Research Article

Predicting the Susceptibility of Bamenda Escarpment Zone, NW Cameroon to Landslides Using Logistic Regression Model

Roland Ngwatung Afungang[1*], Roger Ngoufo[2], Clement AnguhNkwemoh[2]

[1]Department of Geography, University of Porto, Portugal

[2]Department of Geography, University of Yaoundé, Cameroon

***Corresponding author:** Roland Ngwatung Afungang, Department of Geography, University of Porto, Portugal. Tel: +351920561034: Email: afungang@gmail.com

Abstract

The Bamenda escarpment zone which is about 8.4 km2 is one of the most hazardous zones along the North-West stretch of the Cameroon Volcanic Line. Landslide events in this area cause harm to humans and economic disruptions almost on a yearly basis. The impact of this hazard is highly felt in Bamenda I, II and III municipalities especially those communities leaving at close proximity to the escarpment. The steep slopes dominated by volcanic rocks and highly weathered residual soils are easily pulled down by gravity during the rainy season. The objective of this study is to use a quantitative technique (logistic regression model) in assessing landslides susceptibility in the area. Landslides were identified using aerial photographs, satellite images and systematic field recognition. A total of 110 landslides mostly shallow translational landslides were registered during the landslide inventory process. This landslide dataset were randomly divided into two groups and the first group was used in building the predictive model and the second in its validation. Ten landslide conditioning factors were initially considered for the assessment and after being tested using the accountability and reliability indices, six out of the ten factors were retained for the susceptibility modeling. The slope was the most important geo-environmental factor influencing landslides and had a coefficient of 0.927. The model was validated using the success and prediction rates methods and the Area Under the Curve (AUC) was used to evaluate the performance of the model. The training model had a success rate of 81% and the validation model had a prediction rate 88%. The predictive power and accuracy of the model was evaluated using the error index of the ROC curve. The model had a True Positive rate of 0.554, a True Negative Rate of 0.879, an accuracy of 0.879 with a precision of 0.006. From this result, we concluded that landslide remains a serious threat in the area and can be predicted statistically.

Introduction

Landslide hazard is one of the most destructive geological hazards that cause damage on the physical environment [1]. As described by Varnes and the IAEG [2], a landslide hazard is the probability of occurrence of a landslide within a given area and at a specific time. Varnes [3] defines landslide itself as "A downward and outward movement of slope forming materials under the influence of gravity". From these definitions, the "Where" and "When" factor of landslide occurrence forms the basis for landslide hazard assessment [4]. The prediction of landslides events has been a subject of discussion for geomorphologists and land-use planners for many decades [5,6] and remains an active subject stimulating scientific discourse today. [1,7] in their research on hazard publication in international journals noted that there is significant growth of landslide hazard related articles in different parts of the world especially those induced by man activities. These publications shows that slope movements can be modeled through a number of approaches. These includes the statistical approach [6,8-11], the physical or process-based approach [12-17] and a combination of the two approaches [16,18,19]. Whether it is statistically or physically based, each method has its own weaknesses [11,15,18,19]. Due to some limitations such as the lack of proper knowledge about the mechanical properties of rocks, the mechanisms physical laws controlling slope [20] and the difficulties involved in applying physical and deterministic models on very large scales [17,21], statistically-based models have most often preferred. In the past years, studies on landslide hazard had focused on relating

landslide occurrence to several independent variables using statistical techniques [5,19,22-25]. This easy-to-use method has led to an increase in the publication of susceptibility assessment articles as noted by Gokceoglu and Sezer [7].

The Bamenda escarpment just as other landslide hazard affected areas in the world has seen an uptick in landslide events in recent years with severe impact on landscape, humans and biodiversity [26]. The impact of human action on steep slopes plays an important role in weakening the regolith [27] and is thought to be the main driving force behind this increase in slope failures. Landslide occurrences in the Bamenda area have most often been assessed qualitatively [28,29] with highly subjective results since judgment is based solely on the experience of the geomorphologist. The objective of this study is to use the logistic regression model [30] which is a database and multivariate quantitative technique to assess the susceptibility of the area to landslides and to evaluate the impact on land-use on slope failure. The study assumes that past behaviour trends of natural systems may be extrapolated into the future causing the same effects [10,20] meaning that past failures is key in predicting future failures.
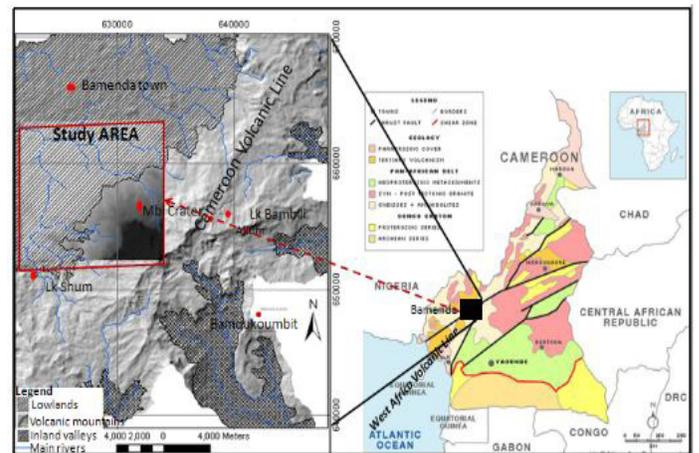
Study area

The Bamenda escarpment is part of a geomorphologic system in the region called the West-Cameroon Highlands and lies along the Cameroon Volcanic line [31-33] which is an extension of the West-African Volcanic line. The study area (Figure 1) covers an approximate area of 8.4 km2 and includes the escarpment proper and its periphery. Although landslides events have largely been recognized along very steep escarpment slopes, the gentle slopes around the escarpment also had a number of landslides. The escarpment fault divides the study area into two main relief units with the volcanic mountain to the south and the lowland (plain) to the north (Figure 2). It is a highland area with very high hypsometric difference. For instance, the peak altitude at the volcanic mountain is 2670m and the lowest altitude point in the plain is 1034m (ALS).

The general geological structure includes; basanite, hawaiite, mugearite, benmoreite, trachyte, rhyolites and ignimbrites [33] and the escarpment itself is characterized by an alignment of volcanic massifs and orogenic plutonic complexes. It is composed of trachyte; rhyolite and ignimbrite are covered by lateried basalt and felsic lavas which form the materials of most debris flows. The lowland is mainly composed of plutonic granite and gneiss rocks with patches of ignimbrite and phonolite which are rigid formations. Altitude is an important causal factor of landslide here as weathering is very high at the higher altitudes and decreases with elevation [22,27]. Rock type sequence is an important factor influencing slope movements [23,34] because it influences the slope stability [35] and weathering processes.

