

Applying Logistic Regression for Landslide Susceptibility Mapping. The Case Study of Krathis Watershed, North Peloponnese, Greece

Dionysis Horafas^{1,*}, Theodora Gkeki²

¹Civil Engineer, Xylokaastro, Greece

²Civil Engineer, Derveni, Greece

Abstract The main objective of the present study was to produce a landslide susceptibility map by applying a Logistic regression model in the watershed of Krathis River that is located in the Achaia County, North Peloponnese, Greece. Five parameters were analyzed, namely: engineering geological units, slope angle, slope aspect, distance from faults and distance from river network. Each parameter was classified into different classes and weighted according to their susceptibility to slide. It was evaluated that the developed model classified correctly over 80% of the validation data. The developed model could be considered as a useful tool for the national and local authorities in order to evaluate strategies to prevent and mitigate the impact of landslides.

Keywords Landslide susceptibility, Logistic regression, Krathis River, Greece

1. Introduction

Landslides are geological phenomena that are characterized by a wide range of soil, debris or rock mass movements that may occur in offshore, coastal and inland areas, driven by the force of gravity and the aid of water [1]. Landslides are the result of the progressive or extreme evolution of natural events that occur due to the action of geological, tectonic, geomorphological and climatic processes.

The methods and techniques that are used in landslide susceptibility assessments, which are defined by the spatial component of landslide occurrence, could be classified into two main approaches; the data driven approach that is based on the exploration of data and the knowledge driven approach that is based on the assessment of knowledge [2]. The knowledge driven approach methods are based on the site specific experience of experts with the landslide susceptibility determined directly in the field or by combining different layered index maps, while the data driven approach methods perform statistical and probabilistic analysis or follow deterministic approaches [3].

Among the wide range of statistical methods proposed in the assessment of landslide susceptibility, logistic regression

analysis (LR) is one of the most reliable approaches [4-7]. LR is a statistical technique that involves one or more independent variables in order to predict the probability of a binary or categorical dichotomous dependent variable [8]. The objective of LR analysis is to identify the best predictive model which describes the relations between the dependent variable and multiple independent variables [9]. Thus, by utilizing LR it could be possible to model the probability of presence and absence of the dependent variable. The main advantage of LR model over linear and log-linear regression models is that it does not assume normality among variables.

In this context, the present study utilizes the logistic regression method to establish a landslide susceptibility map. As a case study, Krathis water basin at North Peloponnese, Greece has been selected.

2. Study Area and Data

The study area is located at the northern part of Peloponnese, Greece. It concerns the Krathis watershed, approximately 145 km² (Figure 1). Concerning the morphological settings, the relief of the wider area is influenced by the geological structure, the recent tectonic activity and the ongoing weathering and erosion mechanisms. The area is characterized as mountainous with strong relief, massive rocky limestone ridges and high peaks. In particular, the highest observed altitude is 2,310 m, with a mean elevation 985 m. Areas with slopes greater than 46° cover approximately 3.0% of the total area, while areas with slope angle less than 15° cover about 25%.

* Corresponding author:

horafas13@gmail.com (Dionysis Horafas)

Published online at <http://journal.sapub.org/ajgis>

Copyright © 2017 Scientific & Academic Publishing. All Rights Reserved

The climate type of the area is Mediterranean (Csa) with mild winters and dry and hot summers. The rainy season is from October to May, with December as the rainiest month (128.9 mm) followed by November (124.7 mm), while the driest month appears to be August (7.0mm) followed by July (8.8 mm). The climate data were obtained from the University of East Anglia Climate Research Unit (CRU) and referred to a period over 100 years between 1901 and 2008 [10].

A detail database concerning 36 rotational and translational slides, and rockfalls that also provided the date of slide, the type, the triggering factor and the severity of the phenomena was available. The geo-environmental conditions in those locations were analysed concerning five parameters: engineering geological units, slope angle, slope aspect, distance from faults, distance from river network.

The geological formations that are present and cover the wider area are [11]: Quaternary formations (loose coarse grained deposits, loose deposits of mixed phases), Plio-Pleistocene deposits (coarse grained sediments, fine

grained sediments), flysch formations, limestones and dolomites, shale and cherts formations and volcanic rocks (Figure 2).

