Geophysical Research Abstracts
Vol. 21, EGU2019-8057, 2019
EGU General Assembly 2019

Landslide susceptibility assessment using Machine Learning: the Valais Canton (Switzerland) case study

Marj Tonini (1), Federico Amato (1), and Nicola Deluigi (2)

(1) Institute of Earth Surface Dynamics, Faculty of Geosciences and the Environment, UNIL, Lausanne, Switzerland (marj.tonini@unil.ch, federico.amato@unil.ch), (2) Stream Biofilm and ecosystem research laboratory, ENAC, EPFL , Lausanne, Switzerland

Landslide susceptibility measures the tendency of an area to be affected by landslides, ranging from low to high. According to the basic assumption that "the past is the key to the future", new landslides are expected to occur under similar conditions as observed past events; susceptibility maps can thus be produced based on predisposing factors and on inventories of historical events.

The main objective of the present study is the elaboration of landslide susceptibility maps for the Valais Canton (Switzerland) using a machine learning (ML) algorithm. ML are statistical-based approaches capable of learning from data and of making predictions from the acquired knowledge on the studied phenomenon. The modelling process extrapolates the hidden relationships existing between a set of input and output variables. After a training procedure to calibrate the model parameters, susceptibility maps can be displayed. Belonging to a data-driven framework is one of the main benefit of ML approaches; hence, they do not need a priori knowledge on the process. The available landslide inventory gathers various types of observed gravitational slope deformations of the study region (Pedrazzini et al., 2016). The environmental variables considered as predisposing factors are: slope angle, curvature, profile curvature, plan curvature, lithology, NDVI, river proximity and road proximity.

For this study, we adopted Random Forest (Breiman, 2001), an ensemble ML algorithm based on decision trees. Computationally, a subset of the training dataset is generated by bootstrapping (i.e. random sampling with replacement). For each subset a decision tree is grown and, at each split, the algorithm randomly selects a number of variables and it computes the Gini index to identify the best one. The process stops when each node contains less than a fixed number of data points. The prediction of new data is finally computed taking the average value of all decision trees for regression and the maximum voting for classification. The prediction error is assessed by evaluating predictions on those observations that were not used in the subset, defined as "out-of-bag". Our model includes the implementation of the landslide pseudo-absences, which are the areas where the hazardous event did not take place (i.e. landslide location is known and the mapped footprint areas are available, but the non-landslide areas have to be defined). Indeed, to assure a good generalization of the model and to avoid the overestimation of low classes, pseudo-absences need to be generated in all the cases where they are not explicitly expressed.

The implemented model successfully allowed to elaborate a landslide susceptibility map for Valais Canton and it was able to define a variable importance ranking. More generally, the proposed approach proved to be particularly effective to deal with the large and high dimensional employed spatial dataset, which reflects the great flexibility of ML methods.

REFERENCES

Pedrazzini, A., Humair, F., Jaboyedoff, M., Tonini, M. 2016, Characterisation and spatial distribution of gravitational slope deformation in the Upper Rhone catchment (Western Swiss Alps), Landslides, 13:2, 259-277.
Breiman L, (2001). Random forests. Machine Learning, 45, 5-32