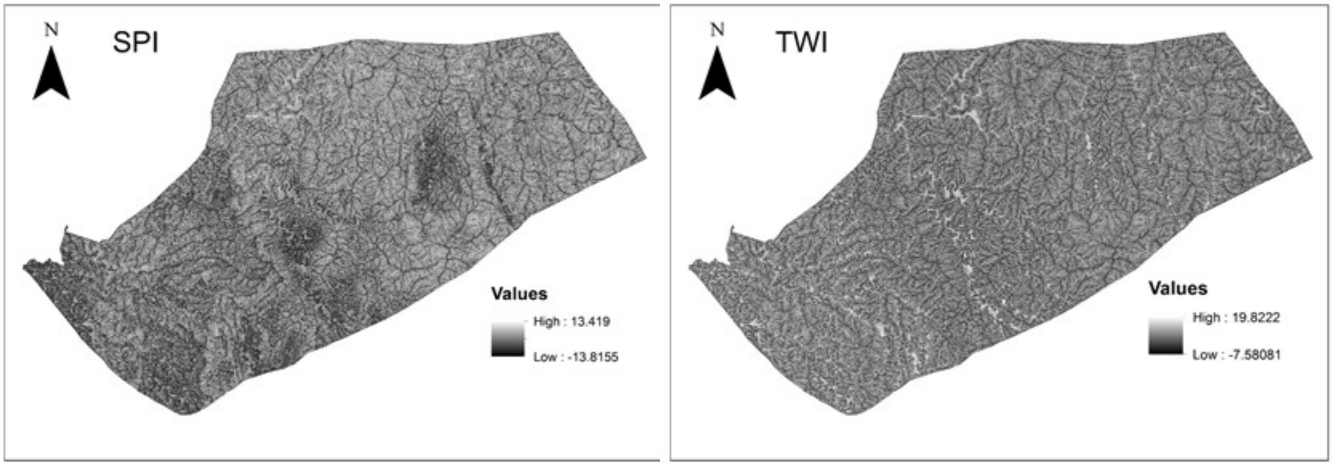

Figure 2 (continued). Thematic factor maps of morphometric variables related to landslide occurrence: SPI, and TWI raster maps of Antipolo City.