About 66% of the study area has rectilinear slopes where most slope failures have occurred in the past and concave slopes make up 28%. Convex slopes make up about 16%of the study area and are mostly known for rock falls. For instance, the Mbi crater (volcano) is a large anticlinal (quasi perpendicular) structures found in the area and famous for its spectacular rock falls. Slope curvature as important landslide conditioning factor [22,34] is also an important conditioning factor in the area.

Figure 1: Location of the study area.

Slope steepness is widely accepted as one of the most important determining factors of slope instability [5,17,34,36]. The steepness of the slope also determines the accumulation and transport of unconsolidated materials down the slope [27]. The steep slopes have an outcropping geology with bear soils and the moderate slopes are covered by scree or talus deposits from fragmented rocks that originate from the cliff face and deep weathered mantle material. Most devastating landslides on the volcanic mountains have been recorded in areas where thick clayey materials of several meters lay above. Very fine-grain debris developed from rhyolite or trachytic parent materials are capable of trapping large amounts of water in their structures during the rainy season. Although clays are almost impermeable, the presence of this daily recycling material beneath clay layers makes it easy for them to slide. Colluvium dominates the low slopes composed predominantly of fine grain particles [31] and these unconsolidated particles are mostly plastic in nature capable of holding much water in its upper layers. Colluvium is known to provide favorable conditions for shallow landslides [13]. The terraces on the slopes, fluvial terraces and some valleys are dominated by sandy soils and alluvium deposits with depths between 1-3 meters deep [37]. These materials allow water to go deep into their profile and are believed to be responsible for some of the deep-seated landslides that had occurred in the area. Geotechnical analysis show that the soil has a low bulk density of 1.32-1.59, low specific density of 2.20-2.58, high porosity of 47.92-64.28%, a water content exceeding 35.2% coupled with low cohesion ranging between 2.60-7.20 Kilopascal (kPa) with an an-

gle of internal friction ranging between 25.5° to 28° [31,33] which makes the soil highly exposed to landslides.

The climate of the study area is generally cool and dry compared to the rest of the country with temperatures ranging between 13 and 22°C [26]. The region receives about 2500mm of rainfall annually [4,26,28] and rain falls last from March to October with July and August being the rainiest months. This sub-tropical forest biome climate facilitates weathering [38] while the rugged relief and the dense dendritic drainage system accelerate the transportation of sediments down the slope. The increase in population and land-use along the escarpment and its peripheries has a negative impact on the land cover and the ecology of the area. For instance, the small upland watersheds seen in the 1930s have degraded from moist, evergreen, sub-mountain forest to farmland and rough pastures [26,28] to almost bare soil.

Material and Methods

The landslide inventory used in the study was built from the interpretation of aerial photographs, satellite images and direct ground survey carried out in 2009, 2012 and 2013. Aerial photo interpretation and field investigations are traditional methods largely used in the identification of geomorphologic features [39] and information gathering on natural system [40]. The aerial photos that were used were taken in 1965 and the satellite images used were taken from 2001 to 2011. The aerial photos show landforms produced by landslide events on and before 1965 and the satellite show some landforms produced before 2001 and those from 2001 to 2011.Landslides that were visible on the satellite images and absent on the air photos were believed to have occurred between 1965 and 2001. The total number of landslides identified from satellite and aerial photos were 69. Other 41 landslides were recognised during field work giving a total of 110 landslides. The largest landslides had a size of 5,565m2 and the smallest measures 912m2 with the average slides estimated at 287m2. Only the landslide scarp was considered in the model leaving out the run-out area even though the entire landslides were mapped. Landslide types found in the area included; debris flows, rock slide, debris slides, debris flows and earth flows. Landslides used in the modeling were defined according to movement type and shallow translational landslides were those considered in the study. For modeling purposes, the landslide dataset was randomly divided into two groups so that the former can be used in training the model and the latter in the validation. There are three main types of landslide division used in the validation of susceptibility model. This include: the spatial, temporal and random division [10,24,41,5] although this divisions have their weaknesses [11,42]. The spatial criterion which classifies slope failure with variation in space could not be used because some landslides sites were inaccessible and the temporal criterion could not because some landslide dates were unknown. Thus, the random criterion was the only option that could be applied in separating the landslide dataset. The first group was named the train-

ing group and the second was referred to as the validation group. Each of these groups contained 55 landslides scarps each. The first group containing 896 grid cells and the second had 858 grid cells with each grid cell measuring 10m2.

The logistic regression model was used in the modeling process. This model uses a dichotomous dependent variable (Y) and this requires two groups of variables to be represented. Since the independent variable (Y) is dichotomous, the first value (Y=1) represented areas with landslides (positives), and the other value (Y=0) represented areas with no slides (negatives). This was achieved by randomly generating 110 "Pseudo" points called non-points within the study area with a 400m buffer around landslide scars using the buffer creation tool in ArcGIS 10.2. This buffer distance was based on the minimum distance that was recognized between landslides in the field. Randomly extracted point that fell within the buffer perimeter were immediately eliminated. Areas without slope failure were represented by 220 points containing 1792 grid cells which is double the number of the registered landslides. Although most published papers adopt a balance dataset with equal number of positives and randomly extracted negatives, we double the number of negative because we hold the opinion that logistic regression works well when the negatives are widely spread and fully represent the study area [43,44] and since landslide area is usually very small compared to the study area [45]. However, some authors [27,46] due argue that the doubling of grid cells for areas without landslides might cause an over-estimation in the model.
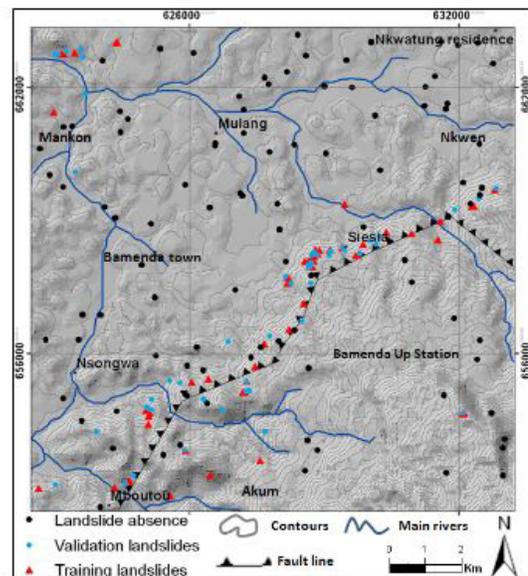

Figure 2: Landslide source area on shaded relief.

