Geophysical Research Abstracts
Vol. 20, EGU2018-7894, 2018
EGU General Assembly 2018

The influence of data scaling in modelling landslide scarps

Flavius Sirbu (1), Lucian Dragut (1), Takashi Oguchi (2), and Yuichi Hayakawa (2)

(1) West University of Timisoara, Department of Geography, Romania (flavius.sirbu@gmail.com), (2) The University of Tokyo, Center for Spatial Information Science, Japan

Random forest (RF) is quickly becoming one of the most used algorithms in the modelling of landslides, mostly because it is relatively easy to use, computationally efficient and produces models with a high level of accuracy. It is acknowledged that the input data hold the most important influence on the RF prediction accuracy, through characteristics such as sampling strategy, spatial distribution, size and split of data and scale of the predictors.
In this study we focus on the last one, by evaluating the sensitivity of RF results to the scale of predictors when modelling landslide scarps. While previous studies have used different strategies (e.g. resampling) to fit all the predictors to the same scale of the modelled variable, we hypothesize that predictors would perform best at specific scales. Thus we present a novel approach in which we scale each predictor to best fit the landslide scarps and then use each one at its best scale. Each predictor was up-scaled using focal mean statistics in a moving window, starting from 3x3 and growing until the predictor was found to best fit the scarps. The degree of fitting was determined with logistic regression. The predictors at the scales that achieved the best degree of fitting made the input data for RF modelling.
The experiments were carried out in a test area of 20 sq. km in Shizuoka Prefecture, Japan, where a LiDAR DEM at 5 m spatial resolution, as well as a landslide inventory were available. A number of 13 land-surface variables (LSVs) were derived from the DEM and up-scaled as described previously, then entered as predictors in the RF modelling. The modelling was performed using the package “randomForest” from the software R, with the setting ntree = 501 and mtry = 3 (number of trees and number of candidate variables, respectively). The model was run for 25 times and the accuracy was tested using the built-in OOB (out of bag) error, by averaging the resulted error over all the runs and subtracting it from 100. The modelling was repeated with the same LSVs at the default scale, i.e. derived in a 3x3 window, for comparison.
The results show an improvement in the overall accuracy of the modelling, when using the scaled input data, from 81.11% to 92.46%. These preliminary results confirm the previous findings on the impact of data scaling in RF modelling and add the importance of fitting the predictors to their specific scales.