3.1.2 *Non- morphometric Preparatory Factors*

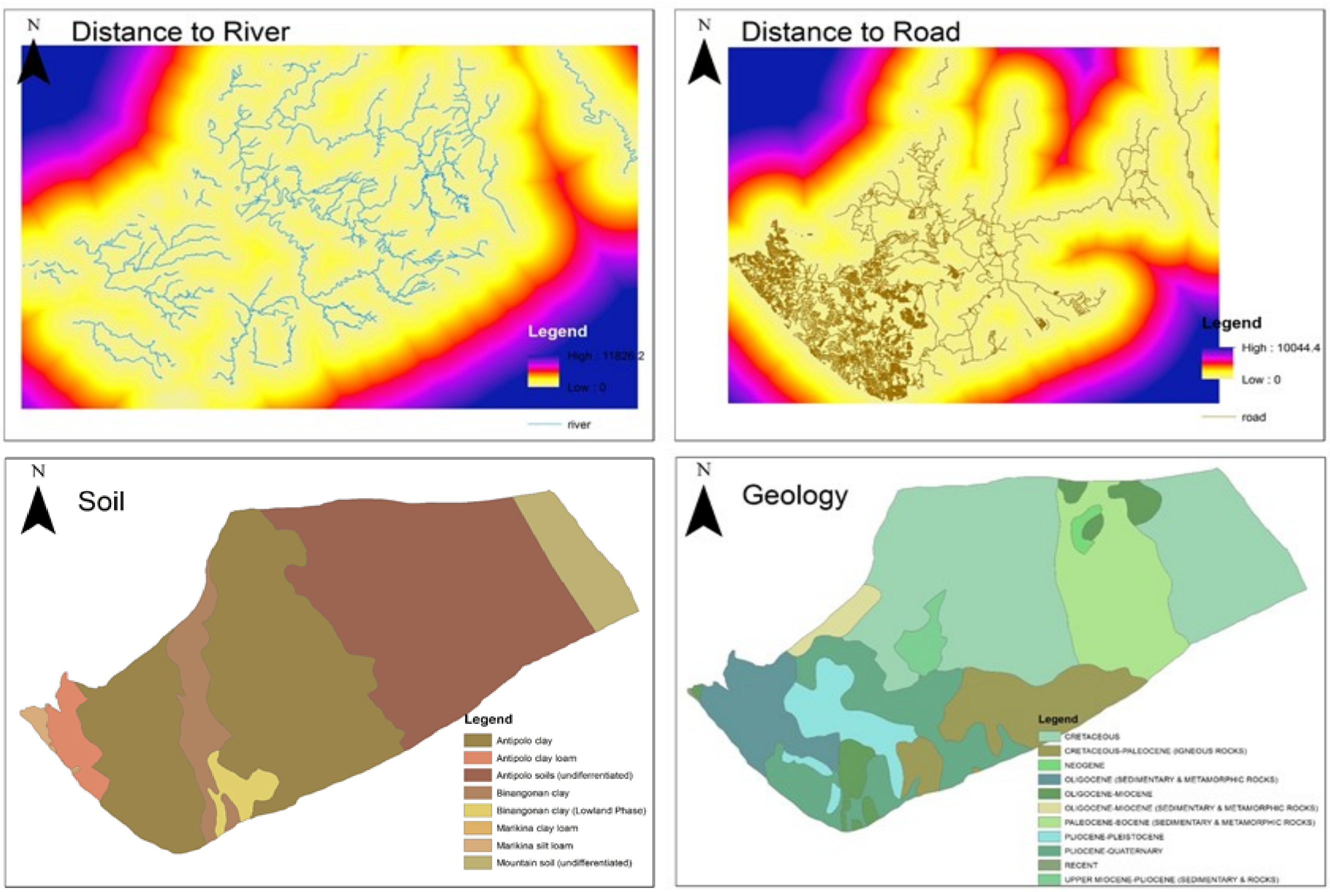

Figure 3. Non-morphometric factors considered in this study: Distance to river, Distance to road, Soil and Geology raster maps.

3.3 *Statistical Analysis – Logistic Regression*

The estimated regression coefficients for Models 1 to 5 are summarized in Table 1. The results of multi collinearity tests show that there is strong interdependence among slope, stream power index, and topographic wetness index, as exhibited by tolerance values of less than 0.2 (VIF > 5). The summary shows only factors that were found to be effective in determining the landslide occurrence. A positive coefficient indicates a positive relationship between the probability of landslide and the factor involved. Factors that were identified to be ineffective for the prediction were those with significance values greater than 0.05. A coefficient of zero has a transformed log value of 1. It means that this coefficient does not affect the odds of the event. It can be seen that that slope and erosion potential is the strongest contributor in landslide occurrence.

Table 1. Summary of Regression Coefficients for Models 1 to 5

	1:1 (LS:NLS)				
Paremeters/ Coefficients	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
SPI	0.025	0.045	0.000	0.034	0.000
Flat	-2.854	-2.591	-4.457	-2.890	-2.361
North	-1.032	-0.803	-0.782	-0.789	-0.736
Northeast	-0.991	-0.972	-0.903	-0.831	-0.890
East	-0.700	-0.580	-0.520	-0.610	-0.490
Southeast	-0.272	0.000	0.000	0.000	0.000
South	-0.281	-0.281	0.000	-0.263	0.000
West	-0.860	-7.330	-0.686	-0.790	-0.671
Northwest	-1.910	-1.950	-1.713	-1.885	-1.668
Oligocene-miocene	-2.393	-2.200	-2.317	-2.518	-2.226
Paleocene-eocene	-2.268	-2.272	-2.086	-2.235	-2.302
Oligocene	-2.680	-3.103	-2.783	-2.621	-2.524
Neogene	-2.999	-2.766	-2.923	-2.362	-2.790
Upper miocene-pliocene	-0.702	0.000	-0.803	-0.964	-0.713
Pliocene-quaternary	0.590	0.671	0.596	0.650	0.640
Pliocene-pleistocene	-1.976	-2.121	-2.034	-1.771	-1.949
Cretaceous-paleocene	0.705	0.782	0.681	0.823	0.806
Plan convex	0.169	0.000	0.213	0.000	0.252
Antipolo soil	-0.907	-0.923	-0.967	-1.115	-0.908
Antipolo clay	-1.556	-1.668	-1.551	-1.721	-1.593
Binangonan clay	-1.950	-2.141	-2.075	-2.225	-2.065
Binangonan clay (lowland)	-2.594	-3.699	-2.749	-3.522	-2.600
constant	-9.145	-9.026	-9.371	-9.076	-9.352

The success rates of all models are found to be greater than 90%, as shown in Table 2. Chance accuracy is still expected even if there is no relationship between landslide occurrence and the independent factors. The chance accuracy rate is computed as the sum of the squared percentage of landslide and non-landslide events. Since the input training data set contain equal landslide and non-landslide events, and imposing a 25% increase in chance accuracy rate, the resulting rate is equal to $1.25* (0.5^2 + 0.5^2) = 0.625$ or 62.5%. All model success rates are greater than 62.5%, which signifies that the models are not based on random fit.