Selection of conditioning factors used in the study

To derive detail information on predictive variables to be used in the modeling process, each factor was divided into sub-

classes (Table 1) depending on their physical characteristics and standard deviation. As earlier noted, seven landslide conditioning factor were selected (Figure 2) and those that were in vector form were transformed into raster maps with a unique cell size of $10m^2$. Each of these raster maps were then overlay with the landslide inventory map from which negatives representing non-landslides pixels were extracted.

Landslide conditioning factors (geo-environmental factors) used in the assessment came from three sources. The first group of factors were derived from the Digital Elevation Model (DEM) and included: slope gradient (steepness), slope curvature (plan curvature), slope aspect, contributing upstream area and elevation). The second group was derived from a 1:5.000 scale map constructed using satellite images and included: land-use, and geomorphology, distance to roads and streams. The third group of factors including lithology and distance to faults was constructed from a 1:50000 scale geological factor map.

It's largely accepted that the choice of landslide conditioning factors can improve the result of the prediction model and thatthe selection of minimum conditioning factors have sometimes been difficult for geomorphologists [11,20,24]. This is because not all factors play a rolein the occurrence of each landslide or landslide type [47]. Two basic weights estimation methods and the Chi square 2x2 contingency table were used to select the best variables influencing landslides in the area. This includes the "Accountability and Reliability" indices [45,48] and Chi square test of contingency. The accountability indices estimates the total number of landslide pixels in sub-classes of the factor map whose average landslide density is greater than the average density in the factor map divided by the sum of landslide pixels in the entire area and multiplied by 100. It is mathematically expressed as;

$$Accountability = \frac{\Sigma (Si>AvDA)}{\Sigma (SiDA)} \times 100 \qquad (1)$$

Si>AvDAis the area of sub-classes in factor map with landslide density is equal or greater than the general average of the factor map; SiDA is total area with landslides in the factor map. The Reliability index estimates the sum of landslide cells of those subclasses with landslide densities greater than the average density of the factor map (study area), divided by the total area of those classes and multiplied by 100. This was expressed as;

$$Reliability = \frac{\Sigma (Si>AvDA)}{\Sigma (NSi)} \times 100 \qquad (2)$$

NSi = total area of classes with landslide density greater than the average density in the map. The Chi square test on contingency table (Eq. 6), was also used to select the factors to be used in the final model. This statistical significance test measured the "goodness of fit" of observed and expected values. The maximum likelihood function was used to estimate the parameters most likely to cause an event (observed data) with the odds reflecting the probability of landslide occurring, to the probability that it will not occur. The odds ratio using "Chi Square" was computed using the following expression;

$$X^2 = \sum \frac{(o-e)^2}{e} X^2 = \sum \frac{(o-e)^2}{e} \qquad 3a$$

Where "o" represents observed frequencies and "e" represents expected frequencies

$$G = 2\sum f. \ln(\frac{f}{f_i}) G = 2\sum f. \ln(\frac{f}{f_i}) \qquad 3b$$

Where "G" represents the Likelihood Ratio statistic, f represents observed values, f_i represents expected values, and "ln" indicates the log to be taken. Variables whose computed p-value were lower than the significance level alpha=0.05 meant a rejection of the null hypothesis H0, and acceptance of the alternative hypothesis Ha. Those variables whose rows and columns were independent (H0), were accepted.
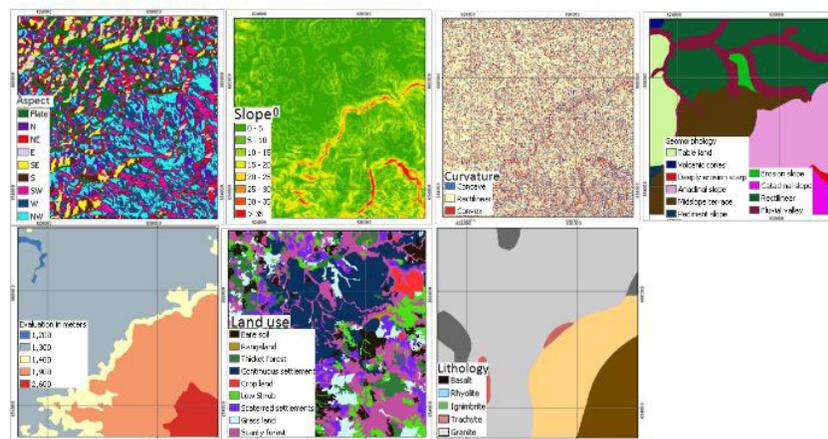

Figure 3: Selected landslide conditioning factors used in the study.

The results of the two variables selection methods had some differences. For instance, the accountability and reliability index showed that the lithology and fault factors weren't significant landslide conditioning factors. Distance to faults and curvature were to be rejected in Chi square test. Amongst the eleven variables that were initially considered, seven of them including: slope gradient, aspect, curvature, elevation, geomorphology, land-use and litho logy were selected (Figure 3) to be used as independent prediction variables in the logistic regression model.

Computing Process

The logistic regression model also called the logit model was used to analyse the relationship between multiple geo-environmental factors (X_1, X_2, X_n) and landslides dataset (y). It is a multivariate regression between a dependent variable and several independent variables [44]. This modelhas proven to be reliable in analysing the relationship between multiple independent variables and a dependent variable that are dichotomous [27]. The presence or absence of landslides denoted as Y=1 and Y=0 are represented by the binary variable Y and n is the number of covariates expressed as X_1, X_2, X_n. The probability of landslide occurrence expressed as Y=1 is a function of X_1, X_2, X_n using the samples of m (X_{11}, X_{12}, $X_{13...}$, X_{1n}; Y_1),…., (X_{m1}, X_{m2}, $X_{m3...}$, X_{mn}; Y_m). In the Logistic regression model, P(Y=1) is the expected value of Y given the n independent variable (X_1, X_2, X_n) as Y is the indicator variable. To avoid that the P(Y=1) should not lie between zero and 1, we use the odds that Y=1 instead of the probability that Y=1. This is expressed as:

$$Odds(y = 1) = \frac{P(Y=1)}{1-P(Y=1)} \qquad (4)$$

Figure 4: Illustration of the Logistic Regression procedure used in the study.

