Geophysical Research Abstracts
Vol. 20, EGU2018-8551, 2018
EGU General Assembly 2018

Incompleteness matters – An approach to counteract inventory-based biases in statistical landslide susceptibility modelling

Stefan Steger (1,3), Alexander Brenning (2), Rainer Bell (3,4), and Thomas Glade (3)

(1) Eurac Research, Institute for Earth Observation, Bolzano-Bozen, Italy, (2) Friedrich Schiller University, Department of Geography, Jena, Germany, (3) University of Vienna, Geography and Regional Research, Vienna, Austria, (4) University of Innsbruck, Department of Geography, Innsbruck, Austria

A number of studies highlight that the explanatory power of a statistical landslide susceptibility model is conditional on a truthful spatial representation of past landslide occurrences. However, landslide inventories are often incomplete and subject to known or unknown systematic errors. We assume that ignoring such common landslide inventory biases might result in misleading modelling results, especially when crucial modelling decisions (e.g. variable selection) are solely based on obtained predictive performances.
The present research explored (i) under which conditions an inventory-based bias propagates into a landslide susceptibility model and (ii) whether and how such input data biases can be diminished.
Modelling experiments were performed with real-world data (test site in Lower Austria; 100 km^2) and synthetic data ("error-free" reference data). A simulation of various mapping biases (i.e. gradual deletion of landslide data within forests and specific administrative units) resulted in differently biased response variables. These differently error-prone data sets served as a basis for the subsequent modelling while we additionally tested the effects of various classifier-predictor combinations. All results were evaluated by means of state-of-the-art quantitative procedures (e.g. spatial cross-validation). In this context, a specific emphasize was placed on a confrontation of "error-prone" models (generated with incomplete inventories) with their references (generated with unmodified inventories). Mixed-effects logistic regression was tested for its ability to diminish potential bias-effects.
The inspection of 34 models highlighted that an inclusion of a bias-describing predictor (i.e. an environmental variable that closely describes an incompleteness) favors biased landslide susceptibility models and misleading (i.e. highest) predictive performances. However, an exclusion of such variables provoked confounded relationships because of omnipresent spatial correlations between environmental variables (e.g. land cover and topography). Both of these drawbacks (i.e. biases and confounding) were counteracted by including a bias-describing categorical variable as random effect within a mixed-effects logistic regression model. Among all models produced with substantially incomplete inventories, the newly introduced mixed-effects approach performed best to reproduce the reference data (i.e. prediction surfaces, spatial distribution of the unmodified response). Thus, the results underscored the efficiency of mixed-effects models in the presence of incomplete landslide inventories. We conclude that modelers should reject predictors that are directly associated with an inventory-based incompleteness, even if estimated predictive performances or automated variable selections might suggest the opposite. If systematically missing landslide information can be attributed to a spatial variable (e.g. administrative units, mapping domains, land cover) mixed-effects modelling might be the right choice.