Table 2 . Success Rate for Models 1 to 5

	Model 1	Model 2	Model 3	Model 4	Model 5
Success Rate	90.7%	90.8 %	90.7 %	91.1 %	90.7 %

The four cases considered in creating a confusion matrix are correctly predicted landslide occurrence, correctly predicted landslide presence, landslide event is predicted but not observed, and landslide absence is predicted but not observed. As an illustration, the confusion matrix corresponding to Model 1 is presented in Table 3.

Table 3. Confusion Matrix For Model 1

	PREDICTED		Percentage Correct
OBSERVED	**Absence**	**Presence**	
Absence	1515	205	88.08
Presence	82	1638	95.23
Overall Accuracy			**91.66%**

The results of analysis for overall accuracy, AUROC with standard error, and root-mean-square error (RMSE) is presented in Table 4. It can be seen that Model 1 generated the highest overall accuracy and AUROC with corresponding lowest RMSE, suggesting that this model is the best estimator among the five trained models.

Table 4. Summary of Model Performance Assessment for LS:NLS =1

	Model 1	Model 2	Model 3	Model 4	Model 5
AUROC	0.908	0.887	0.907	0.903	0.907
Standard Error	0.002	0.003	0.002	0.002	0.001
RMSE	0.2478	0.2937	0.2540	0.2650	0.2545
Overall Accuracy	91.66%	88.49%	90.20%	89.88%	90.29%

The results of final run for training of Model 6 (LS:NLS=2) to 8 (LS:NLS=4) are summarized in Table 5. I can be observed that as the landslide to non-landslide ratio increases, the overall accuracy decreases. This is also reflected by the decrease in AUROC from model 6 to 8, with a corresponding increase in RMSE. Hence, increasing the landslide to non-landslide ratio in the input training data also makes the generated model more unreliable.

Table 5. Summary of Model Performance Assessment LS:NLS =2,3 and 4

Based on the above performance assessment, Model 1 is implemented as the final model in generating a landslide susceptibility map for Antipolo City using logistic regression. The computed probabilities were classified into four susceptibility classes (low, moderate, high, very high) using by natural breaks.

	Model 6	**Model 7**	**Model 8**
AUROC	0.889	0.881	0.877
Standard Error	0.001	0.001	0.001
RMSE	0.3296	0.3278	0.4152
Overall Accuracy	85.18%	84.26%	81.26%

Based on the above performance assessment, Model 1 is implemented as the final model in generating a landslide susceptibility map for Antipolo City using logistic regression. The computed probabilities were classified into four susceptibility classes (low, moderate, high, very high) using by natural breaks.