To avoid that the estimated probability should not exceed the maximum or minimum possible values of the probability, the natural logarithm of the odds that Y=1 was used as the dependent variable [27] and expressed as:

$$L(Y) = \alpha + \beta_1 X_1 + \cdots.. + \beta_n X_n \qquad (5)$$

Landslide modeling through Logistic Regression Analysis (LRA) makes use of estimated coefficients, standard errors, and p-values and the model was applied following the principles laid down by Hosmer and Lemeshow [49].

Using the laws of exponent and log, (i.e., converting Logit (Y) to odds and odds to Y=1), the probability of landslides (P) to occur in the m sample is expressed as:

$$P = \frac{1}{1+ Exp \left[-(\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)\right]} \qquad (6)$$

Where, P (0, 1) = Probability of an area to be stable or unstable; X_1, X_2, X_n are the Independent variables; $\beta_1 \beta_1$, $\beta_2 \beta_2$,.... $\beta_n \beta_n$ Are the regression coefficients; $\alpha \alpha$ is the value of the intercept.

The model was applied using 50% of the inventory (training landslides) extracted randomly using the spatial analysis tool in ArcGIS as explained in material and methods section and applied following the scheme shown in (Figure 4).

Results

Evaluation of conditioning factors and Susceptibility mapping

The overlay of negatives or non-landslide points and landslides (positives) over each conditioning factors maps helped train the model in evaluating the relationship between each predictive variable and landslides. The contribution of each class of the predictive factor to landslides was measured by the derived weight values of those classes. This was determined by the number of positives found on each unit of the classified factor map. The higher the number of classes with high derived weight values, the more important is the conditioning capacity of that factor and the higher its influence on the susceptibility index map. The maximum likelihood convergence algorithm of the logistic regression (Eq. 3) was used to get the best fitting of the model. This was then used to infer the weight of each of the variables. Each class of the classified predictive variable map had a weighted value measured in coefficients that ranged from 0 to 1. The coefficient (weight values) of conditioning factors (Table1), range from -2.775-0.927 and expresses the relationship between stable and unstable areas of the map.
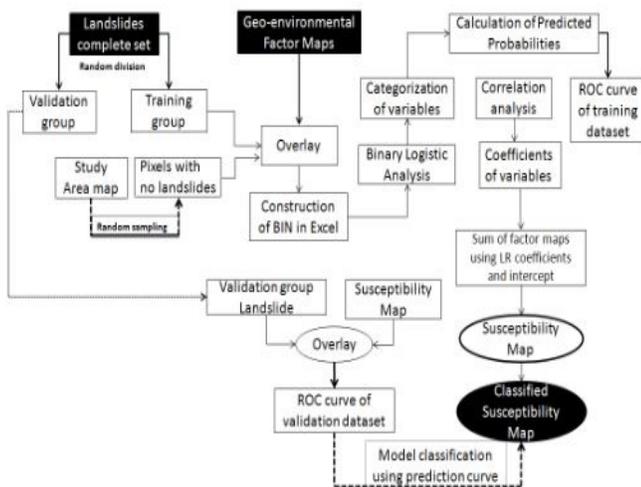
Data set	Sub-classes	Coefficient of factor map
Slope steepness	(1) <5; (2) 5-10; (3) 10-15; (4) 15-20; (5) 20-25; (6) 25-30; (7) 30-35; (8) >35	0.927
Slope elevation	(1) 1190-1200; (2) 1200-1300; (3) 1300-1400; (4) 1400-1900; (5) 1900-2600	-0.001
Slope aspect	(1) Plain; (2) N; (3) NE; (4) E; (5) SE; (6) S (7) SW; (8) W; (9) NW	0.005
Slope curvature	(1) Convex slopes; (2) Rectilinear slopes; (3) Convex slopes	-0.440
Geomorphology	(1) Anaclinal slope; (2) cataclinal slope; (3) Deeply eroded scarp; (4) Mid-slope terrace; (5) Rectilinear slope; (6) Fluvial valley; (7) erosion zone; (8) table land; (9) pediment slope; (10) Volcanic cones	-0.037
Lithology	(1) Basalt; (2) Rhyolite; (3) Ignimbrite; (4) Trachyte; (5) Pan African Granite	0.124
Lands use	(1) Bare soil; (2) scattered settlements (3) Low shrub; (4) Grass land; (5) Crop land; (6) Rangeland; (7) Scanty forest; (8) Continuous settlement (9) Thicket woodland	0.001
	Logistic regression	-2.775

Table 1: Conditioning factors used in the study and their coefficients.

Results from the modeling scheme (figure 2), showed that the slope, lithology, aspect land-use are the most important causal factors of landslides. As earlier noted by Dietrich and Montgomery, [36] the slope is the most important factor determining factor of slope failure and in this case, it has a coefficient of 0.927. Within the slope factor, some classes had higher coefficients which reflected the number of positives in the class. For instance, the classified slope map shows that the 15o-20o slope class has 37% of landslides and the 25o-30o slope class has 16% of landslides. The high coefficient of these classes shows their significant in the model. Within the aspect factor, the North-West slope with very few landslides has a low coefficient like the free fall face of the cliffs which are known for rock fall [35]. The bare rock formation has a very thin layer of coarse rock fragmentsor debris which is resistant with little soil cover which do not constitute enough material for a slide. Aspect had a positive weighted coefficient of 0.005 meaning that it is influential in slope instability. The geological factor has a significant coefficient of 0.124 which shows that petrogenetic processes are an important determining factor of slope movement. Large parts of the escarpment slope where most of the slides were registered are covered by trachyte and rhyolite formed in the quaternary and underneath these rocks are older and more resistant volcanic rocks. The porous nature of weathered materials developed from these rocks facilitate the infiltration of water into greater depths. This weakens the cohesion in rock material and making them unstable and more likely to move under gravity. The peripheral areas of the escarpment have moderate slopes composed of granites rocks with deep soil profiles. These slopes are less likely to fall since a large amount of water is required to trigger landslides and this is only possible when rainfall is prolonged and heavy. The coefficient of the Land-use factor (0.001) was not as impressive as expected given its multifunction role in causing slope instability in the area. Anthropogenic activities are quite intense around the peripheries of the escarpment whose slopes are very steep holding large amounts of unconsolidated material from weathering processes. The Land-use factor is not only a conditioning factor but is also a trigger of slope movements in the area. For instance, intensive agricultural activities along the escarpment wall has reduced the cohesion of the rhyolite exposing it to gravitational pull. The year 2009 saw the greatest number of landslides in the area for decades [4] and was linked to the undercutting of slopes and soil vibration during the construction of the Bamenda Up station-Down town new road. This shows the role of land-use as a trigger of slope failures. About half (46%) of the 110 registered landslides were on slopes affected to some degree by human activities. Land-use also has an impact on other conditioning factors like the slope, curvature, geomorphology and soil cover and its influence is reflected in these factors.