For the purpose of the present study the slope aspect layer was classified into eight classes (Figure 3). North (337.5-22.5), Northeast (22.5-67.5), East (67.5-112.5), Southeast (112.5-157.5), South (157.5-202.5), Southwest (202.5-247.5), West (247.5-292.5), Northwest (292.5-337.5). As for the slope angle layer it was also classified into four classes according to the local geological and geotechnical conditions. Class A (<15°), class B (16°-30°), class C (31°-45°) and class D (>46°) (Figure 4).

The tectonic characteristics, mainly faults, thrusts and overthrusts were mapped based on an existing geological map, scale 1:50.000 [11]. The research area was classified into a three class layer, areas that cover zones that have distance less than 250 m from tectonic features, areas that cover zones that have distance between 251 and 500 m, and areas with distance greater than 501 m from tectonic characteristics (Figure 5).



Figure 1. Study area

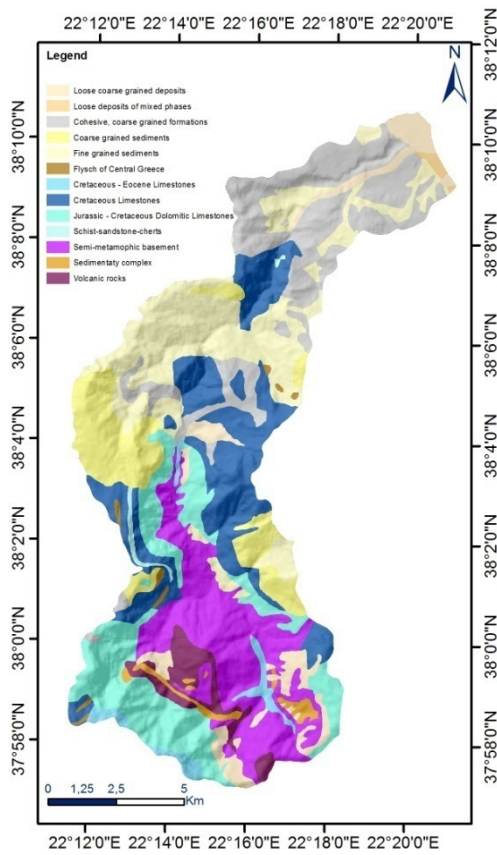


Figure 2. Engineering geology units

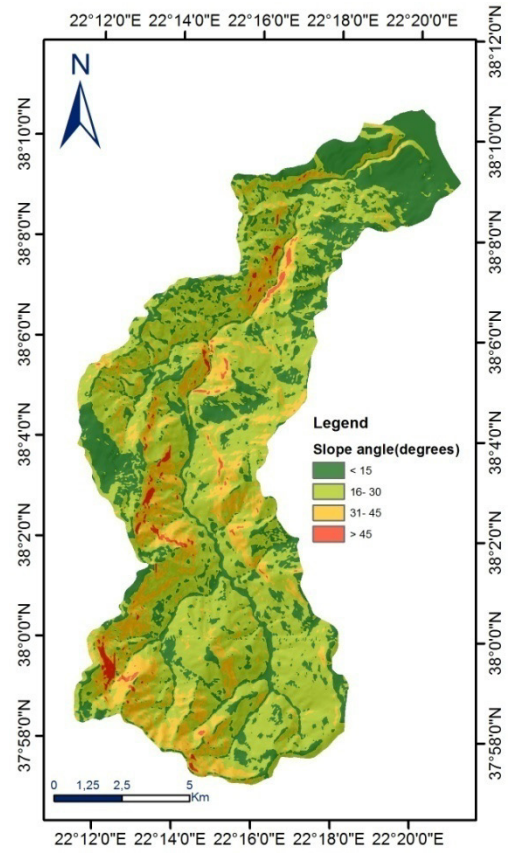


Figure 4. Slope angle

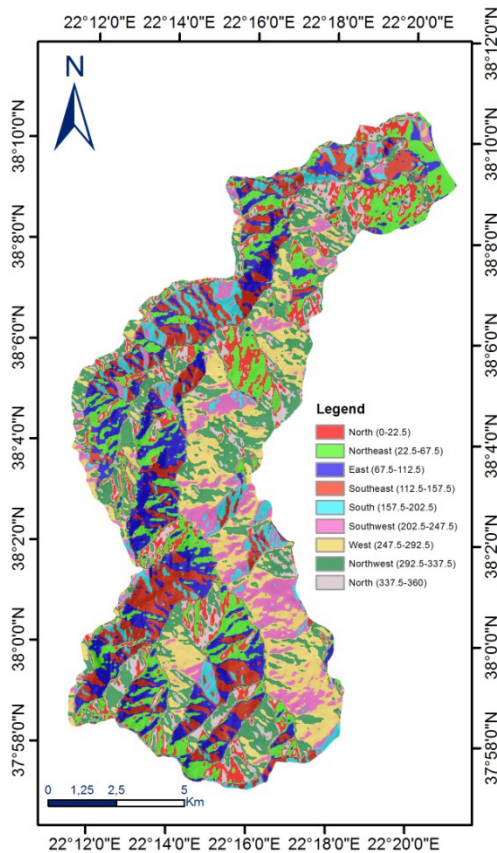


Figure 3. Slope aspect

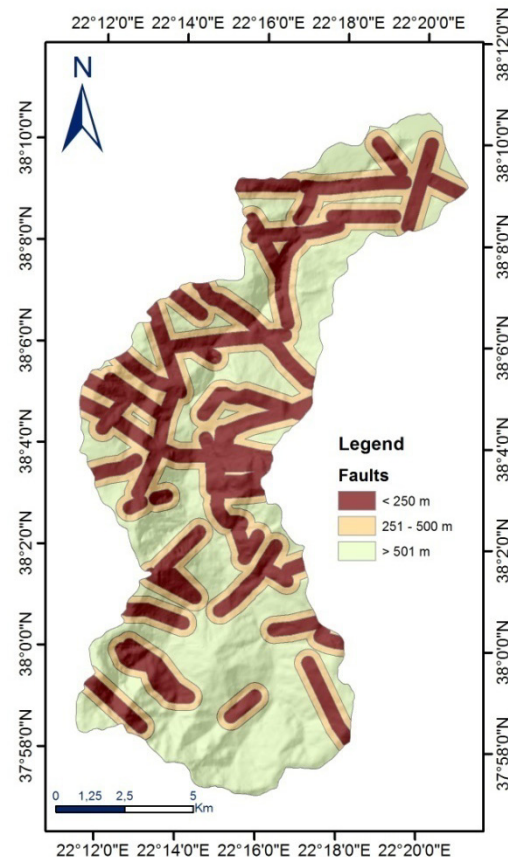


Figure 5. Distance from faults

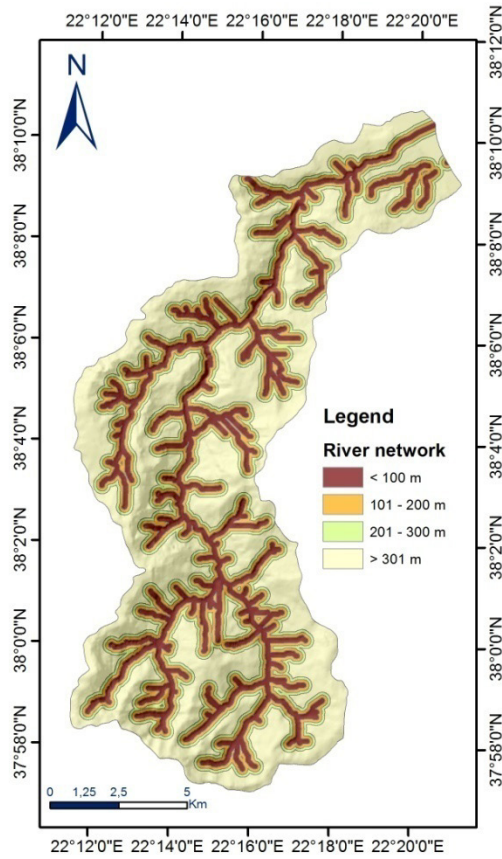


Figure 6. Distance from river network

Finally, the distance to the river network was classified into a three class layer using the Euclidean distance between the sample grid cell and the nearest hydrographic network (<150 m, 151 – 300 m, >301 m) (Figure 6).