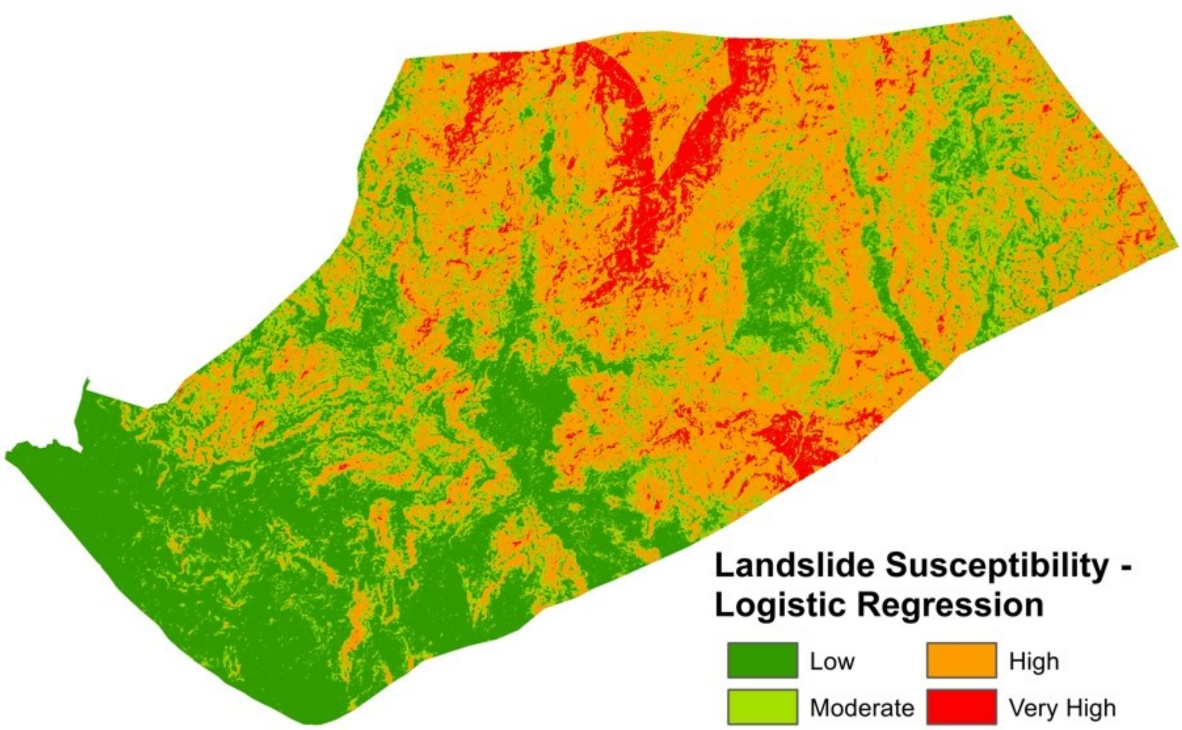

Figure 4. Final Landslide Susceptibility Map Using Statistical Approach

3.4 *Multi Method Comparison*

The distribution of the total area with respect to the four susceptibility classes for the two methods is presented in Table 6. It can be seen that logistic regression identified a larger area that is highly susceptible to landslide occurrence with 30.43%. However, considering the total high and very highly susceptible areas, there was no significant difference in the assessment of both methods; at 37.94% and 37.65% of the total area for the statistical and deterministic approach, respectively.

Table 6. Area Distribution Among the Susceptibility Classes

	Statistical Approach (Logistic Regression)		**Deterministic Approach (SINMAP)**	
Class	Total Cells	% of Total	Total Cells	% of Total
Low	1,900,847	53.89	1,465,905	41.56
Moderate	288,463	8.18	799,321	22.66
High	1,073,473	30.43	470,882	13.35
Very High	264,752	7.51	791,427	22.44

The complete landslide inventory map was overlaid to each susceptibility map and the classification of each landslide pixel with respect to each model was identified, as presented in Table 7. The statistical model had a better performance in classifying the landslides having less than 15% of the total landslides being assigned to area of low to moderate landslide susceptibility, while the deterministic model 20.76% for these classes.

Table 7. Landslide Distribution Among the Susceptibility Classes

	Statistical Approach (Logistic Regression)		Deterministic Approach (SINMAP)	
Class	**LS Pixels**	**% of Total**	**LS Pixels**	**% of Total**
Low	305	2.83	691	6.41
Moderate	901	8.36	1,547	14.35
High	5,640	52.31	5,123	47.52
Very High	3,935	36.50	3,420	31.72

IV. SUMMARY AND CONCLUSIONS

The logistic regression modeling implements a statistical approach based on maximum likelihood of landslide occurrence based on combinations of identified causative factors. K-fold cross validation approach required generating five landslide susceptibility models, which are based on input consisting of one to one relationship between landslide and non-landslide pixel. A positive regression coefficient indicates a positive relationship between the probability of landslide and the factor involved. Among the five models, designated Model 1 was chosen as the best predictor model, with corresponding success rate of 90.7%, 91.66% over-all accuracy, AUROC of 0.908 (standard error = 0.002), and lowest root mean square error (RMSE) of 0.2478. Sensitivity analysis was also performed using input data with LS:NLS values of 2, 3 and 4. Based on the analysis, increasing this ratio results to a decrease in accuracy resulting to lower AUROCs. There was also a significant increase in RMSE, making the models more unreliable. For a relatively large area of assessment, it was shown that statistical methods perform better than data-intensive slope stability modeling. This study provides a platform to perform a city-wide landslide hazard and risk assessment of Antipolo City. The results provide an opportunity to prioritize critical structures (i.e. schools, hospitals etc.) in site specific and detailed slope stability assessment.