The slope elevation, geomorphology and curvature factor had less beta values with -0.001, -0.037 and -0.440 coefficients respectively (Table 1). The differences in high and low slopes didn't have any real impact on landslides. Changes in the morphology and slope instability were more reflected in the land-use since most of the area is undergoing rapid urbanization. Slope failure was more associated to land-use because landslides were registered on slopes (concave, convex, and rectilinear) irrespective of the characteristics.

The final susceptibility of the area to landslides was developed by multiplying the predictive variable raster maps (Figure 3) by the weighted coefficient of their beta values (Table 1) and combining these maps to get the susceptibility index map (Figure 5). This involved the integrated of the weighted maps in raster calculator and summing them (Eq.7).

$$\text{Susceptibility} = (Slp_{wc} + Elev_{wc} + Asp_{wc} @ + Cvt\_wc + Geopho\_wc @ + Lith\_wc + Land\_wc) \quad (7)$$

Where Slp is slope, Elev is elevation, Asp is aspect, Cvt is curvature, Lith is lithology, Land is land-use and Wc is weighted coefficient. The combination of the spatial probability layers of

conditioning variables (Eq. 7) gave the susceptibility index map (Figure 5a).
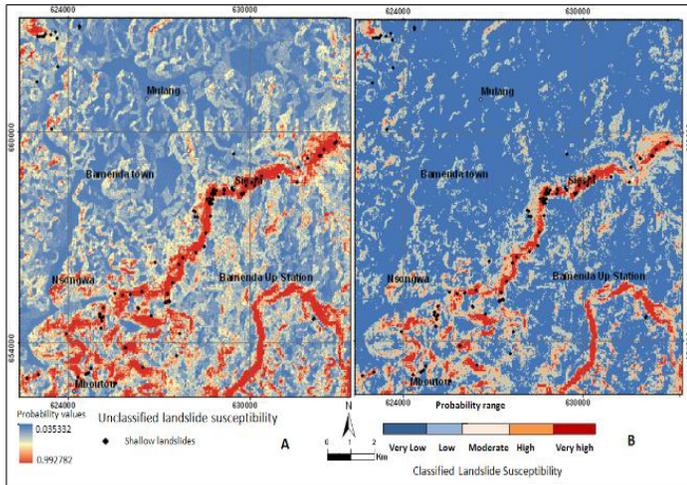
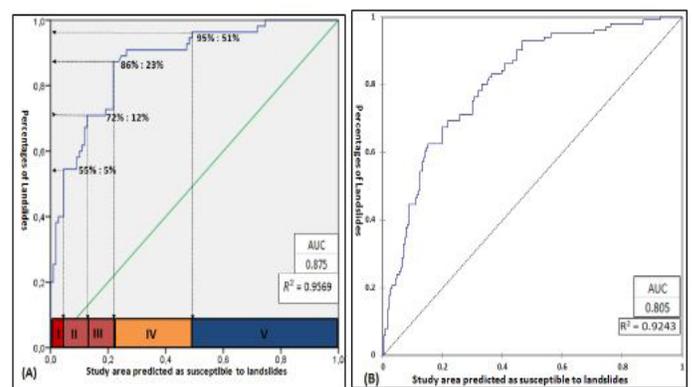**Figure 5:** Landslide Susceptibility maps of Bamenda escarpment zone.

The susceptibility map shows the probability of landslides ranging from 0.035332 to 0.99278. Slopes with many landslides (unstable zones) were classified by the model and separated from those with few or no slides (stable zones). To make the map better understood especially by non-specialists, the susceptibility map (Figure 5a) was reclassified using predicted break values of the ROC curve (Figure 5b). The closer the susceptibility value nears zero the lesser the probability of landslides and the closer it moves to towards 1, the higher the probability of having a landslide event. The Bamenda escarpment line and the Mbatu Mountain that are frequently hit by landslides were shown to be highly susceptibility to future slope failure.

Validation of the susceptibility model

Many authors hold the opinion that the validation of susceptibility models is important in evaluating the modelling process and its ability in forecasting future landslides [20,41,42]. Tien Bui et al. [43] sited Chung and Fabbri [41] as even saying that, when prediction models are not validated, they become useless and have no scientific significance. Thus, the model was validated to evaluate it's performed in classifying stable and non-stable areas. The validation was done through the success and prediction rates method [41,43,50]. To estimate the success and prediction rate, the susceptibility index map was classified. Some common used classification methods include: equal Interval method, quantile method, the natural break method and the standard deviation method. Although there is no standard method, some authors prefer the equal area classification [43,51,52]. There is also no standard number of classes but some authors [12,50,43] have used 3-5 classes. The optimum cut-off value was taken based on the procedure described by Chung and Fabbri [41] and Frattini et al. [53], where the ROC

curve "Turns over" to the right (Figure 6b). At this point, the cut point is lowered by a large amount to catch only a few more landslide pixels. Five susceptibility levels were adopted to visualize the Landslide susceptibility index map.