3. Methodology

3.1. Data Preparation

For the purpose of the study, the landslide dataset was randomly divided into two subsets: 80% of the landslide data were used for training and the remaining 20% for validating the developed model. Thirty six non-landslide areas were selected randomly within the research area and separated also into training and validating data.

3.2. Logistic Regression

When performing LR analysis the objective is to correlate the probability of landslide occurrence, that can take values from 0 to 1, to the “logit” Z ($-\infty < Z < 0$ for higher odds of non - occurrence and $0 < Z < \infty$ for higher odds of occurrence). The probability of landslide occurrence is expressed by the following equation:

$$p = 1 / 1 + e^{-z} \quad (1)$$

The logit Z is assumed to express the independent parameters on which landslide occurrence may depend. The

LR analysis assumes the term Z to be a product of the independent set of parameters X_i ($i = 1, 2, \dots, n$) acting as potential causal factors of landslide phenomena. Z is expressed by the following linear equation:

$$Z = \beta_0 + \beta_1 x X_1 + \beta_2 x X_2 + \dots + \beta_n x X_n \quad (2)$$

where coefficients β_i ($i = 1, 2, \dots, n$) are representative of the contribution of single independent variables X_i to the logit Z and β_0 is the intercept of the regression function. The LR methodology does not imply linear dependencies between the dependent variable and the independent set of variables; instead an exponential function is involved. The coefficients β are calculated through the maximum likelihood criterion and correspond to the estimation of the more likely unknown factors.

3.3. Landslide Susceptibility Mapping

The produced map was classified into five categories of susceptibility, namely very high susceptibility (VHS), high susceptibility (HS), moderate susceptibility (MS), low susceptibility (LS) and very low susceptibility (VLS), using the natural break method for the determination of the class intervals [12]. The computation process was carried out using SPSS for applying logistic regression and ArcGIS 10.3 was used for compiling and analysing the data and also for producing the landslide susceptibility maps.

4. Results and Discussion

The relative importance of the independent parameters was assessed using the coefficients of the logistic regression function (Table 1). According to the findings, the variables of engineering geology units, slope angle and distance from river network had a positive effect on the LR function. On the other hand, the slope aspect and distance from faults had a negative effect on landslide occurrence. Slope angle and engineering geologic units were found to be the most important variables that contribute to slope instability as they have the highest coefficients, 3.890 and 3.171 respectively. The outcomes of the present study is in agreement with the majority of the LS literature concerning landslide assessments [13-15], which have found that variations in landslide distribution are highly dependent on geological formations and slope angle. Concerning the research area, according to [16], fine grained Plio-Pleistocene sediments, which consist of alternations of clayey marls, marls, silty sands and weak sandstones, appear to be much more susceptible in rotational slides. Concerning, limestone formations, they appear susceptible to rockfalls, that are influenced by the degree of weathering and fragmentation, the orientation of the discontinuities surfaces and the intense morphological relief.

The training dataset was evaluated using Cox and Snell R^2 (0.567) and Nagelkerke R^2 tests (0.756) (Table 2) indicating a good performance.

Table 1. Coefficients of independent variables in logistic regression

Variables	Coefficient
Intercept (β_0)	-11.632
Engineering geology units	3.171
Slope aspect	-1.123
Slope angle	3.890
Distance from faults	-1.767
Distance from river network	2.318