References

[1] F. Guzzetti, "Landslide Hazard Assessment and Risk Evaluation: Limits and Prospectives," in 4th EGS Plinius Conference, Spain, 2002.
[2] M. Safaei, H. Omar, B. K. Huat, Z. B. M. Yousof and V. Ghiasi, "Deterministic Rainfall Induced Landslide Approaches, Advantage and Limitation," EJGE, pp. 1619-1650, 2011.
[3] ESRI 2011. ArcGIS Desktop: Release 10, Redlands CA: Environmental Systems Research Institute..
[4] R. A. Lohnes and R. Handy, "Slope Angles in Friable Loess," he Journal of Geology , pp. 247-258, 1968.
[5] F. C. Dai and C. F. Lee, "Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong," Geomorphology, vol. 42, no. 3-4, pp. 213-228, 2002.
[6] A. Buckley, ESRI, Mapping Center Lead, 2010. [Online]. Available: https://www.esri.com/arcgis-blog/products/product/imagery/understanding-curvature-rasters/.
[7] R. C. Sidle and H. Ochiai, Landslides: Processes, Prediction, and Land Use, American Geophysical Union, 2006.
[8] I. D. Moore, R. B. Grayson and A. R. Ladson, "Digital terrain modeling – a review of hydrological, geomorphological, and biological applications," Hydrol. Processes, vol. 5, pp. 3-30, 1991.
[9] S. Lee and K. Min, "Statistical analysis of landslide susceptibility at Yongin, Korea," Environmental Geology, vol. 40, no. 9, pp. 1095-1113, 2001.
[10] E. A. C. Abella and C. J. V. Westen, "Generation of a landslide risk index map for Cuba using spatial multi-criteria evaluation," Landslides, vol. 4, no. 4, pp. 311-325, 2007.
[11] S. a. Mossa, D. Capolongo, L. Pennetta and J. Wasowski, "A GIS-based assessment of landsliding in the Daunia Apennines, Southern Italy," in Mass movement hazard in various environments, Cracovia, Polonia, 86-91.
[12] L. Ayalew and H. Yamagishi, "The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan," Geomorphology, vol. 65, no. 1-2, pp. 15-31, 2005.
[13] Google, Google Earth Version 7.1, 2013.
[14] A. Nandi and A. Shakoor, "Application of logistic regression model for slope instability prediction in Cuyahoga River Watershed, Ohio, USA," Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards , vol. 2, no. 1, pp. 16-27, 2008.
[15] S. Lee, "Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data," International Journal of Remote Sensing, vol. 26, no. 7, pp. 1477-1491, 2005.
[16] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression, John Wiley & Sons, Inc, 2000.
[17] S. Menard, Applied Logistic Regression Analysis, SAGE University, 2001.
[18] I. Corp, IBM SPSS for Windows, Version 22.0, Armonk, NY: IBM Corp, 2013.
[19] H. Petschko, A. Brenning, B. R., G. J. and G. T., "Assessing the quality of landslide susceptibility maps – case study Lower Austria," Natural Hazards and Earth System Science, vol. 14, no. 1, pp. 95-118, 2014.
[20] M. Kuhn and K. Johnson, Applied Predictive Modeling, Springer, 2013.
[21] A. Bayaga, "Multinomial Logistic Regression: Usage and Application in Risk Analysis," Journal of Applied Quantitative Methods, vol. 5, no. 2, pp. 288-297, 2010.
[22] M. Zweig and G. Campbell, "Receiver-Operating Characteristic (ROC) Plots: A Fundamental Tool in Clinical Medicine," Clinical Chemistry, vol. 39, no. 4, pp. 561-577, 1993.
[23] T. Chai and R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," eoscientific Model Development, vol. 7, no. 3, pp. 1247-1250, 2014.
[24] C. V. Western, "Geo-Information tools for landslide risk assessment: an overview of recent developments," in 9th International Symposium on Landslides. Landslides: Evaluation and Stabilization, 2004.
[25] R. Pack and C. G. D. Tartabon, "Terrain stability mapping with SINMAP," Teretech Consulting, B.C. Canada.
[26] "pro.arcgis.com," ESRI, [Online]. Available: http://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/data-classification-methods.htm. [Accessed 2018].
[27] C. v. Westen, Application of geographic information systems to landslide hazard zonation, International Institute for Aerospace Survey and Earth Sciences, Enschede, 1993.