Geophysical Research Abstracts,
Vol. 10, EGU2008-A-06367, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-06367
EGU General Assembly 2008

GIS-based landslide susceptibility mapping with comparisons of results from machine learning methods process versus logistic regression in Bailongjiang river basin, China

S. B. Bai (1), J. Wang (1), G. N. Lu (1), M. Kanevski (2), A. Pozdnoukhov (2)

1. National Education Administration Key Laboratory of Virtual Geographic Environments, Nanjing Normal University, Nanjing, China (Shibiaobai21@163.com; jwang169@vip.sina.com; gnlu@njnu.edu.cn)
2. Institute of Geomatics and Analysis of Risk, University of Lausanne, Amphipole, , Lausanne, Switzerland (Mikhail.Kanevski@unil.ch; Alexei.Pozdnoukhov@unil.ch)

Landslide susceptibility map is one of the study fields portraying the spatial distribution of future slope failure susceptibility. The purpose of this study is to evaluate and to compare the results of Machine Learning Methods and logistic regression model for basin scale landslide susceptibility mapping based on GIS. The logistic regression (LR) approach is further elaborated on by crosstabs method, which is used to analyze the relationship between the categorical or binary response variable and one or more continuous or categorical or binary explanatory variables derived from samples. It is an objective assignment of coefficients serving as weights of various factors under considerations while expert opinions make great difference in heuristic approaches. Different from deterministic approach, it is very applicable to regional scale. Nowdays the data-driven models are becoming more and more important. Particularly, machine learning methods provide promising perspectives in the landslide susceptibility mapping, being well-suited to non-linear high dimensional data modeling problems.

This work describes the application of LR and the Support Vector Machines (SVM) for landslide susceptibility mapping and validation in a 1361 km^2 study area mainly

on the Bailongjiang river basin, in northwest China. The site includes 6 subbasins and the landslide area covers an area of 19.67 km^2. A detailed landslide inventory map of the study area was identified by interpreting of 1:20,000-scale color aerial photographs and extensive field surveys. Four data domains are used in this study: remote sensing products, thematic maps, geological maps, and topographical maps. The size of pixels for all of the data layers was 30 m$\times$30 m. The study area was subdivided into calibration area (3 subbasins of the study zone, 924 km^2) and validation area (2 subbasins of the study zone, 437 km^2).

The factors chosen that influence landslide occurrence were: elevation, slope, aspect, curvature, distance to river, geological formation, distance to fault, distance to road, distance to settlement, land cover, vegetation index and precipitation distribution. This work was conducted among parameters, not classes of parameters. After exluded of the highly correlated dependent variables, stepwise logistic regression was carried out in SPSS in order to incorporate only the predicator variables with an important contribution to the presence of landslides. For the calibration area, significant model results were obtained, with geological formation, aspect, land cover, the summer precipitation, proximity to river and proximity to fault are main factors triggering landslide occurrence in this area.. Receiver Operating Characteristic (ROC) curves and the Kappa index were used to validate the model. Both show a good agreement between the observed and predicted values of the validation dataset.

The susceptibility maps produced by these two methods were validated and compared. Based on a qualified judgement, the created landslide susceptibility map based on the two method are classified into four classes, i.e., very high, high, moderate and low susceptibility. And the comparisons of the classified result susceptibility map and validation area landslide contribution were also processed.