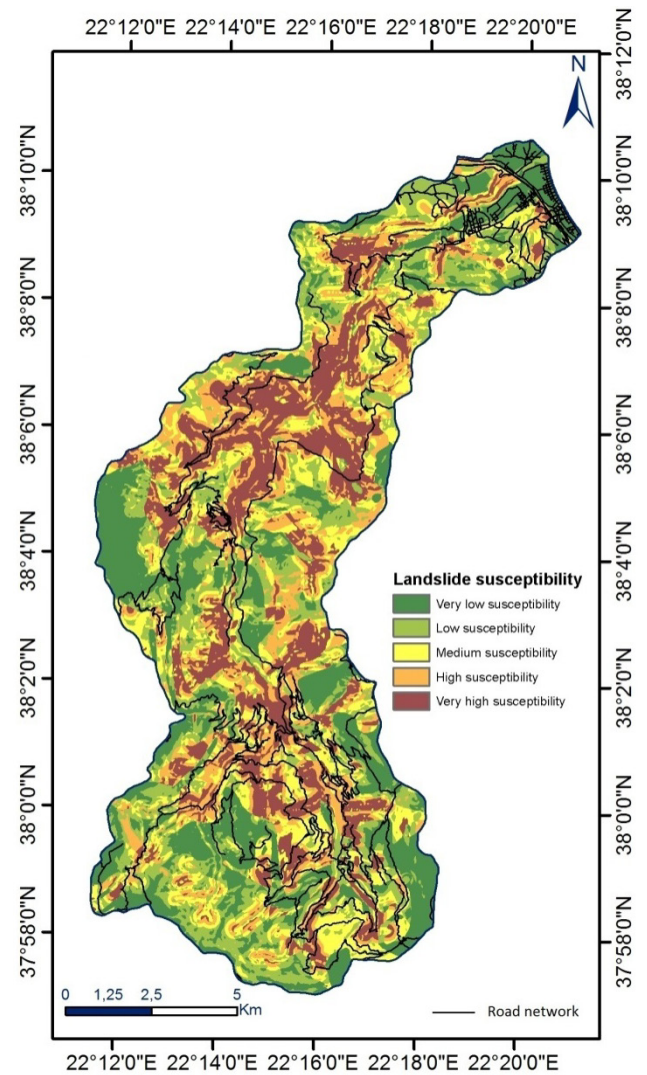
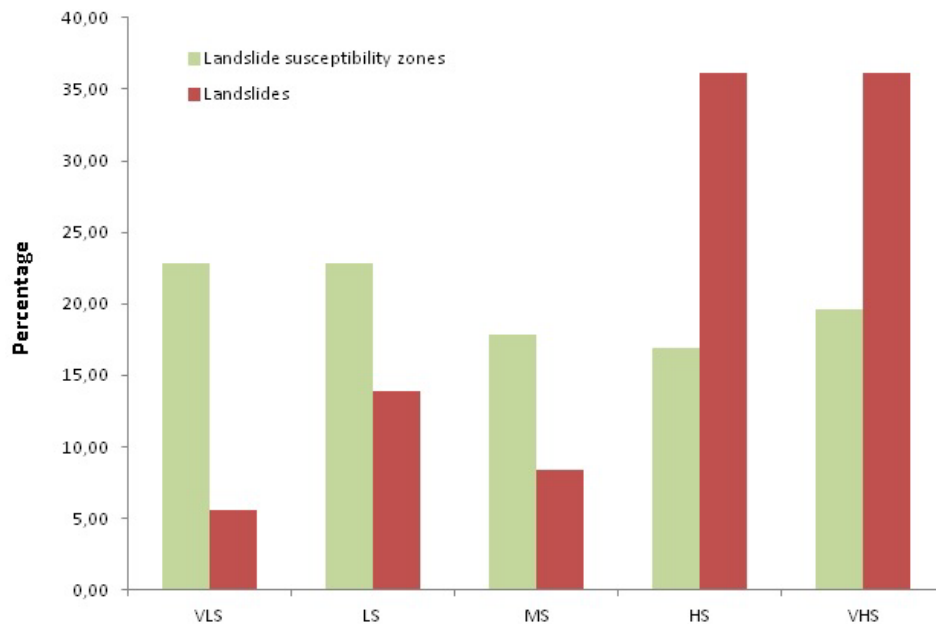
Table 2. Statistics

-2log likelihood	Cox and Snell R^2	Nagelkerke R^2
31.844	0.567	0.756

The outcomes of the experiment showed that 82.80% of the instances during the training phase were correctly classified. During the validation phase, the LR model achieved an accuracy of 85.70% (Table 3).

Table 3. Confusion matrix of the training and validating dataset

Training dataset		Predicted		
Observed		non-landslide	landslide	%
	non-landslide	22	6	78.60
	landslide	4	26	86.70
				82.80
Validating dataset		Predicted		
Observed		non-landslide	landslide	%
	non-landslide	7	1	87.50
	landslide	1	5	83.30
				85.70

**Figure 7.** Landslide susceptibility map**Figure 8.** Study area

The logit of $f(x)$ function was calculated for all of the grids of the research area, in which zero corresponds to no susceptibility and one to total susceptibility. Based on constant values that were calculated, the LR model was compiled according to equation 1, while the possibility of landslide occurrence in each grid was calculated from equation 1 the outcome of which produced the landslide susceptibility map (Figure 7).