To construct the success rate curve of the training model, the computerized probabilities of the predictive factor maps using the training landslide dataset were combined with the binary values of the training model using the Receiver Operation Curve (ROC) creator algorithm of the SPSS software. From here the mapping units (landslides) and terrain units' correctly classified (Figure 5b) were estimated and the results reflected in the ROC curve. The horizontal axis (sensitivity or true positive) of the ROC curve is the estimated probability of terrain units or proportion of landslides rightly classified while the horizontal axis (1-specificity or false negative) expresses the probability of the area with no landslides correctly classified. The plotting of the accumulative frequency of landslides and accumulated area gave the spatial probability of landslide occurrence in the area. The success rate measures the goodness of fit for the landslide model to the training data. However, it is not suitable to use the success rate in measuring the real efficiency of the training model since it was built using the training group of landslides [43,50,52]. If done, it will turn to overestimate the capacity of the model. Because of this, the validation of the model was done using the prediction rate constructed from the validation group of landslides. The prediction ROC curve was obtained by overlapping the validation dataset of landslides (896 grid cells) with the susceptibility index map. The ROC creator algorithm of the SPSS software was again used to construct the prediction curve. Contrarily, to the success rate, the prediction rate explains the success rate [54] and how well the conditioning factors predict future landslides.

Figure 6: ROC curve for the logistic regression (a) validation model (b) training model.

By overlaying the two groups of landslides with the landslide susceptibility index, the cumulative percentages of existing landslide against landslide susceptibility index values were calculated.

The prediction rate curve (Figure 6a) has a better classification of susceptibility area more than the success rate curve (Figure 6b). The prediction rate curve was used to interpret the susceptibility map. Based on the predicted probabilitiesof the prediction ROC curve, 55%, 17%, 14%, 9% and 5% of slope failureare predicted to occur on 5%, 7%, 11%, 28% and49% of the study area respectively. The frequency of landslidesgenerally decreases with increase in the susceptibility area as landslides are mostly concentrated on particular zones. This means that the predictive value of the ROC curve varies with variation in susceptibility area. The prediction ROC curve indicates thatabout 24% of terrain units has an estimated probability of 0.8 (Figure 6a) making this zone highly susceptible to landslides. The very high, high, moderate, low and very low susceptibility classes (class I, II, III, IV and V) had predictive values of; 0,55; 0,17; 0.14, 0,09 and 0,50 respectively (Table 2). The predictive values show the probability of each class to be affected by future landslides. The very high and high usuceptibility classes are the most dangerous zones. But for civil protection purposes, the moderate class should also be considered as a high potential landslide zone because many devastating events took place in this class.

Susceptibility class		Class Area (m²)	Susceptibility zone	Prediction values of susceptibility class
I	Top 5%	6716085	Very high	0.55
II	5 – 12%	9402519	High	0.17
III	12 – 23%	14775387	Moderate	0.14
IV	23 – 51%	37610076	Low	0.09
V	51 – 100%	65817633	Very low	0.05

Table 2: Characteristics of susceptibilit classes of the classified model.

Interpretation of results and estimation of model's performance

There are many parameters used in interpreting the results of the logistic regression model. Three of them include the "Goodness of Fit", accuracy of the model denoted by R2 and the Area Under the Curve (AUC). The goodness of fit was used by comparing the likelihood L0 for the null model and the likelihood L0 of the fitted model where the slope parameters are zero (0). The number of the independent variables is equal to the degree of freedom of the R2 value. The higher the p-value associated with the R2, the higher the significance of the regression parameters and the better the fit to the data than the null model [27]. The R2 statistic was calculated as:

$$R^2 = -2\{\log(L0) - \log(L1)\} \tag{7}$$

The success curve has a regression fitting function of R2=0.9243 and the prediction curve shows R2=0.9569. The Area

Under the Curve (AUC) is also considered as a good qualitatively measure to assess the accuracy of susceptibility models [41,43,52]. It represents the probability that the landslide susceptibility value for a landslide mapping unit (Y=1) estimated by the model will exceed the result for a randomly chosen no-landslide points (Y=0). The AUC varies from zero (0) to one (1), and the closer the AUC value towards 1, the higher the prediction accuracy of the model. The AUC for the training model is 0.805 while that for the validation model is 0.875. These functions indicate how well the model has classified terrain units with no slides against mapping units with landslides. The higher AUC rate of the validation curve shows that it is more accurate to predict landslides than the success curve.

Notwithstanding the good results of the validation process, the performance of the model needs to be estimated in order to improve its predictive power. This was done using the error index of the ROC curve. As such, different metrics including the True Positive Rate (TPR), True Negative Rate (TNR), the accuracy and the precision rate [12,42] of the model were computed (Eq. 8). These rates were calculated following the formulas below:

$$TPR = TP/(TP+FN)$$

$$TNR \text{ or sensitivity} = TN/(FP+TN)$$

$$FPR = FP/(FP+TN) \tag{8}$$

$$FNR = FN/(FN+TP)$$

$$ACC = (TP+TN)/(TP+FN+FP+TN)$$

$$PPV = TP/(TP+FP)$$

Where: TP is True Positive; TPR is True positive rate; TN is True Negative; TNR is True Negative Rate: FP is False Positive; FPR is False Positive Rate; FN is False Negative; FNR is False Negative Rate; PPV is Precision; ACC is Accuracy.

The true positive and true negative was considered as correct assignments while FP and FN are model errors. Based on the prediction ROC curve, all grid cells with landslide probability less than 0.8were considered as probable stable areas while those above 0.8 were considered unstable (i.e., potential landslide) cells. To optimize and evaluate the efficiency of the modeling process, the unstable (40654 pixels) and stable zone (295238 pixels) were compared with the landslide inventory map. The observed landslide truly identified by the model (hit area) had 227 pixels thus giving a True Positive Rate (TPR) of 0.55446 and the unstable areas defined by the model to be unstable had 295058 grid cells giving a True Negative Rate (TNR) of 0.87949. Model errors including FP (40430 pixels) and FN (180 pixels) had rates of 0.12051 and 0.44554 respectively (Table 3). The accuracy rate of the model stood at 0.879 and the precision was 0.006 calculated from the positive and negative values (Table 3).

Indicators	TP TPR		TN TNR		FP FPR		FN FNR		ACC	PPV
Prediction rates	224	0.56	295058	0.88	40430	0.12	180	0.45	0.88	**0.006**

Table 3: Estimator of model performance.

