Geophysical Research Abstracts
Vol. 19, EGU2017-15560, 2017
EGU General Assembly 2017

Producing landslide susceptibility maps by utilizing machine learning methods. The case of Finikas catchment basin, North Peloponnese, Greece.

Paraskevas Tsangaratos (1), Ioanna Ilia (2), Constantinos Loupasakis (3), Michalis Papadakis (4), and Antonios Karimalis (5)

(1) Ph.D., Mining and Metallurgical Engineer, National Technical University of Athens, School of Mining and Metallurgical Engineering, Department of Geological Studies, Athens, Greece (corresponding author)., (2) Ph.D., M.Sc., Geologist, National Technical University of Athens, School of Mining and Metallurgical Engineering, Department of Geological Studies, Athens, Greece, gilia@metal.ntua.gr , (3) Associate Professor, Ph.D., M.Sc., Geologist, Civil Engineer, National Technical University of Athens, School of Mining and Metallurgical Engineering, Department of Geological Studies, Athens, Greece, cloupasakis@metal.ntua.gr, (4) Ph.D., Tunnel construction engineer, J&P AVAX S.A., Athens, Greece, mikpap@yahoo.com , (5) Tunnel construction engineer, J&P AVAX S.A., Athens, Greece, akarimalis@yahoo.gr

The main objective of the present study was to apply two machine learning methods for the production of a landslide susceptibility map in the Finikas catchment basin, located in North Peloponnese, Greece and to compare their results. Specifically, Logistic Regression and Random Forest were utilized, based on a database of 40 sites classified into two categories, non-landslide and landslide areas that were separated into a training dataset (70% of the total data) and a validation dataset (remaining 30%). The identification of the areas was established by analyzing airborne imagery, extensive field investigation and the examination of previous research studies. Six landslide related variables were analyzed, namely: lithology, elevation, slope, aspect, distance to rivers and distance to faults. Within the Finikas catchment basin most of the reported landslides were located along the road network and within the residential complexes, classified as rotational and translational slides, and rockfalls, mainly caused due to the physical conditions and the general geotechnical behavior of the geological formation that cover the area.

Each landslide susceptibility map was reclassified by applying the Geometric Interval classification technique into five classes, namely: very low susceptibility, low susceptibility, moderate susceptibility, high susceptibility, and very high susceptibility. The comparison and validation of the outcomes of each model were achieved using statistical evaluation measures, the receiving operating characteristic and the area under the success and predictive rate curves. The computation process was carried out using RStudio an integrated development environment for R language and ArcGIS 10.1 for compiling the data and producing the landslide susceptibility maps. From the outcomes of the Logistic Regression analysis it was induced that the highest b coefficient is allocated to lithology and slope, which was 2.8423 and 1.5841, respectively. From the estimation of the mean decrease in Gini coefficient performed during the application of Random Forest and the mean decrease in accuracy the most important variable is slope followed by lithology, aspect, elevation, distance from river network, and distance from faults, while the most used variables during the training phase were the variable aspect (21.45%), slope (20.53%) and lithology (19.84%). The outcomes of the analysis are consistent with previous studies concerning the area of research, which have indicated the high influence of lithology and slope in the manifestation of landslides. High percentage of landslide occurrence has been observed in Plio-Pleistocene sediments, flysch formations, and Cretaceous limestone. Also the presences of landslides have been associated with the degree of weathering and fragmentation, the orientation of the discontinuities surfaces and the intense morphological relief.

The most accurate model was Random Forest which identified correctly 92.00% of the instances during the training phase, followed by the Logistic Regression 89.00%. The same pattern of accuracy was calculated during the validation phase, in which the Random Forest achieved a classification accuracy of 93.00%, while the Logistic Regression model achieved an accuracy of 91.00%. In conclusion, the outcomes of the study could be a useful cartographic product to local authorities and government agencies during the implementation of successful decision-making and land use planning strategies.

Keywords: Landslide Susceptibility, Logistic Regression, Random Forest, GIS, Greece.