Exploring the relative importance of environmental variables across scales in landslide susceptibility mapping

F. Catani, D. Lagomarsino, D. Petracca, S. Segoni, and V. Tofani

University of Firenze, Earth Sciences, Firenze, Italy (filippo.catani@unifi.it, +39 055 2756296)

The purpose of this study is to apply a regression method, based on a specific version of the random forest algorithm, to produce a series of susceptibility maps of the Arno river basin (Central Italy) and to analyze the contribution that each selected preparatory variable has on the final outcome according to varying scales and parameter sets.

Random forest is a combination of tree (usually binary) bayesian predictors that permits to relate a set of contributing factors with the actual landslides occurrence. Being it a nonparametric model, it is possible to incorporate a range of numeric or categorical data layers and there is no need to select unimodal training data.

The study area is divided into three distinct macro-areas, homogeneous from a geological and lithological point of view. Several classical and widely acknowledged landslide predisposing factors have been taken into account as mainly related to: the lithology, the land use the land surface geometry (derived from of DTM). In addition, for each factor we also included in the parameter set the standard deviation (for numerical variables) or the variety (for categorical ones).

The use of random forest enables to estimate the relative importance of the single input parameters and to select the optimal configuration of the regression model. The model was initially applied using the complete set of input parameters at disposal, automatically assigning them a rank by relevance and calculating the ROC curve (with relative AUC value) using an independent testing dataset. Subsequently reduced versions of the random forest model were applied taking into account a progressively lower number of parameters. Step by step the least relevant parameters were discarded and the AUC values of every run was used to assess the effectiveness of the regression model. This procedure has been applied for each area in order to check which parameters need to be taken into account to best evaluate the landslide susceptibility in the study area. Considering the best set of parameters for each macro-area and the impact of scale and accuracy of input variables, the consequences on susceptibility applications are discussed.