Geophysical Research Abstracts
Vol. 16, EGU2014-11407, 2014
EGU General Assembly 2014

Landslide Susceptibility Analysis by the comparison and integration of Random Forest and Logistic Regression methods; application to the disaster of Nova Friburgo - Rio de Janeiro, Brasil (January 2011)

Carlo Esposito (1), Anna Barra (2), Stephen G. Evans (3), Gabriele Scarascia Mugnozza (1), and Keith Delaney (3)

(1) "Sapienza" University of Rome, Department of Earth Sciences, Rome, Italy (carlo.esposito@uniroma1.it; gabriele.scarasciamugnozza@uniroma1.it), (2) Geologist, Rome, Italy (anna.barra1@gmail.com), (3) University of Waterloo, Department of Earth and Environmental Sciences, Waterloo (ON), Canada (sgevans@uwaterloo.ca; kbdelane@uwaterloo.ca)

The study of landslide susceptibility by multivariate statistical methods is based on finding a quantitative relationship between controlling factors and landslide occurrence. Such studies have become popular in the last few decades thanks to the development of geographic information systems (GIS) software and the related improved data management. In this work we applied a statistical approach to an area of high landslide susceptibility mainly due to its tropical climate and geological-geomorphological setting. The study area is located in the south-east region of Brazil that has frequently been affected by flood and landslide hazard, especially because of heavy rainfall events during the summer season. In this work we studied a disastrous event that occurred on January 11th and 12th of 2011, which involved Região Serrana (the mountainous region of Rio de Janeiro State) and caused more than 5000 landslides and at least 904 deaths. In order to produce susceptibility maps, we focused our attention on an area of 93,6 km2 that includes Nova Friburgo city. We utilized two different multivariate statistic methods: Logistic Regression (LR), already widely used in applied geosciences, and Random Forest (RF), which has only recently been applied to landslide susceptibility analysis. With reference to each mapping unit, the first method (LR) results in a probability of landslide occurrence, while the second one (RF) gives a prediction in terms of % of area susceptible to slope failure. With this aim in mind, a landslide inventory map (related to the studied event) has been drawn up through analyses of high-resolution GeoEye satellite images, in a GIS environment. Data layers of 11 causative factors have been created and processed in order to be used as continuous numerical or discrete categorical variables in statistical analysis. In particular, the logistic regression method has frequent difficulties in managing numerical continuous and discrete categorical variables together; therefore in our work we tried different methods to process categorical variables , until we obtained a statistically significant model. The outcomes of the two statistical methods (RF and LR) have been tested with a spatial validation and gave us two susceptibility maps. The significance of the models is quantified in terms of Area Under ROC Curve (AUC resulted in 0.81 for RF model and in 0.72 for LR model). In the first instance, a graphical comparison of the two methods shows a good correspondence between them. Further, we integrated results in a unique susceptibility map which maintains both information of probability of occurrence and % of area of landslide detachment, resulting from LR and RF respectively. In fact, in view of a landslide susceptibility classification of the study area, the former is less accurate but gives easily classifiable results, while the latter is more accurate but the results can be only subjectively classified. The obtained “integrated” susceptibility map preserves information about the probability that a given % of area could fail for each mapping unit.