Discussion

Multivariate regression analysis of binary sequences is often applied in the study of slope failures. The accountability and reliability index were used to select the most important factors affecting landslides in the area. Among the tengeo-environmental variables that were considered, seven were retained based on their favorable relationship to landslides and included: slope, curvature, elevation, aspect, geomorphology, lithology and land use. Instable and stable areas were represented in terms of positives (landslide) and negatives (non-landslide) denoted as 1 and 0. The evaluation of each conditioning factor and their coefficients, (e.g. slope 0.927, aspect 0.005, land-use 0.001, Curvature -0.44, geomorphology -0.037, elevation -0.001, lithology 0.124), and a constant of -2.775), gave the relationship of each factor to slope movement. The correlation between the causal factors and the landslide inventory map showed that the slope is the most important landslide conditioning factor in the area just as in many areas of the world [17,36]. Through the binary function of the logistic regression using the laws of exponent (exp) and log (L), the probability (P) of the area to be affected by landslides was determined. This reduces subjectivity common with qualitative analyses that have been used to assess landslides in the area. The susceptibility index map analyses satisfy the spatial probability component of Varnes and the IAEG [2] definition of landslides which focused on "WHERE" landslides will occur. The maximum likelihood using the "Chi Square" 2X2 table was more reliable in examining stable and unstable areas and making relationships with the odds (that Y=1 instead of the probability that Y=1), giving the probability of the area to be affected or not affected in the future by landslides. About 55% of the 22400m2 area of landslides was found on only 4% of the susceptible area while the rest 52% of the area had only 5% landslides. This shows that landslides susceptibility is concentrated on a small area. The classification of the susceptibility index map into low, medium, high and very high susceptibility zones using the ROC curve graduations (Figure 5a), did not only improve the efficiency of the model but made the map easy to understand by non-specialist users such as environmental planners, policy makers and civil protection officers. This is in accordance with other authors findings such as Soeters and Van Westen [5], Zêzere et al. [10]. The quantitative validation of the model made the assessment more valuable scientifically. The estimation of the model's performance showed that the logistic regression model is good means of landslide forecast in the area. This process also enhances the quality of the model by improving our understanding of the landslide process.

However, the assessment had some short comings. The predictive variables (Figure 3) which form an integral part in landslide susceptibility assessment [2,25,34] were constructions at the scale of 1: 50000 (Figure 3) which is too large for the identification of slides <50m2.Although such a scale has been used for regional and sub regional landslide investigations, better results can be obtained using more detailed maps with scales of 1:5000, 1: 10000. Despite the poor nature of the basic information, the success rate (80%) and prediction rate (87%) was quite good and are similar to results(prediction rate of 89%) obtained using the Information Value method in the same area [37]. The slide differences in these rates may have been caused by differences in the form of landslides representation since landslides were represented as point in the logistic regression model and polygons in the Information value model.

Conclusion

The assessment has showed that the logistic regression model can be applied in data deficiency zones using few landslide conditioning factors with basic information on landslides. The physical characteristics of the Bamenda escarpment area and its peripheries also described by some authors [28,29,31,33] is a fertile ground for landslides. About 22400m2 study area have been affected by landslides and this result points to the fact that landslide remains a threat to live and property in the area. Based on recommendations on landslide conditioning factors [35] and guidelines on landslide susceptibility analyses [20], the background information used in the study area needs to be improved upon to achieve good susceptibility assessment since the results of susceptibility models highly depend on the quality of quality of the input data [10].The outcome of the work further strengthens the quantity and quality of data available for use by civil protection units and it lays the groundwork for landslide hazard and risk assessments in the area.

Highlights

► We used matrices to select seven factors out of ten considered to have a link with slope failures

► The contribution of these factors to landslides was analysed using logistic regression model

► We found out that the slope, lithology and land-use were the major causes of landslides

► The model classified the study area based on susceptibility levels with steep slopes being highest

► The model had a high prediction rate and can be used to forecast future landslides in the area

References

1. Pardeshi SD, Autade SE, Pardeshi SS (2013) Landslide hazard assessment: recent trends and Techniques. Springer Plus 2: 523

2. Varnes D, IAEG (1984) Landslide hazard zonation: a review of principles and practice. United Nations Scientific and Cultural Organization, Paris 1-6.

3. Varnes D (1978) Slope Movement Types and Processes. In: Schuster, R.L. and Krizek, R.J., Eds., Landslides, Analysis and Control, Transportation Research Board, Special Report No. 176, National Academy of Sciences, 11-33.

4. Afungang RN (2015) Spatiotemporal Probabilistic assessment of landslide hazard along the Bamenda Mountain region of the Cameroon Volcanic Line. University of Porto.

5. Soeters R, van Westen (1996) Slope instability recognition, analysis and zonation. In: Turner K, Schuster R (eds.), Landslides investigation and mitigation. National Academy Press, Washington 129-177.

6. Swets J (1988) Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

7. Gokceoglu C, Sezer E (2009) A statistical assessment on international landslide literature (1945–2008). Landslides 6: 345-351.

8. Epifanio B, Zêzere JL, Neves M (2013) Identification of hazardous zones combining cliff retreat with landslide susceptibility assessment. Journal of Coastal Research, Special Issue 65: 1681-1686.

9. Bateira C (2010) Evaluation of natural susceptibility in the north of Portugal. Analysis and territorial management. Prospectiva e Planeamento 17: 15-32.

10. Zêzere JL, Reis E, Garcia R, Oliveira S, Rodrigues ML, Vieira G, Ferreira AB (2004) Integration of Spatial andtemporal data for the definition of different landslide hazard scenarios in the area north of Lisbon (Portugal). Natural Hazards and Earth System Sciences 4: 133-146.

11. Guzzetti F, Carrara A, Cardinali M, 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.

12. Teixeira M, Bateira C, Marques F, Vieira B (2014) Physically based shallow translational landslide susceptibility analysis in Tibo catchment, NW of Portugal. Landslides. 12: 455-468.

13. Vieira B, Fernandes N, Filho O (2010) Shallow landslide prediction in the Serra do Mar, São Paulo, Brazil. Natural Hazards Earth System Science 10: 1829-1837.

14. Baum RL, Godt JW, Savage WZ (2010) Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. Journal of Geophysical Research, Earth Surface 115.

15. Crosta G, Frattini P (2003) Distributed modeling of shallow landslides triggered by intense rainfall, Natural Hazards Earth System Science 3: 81-93.

16. Terlien M (1998) The determination of statistical and deterministic hydrological landslide-triggering thresholds. Environmental Geology 35: 2-3.

17. Montgomery DR, Dietrich WE (1994) A physically based model for the Topographic control on shallow landsliding. Water Resource Research 30: 1153-1171.

18. Frattini P, Crosta G, Sosio R (2009) Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. Hydrology Process 23: 1444-1460.

