Evaluation of Landslide Susceptibility using Multivariate Statistical Methods: A Case Study in the Prahova Subcarpathians, Romania

Z. Chitu, I. Sandric, B. Mihai, I. Savulescu
University of Bucharest, Faculty of Geography

ABSTRACT: The shallow landslides affect most of the hillslopes in the Prahova Subcarpathians and are mainly caused by the agriculture practices. Though, these processes do not imply immediately losses for human activity, the landslide susceptibility assessment is necessary to prevent terrain degradation. The evaluation of landslide susceptibility requires understanding of spatial distribution of the factors that control slope instability. It is known that the behavior of landslides is difficult to evaluate because of the various factors that trigger the mass movements. One of the most efficient methods to establish the importance of the factors in landslides susceptibility is the multivariate analysis, like Principal Component Analysis (PCA) and logistic regression. For this study, the PCA was applied to remove the redundant information from the elements that control slope instability like lithology, faults, slope gradient, plan curvature, profile curvature, land use and land cover, soils etc. The PCA analysis had reduced the factors at six principal components and the first three components comprise over 85% of the information. The logistic regression analysis was used to correlate the presence of the past and actual landslides with the PCA components and derive a landslides susceptibility map. To validate the model we used the Receiver Operating Characteristic (ROC) analysis. ROC shows a good correlation between the observed and predicted values of the validation dataset.

Methodology: When many factors are available an issue of information redundancy is arising. To reduce the number of variables and to limit their interdependence we choose to use the PCA analysis (Campana 1986). It is well known that morphometrical parameters derived from DEM present a high interdependence and this, often, lead to underestimation or overestimation of the results. The principal components analysis was used in order to reduce the redundant information from the variables and transform them from correlated variables into uncorrelated variables, named principal components (Gorsevski 2001). To model the susceptibility for landslides we used the logistic regression analysis, which calculates the probability that an individual pixel will contain a landslide. The principal components were used as independent variables and the past and present shallow landslides as dependent variable. We have combined the first three components, then the first four components and the first five components.

Logistic regression analysis: Logistic regression was applied for three different combination of the principal components obtained from the PCA analysis. The principal components were used as independent variables and the past and present shallow landslides as dependent variable. We have combined the first three components, then the first four components and the first five components. In this way the logistic regression for the combinations of the first three, four and five principal components with the past and present shallow landslides are:

logit(SL)=-7.3936+0.643115*PC1+0.383955*PC2-0.104961*PC3 (1)

logit(SL)=-7.0373+0.578237*PC1 + 0.296697*PC2-0.120301*PC3-0.267935*PC4 (2)

logit(SL)=-7.8027+0.611648*PC1+0.330487*PC2-0.110811*PC3+0.310245*PC4+0.352099*PC5 (3)

All three equations have produced susceptibility maps with a ROC value very similar (0.739-0.75). The best results were obtained by combining the first four principal components, having the highest ROC value of 0.75. The highest susceptibility values are present on the hillslopes of Tarsa River and on the left hand side of Prahova River. These areas correspond marls and clay deposits, small vegetation and gentle slope gradients between 10 and 25. Low susceptibility areas correspond to flat areas on top of the terraces with deposits of sand and gravel and under the forest vegetation from the hillslopes of Tarsa River. The cutting threshold between low susceptibility and high susceptibility is set by the logistic regression at 0.06.

DISCUSSIONS: The map represents the landslides susceptibility for shallow landslides in the administrative area of Breaza. The map accurately describes areas that are prone to shallow landslides, but it also presents areas that are not prone to shallow landslides or are affected by deep-seated landslides. The terrace scarp of Prahova, on the east part of Breaza de Sus city, are partially identified with high susceptibility for shallow landslides, but these areas are covered by old deep-seated landslides with a latent state of activity. The areas correctly identified with high susceptibility for shallow landslides are on the hilltops of Tarsa River and in the south and Velea Turscoi village in the north and east of Surdesti village. The values of ROC are higher than 0.5, which indicate a random fir. The ROC value of 0.75 is comparable with ROC values reported in other fields. To validate the model we used the Receiver Operating Characteristic (ROC) analysis. ROC shows a good correlation between the observed and predicted values of the validation dataset.

RESULTS: Upon examination of the normalized eigenvectors, the elements with the absolute largest value indicated that the first principal component have a strong relationship with the slope gradient and energy of the relief, the second principal component have a strong relationship with land-use, land-cover and lithology, the third principal component with width drainage density and plan curvature, the fourth principal component with aspect and the fifth principal component with the profile curvature and faults density. Logistic regression analysis has produced a series of three landslides susceptibility. The logistic regression for the combinations of the first three, four and five principal components with the past and present shallow landslides are:

logit(SL)=-7.3936+0.643115*PC1+0.383955*PC2-0.104961*PC3 (1)

logit(SL)=-7.0373+0.578237*PC1 + 0.296697*PC2-0.120301*PC3-0.267935*PC4 (2)

logit(SL)=-7.8027+0.611648*PC1+0.330487*PC2-0.110811*PC3+0.310245*PC4+0.352099*PC5 (3)

All three equations have produced susceptibility maps with a ROC value very similar (0.739-0.75). The best results were obtained by combining the first four principal components, having the highest ROC value of 0.75. The highest susceptibility values are present on the hillslopes of Tarsa River and on the left hand side of Prahova River. These areas correspond marls and clay deposits, small vegetation and gentle slope gradients between 10 and 25. Low susceptibility areas correspond to flat areas on top of the terraces with deposits of sand and gravel and under the forest vegetation from the hillslopes of Tarsa River. The cutting threshold between low susceptibility and high susceptibility is set by the logistic regression at 0.06.

CONCLUSIONS: The PCA analysis can be used to identify the shallow landslides susceptibility with satisfactory results. ROC analysis enables scientist to validate a model ability to identify a specific location. The ROC values obtained are comparable with ROC values reported by scientist from other research fields.