Geophysical Research Abstracts
Vol. 17, EGU2015-2724, 2015
EGU General Assembly 2015

Why a high statistical performance cannot be equated with a high plausibility of landslide susceptibility maps

Stefan Steger (1), Alexander Brenning (2,3), Rainer Bell (1), Helene Petschko (3), and Thomas Glade (1)

(1) University of Vienna, Geography and Regional Research, Vienna, Austria (stefan.steger@univie.ac.at), (2) Department of Geography and Environmental Management, University of Waterloo, Canada, (3) Department of Geography, Friedrich Schiller University, Jena, Germany

Statistical landslide susceptibility maps express a relative estimate of where landslides are more likely to occur in the future due to a set of geo-environmental conditions. Their predictive capability is regularly deduced by interpreting threshold independent performance measures like the area under the receiver operating characteristic curve (AUROC). These quantitative estimates frequently serve as a decision tool to favour a certain classifier over another and/or to select a suitable combination of predictors. Literature exposes that many authors consider their final maps as a valuable instrument for spatial planners and decision makers. However, most often the susceptibility maps are selected by solely interpreting such quantitative estimates.
We assume that a high statistical quality is necessary but not sufficient in order to produce plausible landslide susceptibility maps. This assumption was tested by quantitatively and qualitatively validating 16 susceptibility models for a study area (1354 km^2) located in Lower Austria. The models were generated by applying two statistical and two machine learning classifiers separately for two landslide inventories and two sets of predictors. Quantitative validation was conducted by estimating the AUROC with non spatial hold-out validation and a repeated spatial cross validation technique. The spatial differentiation of the final maps was evaluated at different scales by interpreting semivariograms. Maps of the location of major variations illustrate the spatial structure of the final susceptibility maps and allowed to deduce the most influential predictors and predictor classes.
According to the hold-out validation, all 16 susceptibility models performed similarly well. However, spatial cross validation revealed considerable differences between models generated by different landslide inventories. Semivariograms exposed that the predicted landslide susceptibility pattern differs substantially between maps generated by different classification methods, landslide inventories and predictor sets. The maps of major variations point out that implausible spatial patterns or inventory based biases are expressed by the final susceptibility maps. Moreover, the results indicate that an inclusion of a bias may enhance the apparent predictive performance of a model if a predictor describes this bias.
The findings provide valuable indications that a purely quantitative validation of the results may be not sufficient to produce plausible susceptibility maps. We conclude that a thorough non-purely quantitative evaluation of the final results may help to avoid misleading interpretations.