19. Gorsevski P, Gessler P, Boll J, Elliot W, Foltz R (2006) Spatially and temporally distributed modeling of landslide susceptibility. Geomorphology 80: 178-198.

20. Fell R, Jordi C, Christophe B, Leonardo C, Eric L, et al. (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Engineering Geology 102: 99-111.

21. Feda J, Bohác J, Herle I (1995) Shear resistance of fissured Neogene clays, Eng Geol 39: 171-184.

22. Duman TY, Can T, Gokceoglu C, Nefeslioglu H (2005) Landslide susceptibility mapping of Cekmece area (Istanbul, Turkey) by conditional probability. Hydrology Earth System Science Discuss 2: 155-208.

23. Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probabilistic landslide hazard assessment at the basin scale. Geomorphology 72: 272–299.

24. Alleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bulletin of Engineering Geologyand the Environment 58: 21-44.

25. Carrara A, Guzzetti F, Cardinali M, Reichenbach P (1998) Current limitations in modeling landslide hazard, editedby Buccianti A, Nardi G, Potenza R, Proceedings of IAMG'98.195-203.

26. Ndenecho EN (2011) Local Livelihoods and Protected Area Management-Biodiversity Conservation Problems in Cameroon. Langaa RPCIG 218.

27. Dai FC, Lee FC (2003) A spatiotemporal probabilistic modeling of storm-induced shallow landslides usingaerial photographs and logistic regression. Earth Surface Processes and Landforms 28: 527-545.

28. Lambi CM (2004) Revisit of the Recurrent Landslides on the Bamenda escarpment. Journal of Applied Social Science 4: 4-14.

29. Eze B, Ndenencho E (2004) Geomorphic and anthropic factors influencing landslides in the Bamenda Highlands, N. W. Province, Cameroon. Journal of Applied Social Sciences 4: 15-26.

30. Cox D (1958) The regression analysis of binary sequences. Journal of the Royal Statistical Society 20: 215-242.

31. Guedjeo C, Kagou DA, Ngapgue F, Nkouathio D, Zangmo TG, et al. (2013) Natural hazards along the Bamenda escarpment and its environs: The case of landslide, rock fall and flood risks (Cameroon volcanic line, North-West Region). Global Advanced Research Journal of Geology and Mining Research 2: 15-26.

32. Déruelle B, Ngounouno I, Demaiffe D (2007) The "Cameroon Hot line" (CHL): a unique example of active alkaline intraplate structure in both oceanic and continental lithospheres. Compte Rendu Géoscience 339: 589-600.

33. Gountié DM, Njonfang E, Nono A, Kamgang P, Zangmo TG, et al. (2012) Dynamic and evolution of the Mounts Bamboutos and Bamenda calderas by study of ignimbritic deposits (West-Cameroon, Cameroon Line). Syllabus Review, Science Series 3: 11-23.

34. Bateira C, Soares L (1997) Mass movement in the north of Portugal: Factors of their occurrence. Territorium 4: 63-77.

35. Eberhardt E (2003) Rock slope stability analysis-Utilization of advanced numerical techniques. University of British Columbia, Vancouver.

36. Dietrich W, Montgomery D (1998) SHALSTAB: a digital terrain model for mapping shallow landslide potential. National Council of the Paper Industry for Air and Stream Improvement (NCASI) Technical Report: 26.

37. Afungang RN, Bateira C (2016) Temporal probability analysis of landslides triggered by intense rainfall in the Bamenda Mountain Region, Cameroon. Environmental Earth Sciences journal.

38. Hawkins P, Brunt M (1965) Soils and ecology of west Cameroon, FAO, Rome 516-722.

39. Hapke B, Shepard MK, Nelson RM, Smythe WD, Piatek JL (2009) A quantitative test of the ability of models based on the equation of radiative transfer to predict the bidirectional reflectance of a well-characterized medium. Icarus 199: 210-218.

40. Redweik P, Matildes R, Marques F, Santos L (2009) Photogrammetric methods for monitoring cliffs withlow retreat rate. Journal of Coastal Research 56: 1577-1581.

41. Chung CJ, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. Natural Hazards 30: 451-472.

42. Raia S, Alvioli M, Rossi M, Baum RL, Godt J, et al. (2014) Improving predictive power of physically based rainfall-induced shallow landslide models: a probabilistic approach. Geoscience Model Development 7: 495-514.

43. Tien BD, Biswajet P, Owe L, Inge R, Oystein BD (2012) Spatial prediction of landslide hazards in HoaBinh province(Vietnam): A comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. Catena 96: 28-40.

44. Atkinson PM, Massari R (1998) Generalized linear modeling of landslide susceptibility in the central Apennines, Italy. Computers and Geosciences 24: 373–385.

45. Blahut J, Van Westen C, Sterlacchini S (2010) Analysis of landslide inventories for accurate prediction of debris-flow source areas. Geomorphology 119: 36-51.

46. Yao X, Dai FC (2006) Support vector machine modeling of landslide susceptibility using a GIS: A case study. IAEG Paper number 793. The Geological Society of London.

47. Che VB, Kervyn M, Suh CE, Fontijn K, Ernst GG, et al. (2012) Landslide susceptibility assessment in Limbe (SW Cameroon): A field calibrated seed cell and information value method. Catena 92: 83-98.

48. Abella E (2008) Multi-scale landslide risk assessment in Cuba. International Institute for Geo-Information Science and Earth Observation. ITC Dissertation 154: 1- 273.

49. Hosmer D, Lemeshow S (2000) Applied logistic regression. Second Edition. New York, NY Wiley.

50. Umar Z, Pradhan B, Ahmad A, Neamah MJ, Mahyat ST (2014) Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. Catena 118: 124-135.

51. Papathanassiou G, Valkaniotis S, Athanassios G, Spyros P (2012) GIS-based statistical analysis of the spatial distribution of earthquake-induced landslides in the island of Lefkada, Ionian Islands, Greece. Landslides 10: 771-783.

52. Lee S (2007) Landslide susceptibility mapping using an artificial neural network in the Gangneung area, Korea. International Journal of Remote Sensing 28: 4763-4783.

53. Frattini P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. Engineering Geology 111: 62-72.

54. Ninu KM, Pratheesh P, Rejith PG, Hamza V (2014) Determining the Suitability of Two Different Statistical Techniques in Shallow Landslide (Debris Flow) Initiation Susceptibility Assessment in the Western Ghats. Environmental Research Engineering and Management 70: 27-39.