From the visual analysis of the landslide susceptibility map, high and very high susceptible zones are located at the central area of the research area with the spatial pattern of the landslide susceptibility following the distribution of the river network. The LR analysis revealed that several sections of the road network fall within the very high susceptibility zone, a finding that should concern the local and national authorities. Specifically, within the watershed of Krathis River, approximately 60 km of the road network (17.70% of a total of 343 km), was estimated to be classified as highly susceptible to landslide.

Concerning the produced landslide susceptibility map, the very high and high susceptibility class was estimated to cover 19.64% and 16.89% respectively, of the total research area. The relative landslide density for the high and very high landslide susceptibility class was estimated to be 72.22% (Figure 8).

5. Conclusions

The presented study focused on the construction of a landslide susceptibility map in the watershed of Krathis River that is located in the Achaia County, North Peloponnese, Greece, through the implementation of logistic regression method. Five landslide conditioning variables were analyzed and included in the study, namely engineering geological units, slope angle, slope aspect, distance from faults and distance from river network. The landslide inventory data contained thirty six landslides that were divided into two subsets, one for training (80% of the total data) and one for estimating the prediction capabilities of the developed methodology. Slope angle and engineering geological unit and distance from river network were among the most susceptible parameters. The LR model achieved an accuracy of correctly predicting landslide occurrence that reached 85.70%, indicating a good predictive performance. The findings of the analysis revealed that several sections of the road network fall within the very high susceptibility zone, a finding that should concern the local and national authorities.

REFERENCES

- [1] Hutchinson, JN., 1995. Keynote paper: Landslide hazard assessment. Proceedings 6th International Symposium on Landslides, Christchurch. Balkema, Rotterdam, pp

- 1805-1841.
- [2] Caniani, D., Pascale, S., Sdao, F., Sole, A., 2008. Neural networks and landslide susceptibility: a case study of the urban area of Potenza. *Natural Hazards* (45): 55-72.
- [3] Pourghasemi, HR., Moradi, HR., Fatemi, Aghda. SM., 2013. Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances. *Natural Hazards*, 69(1): 749-779.
- [4] Dai, FC., Lee, CF., Ngai, YY., 2002. Landslide risk assessment and management: an overview. *Engineering Geology*, 64(1): 65-87.
- [5] Ayalew, L., 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.
- [6] Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., Ardizzone, F., 2005. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72: 272-299.
- [7] Tsangaratos, P., and Ilia, I., 2016. Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena*, 145:164-179.
- [8] Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*, second ed. John Wiley and Sons, Inc., New York, 375p, <<http://www.nesug.org/proceedings/nesug06/an/da26.pdf>>.
- [9] Lee, S., 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* 26 (7): 1477-1491.
- [10] Jones, PD., Harris, I., 2008. Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena. University of East Anglia Climatic Research Unit, NCAS British Atmospheric Data Centre.
- [11] IGME, 2005. Geological map of Greece, at a scale of 1:50,000, Aigion sheet, Athens.
- [12] Feizizadeh, B., Blaschke, T., 2013. GIS-multicriteria decision analysis for landslide susceptibility mapping: comparing three methods for the Urmia lake basin, Iran. *Natural Hazards*, 65(3): 2105-2128.
- [13] Akgun, A., 2012. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at Izmir, Turkey. *Landslides*, 9(1):93-106.
- [14] Gokceoglu, C., Sonmez, H., Nefeslioglu, H.A., Duman, T.Y., Can, T., 2005. The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity. *Eng. Geol.* 81:65-83.
- [15] Lee, S., Pradhan, B., 2007. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* 4:33-41.
- [16] Rozos, D., 1989. Engineering-geological conditions in the Achaia County. Geomechanical characteristics of the Plio-pleistocene sediments. PhD thesis, University of Patras, Patras, pp 453 (in Greek, with extensive summary in English).