Altitude as an indicator of biased sampling design in landslide prediction

Andrei Dornik1, Lucian Drăguț1, Takashi Oguchi2, Yuichi Hayakawa2,3, and Mihai Micu4

1Department of Geography, West University of Timisoara, Romania
2Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8568, Japan
3Faculty of Environmental Earth Science, Hokkaido University, N10W5 Kita-ku, Sapporo, Hokkaido, 060-0810, Japan
4Institute of Geography, Romanian Academy, 023993, Bucharest, Romania

Variables related to terrain morphology are widely used and have proven particularly effective in landslides detection as well as susceptibility modelling. Altitude has often been found as one of the main predictors in landslide modelling, although it does not have clear conceptual or empirical justification as predisposing factor. As most other land-surface variables are derived from it, altitude might be just a surrogate for more meaningful predictors in a statistically-based landslide modeling. For instance, altitude might replace curvature simply because convexities tend to occur in upper parts of a landscape, while concavities are associated with lower altitudes. Our work intends to examine the hypothesis that altitude points out issues in sampling design when appears as a main predictor in landslide modeling. The tests were conducted in two study areas, one in the Buzău County, Romania and the other in the Shizuoka Prefecture, Japan, with landslide inventories available. Two sampling designs were tested in each study area: random sampling over the entire study area (random point allocation within each landslide scarp polygon and the same number of points randomly created outside landslide scarp area, as absence data), and stratified random sampling based on lithological strata. Following stratified random sampling based on lithological homogeneity, three study areas in Buzau and two in Japan resulted. Variable importance analysis and prediction of landslide scarp were conducted with Random Forest (RF) on databases with presence/absence of landslide scarp and associated values of 14 terrain variables. The results of variable importance analysis showed that variable hierarchy changed significantly when using lithological stratified sampling. In the random sampling scenario, altitude showed as the second most important landslide predictor in both study areas. In four out of five cases, the lithologically stratified random sampling led to decrease of altitude importance as landslide predictor, in two cases altitude even being one of least important variables. The results of model performance metrics showed that in four out of five cases the lithologically stratified random sampling significantly improved the prediction. In both areas in Japan, all four metrics show improvement of lithological stratified sampling over random sampling, by 6 and 4 % for AUC, 3% for OOB, 3 and 5 % for OA, and 6 and 10 % for Kappa, respectively. We conclude that landslide modeling is sensitive to lithological homogeneity and the presence of altitude as an important predictor could indicate a bias in the sampling design.