Geophysical Research Abstracts
Vol. 12, EGU2010-10786, 2010
EGU General Assembly 2010

Pitfalls in statistical landslide susceptibility modelling

Boris Schröder (1,2), Peter Vorpahl (1), Michael Märker (3), and Helmut Elsenbeer (1)

(1) Institute of Earth and Environmental Sciences, University of Potsdam, Potsdam, Germany (boris.schroeder@uni-potsdam.de), (2) Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany, (3) Heidelberger Akademie der Wissenschaften, c/o Geographisches Institut der Eberhard Karls Universität Tübingen, Tübingen, Germany

The use of statistical methods is a well-established approach to predict landslide occurrence probabilities and to assess landslide susceptibility. This is achieved by applying statistical methods relating historical landslide inventories to topographic indices as predictor variables.

In our contribution, we compare several new and powerful methods developed in machine learning and well-established in landscape ecology and macroecology for predicting the distribution of shallow landslides in tropical mountain rainforests in southern Ecuador (among others: boosted regression trees, multivariate adaptive regression splines, maximum entropy). Although these methods are powerful, we think it is necessary to follow a basic set of guidelines to avoid some pitfalls regarding data sampling, predictor selection, and model quality assessment, especially if a comparison of different models is contemplated. We therefore suggest to apply a novel toolbox to evaluate approaches to the statistical modelling of landslide susceptibility. Additionally, we propose some methods to open the "black box" as an inherent part of machine learning methods in order to achieve further explanatory insights into preparatory factors that control landslides.

Sampling of training data should be guided by hypotheses regarding processes that lead to slope failure taking into account their respective spatial scales. This approach leads to the selection of a set of candidate predictor variables considered on adequate spatial scales. This set should be checked for multicollinearity in order to facilitate model response curve interpretation.

Model quality assesses how well a model is able to reproduce independent observations of its response variable. This includes criteria to evaluate different aspects of model performance, i.e. model discrimination, model calibration, and model refinement. In order to assess a possible violation of the assumption of independency in the training samples or a possible lack of explanatory information in the chosen set of predictor variables, the model residuals need to be checked for spatial auto¬correlation. Therefore, we calculate spline correlograms. In addition to this, we investigate partial dependency plots and bivariate interactions plots considering possible interactions between predictors to improve model interpretation.

Aiming at presenting this toolbox for model quality assessment, we investigate the influence of strategies in the construction of training datasets for statistical models on model quality.