Sample Position Affect Landslide Susceptibility Models in Hotspot Area of Nam Ma Basin, Lai Chau, Viet Nam

Van Trung Chu, Shou-Hao Chiang, and Tang-Huang Lin
National central university, Center for space and remote sensing research, Taiwan (chuvantrung@tuaf.edu.vn)

The aim of this study is to analyze the effect of landslide sample position with point-based approaches for landslide susceptibility modeling which were conducted in the hotspot of the landslide area located downstream of Nam Ma watershed (Sin Ho, Lai Chau, Viet Nam). Seven hundred fifty-nine landslide polygons that occurred in 2018 were mapped by using google earth integrated with field survey and 84 landslide points extracted from the inventory map conducted in 2013. The state-of-the-art sampling techniques and sample partition approach were applied to produce three subsets of training and testing point-based. Such as the highest position point within landslide polygon (SUB1), the centroid of landslide polygon (SUB2) and the point at the highest position within the seed cell area of the landslide polygon (SUB3). Along with that, the optimal strategy in selecting non-landslide samples was also applied and was first explicitly introduced in this study. Besides, multiple landslide conditioning factors were considered including topographic, geomorphological and hydrological groups. Especially beside of commonly used factors such as slope, elevation, curvature, land use land cover, aspect, etc. the unusual variables also considered such as high above the nearest drainage (HAND - the state-of-the-art terrain) or time series disturbance of land surface index was the first use in this study for landslide analysis and other cutting-edge data processing were proposed in this research arming to optimize the most vital part of whole procedure. The next stage of the analysis is landslide susceptibility modeling. In order to have a more objective judgment about the main issue mentioned above, instead of using only one model, we applied three different models namely Random forest (RF), Logistic regression (LR) and Decision tree (DT) to perform three kinds of scenarios by difference subsets of landslides with five folds of training phase. Subsequently, to compare the abilities of those cases, the model performance was assessed by using the area under the receiver operating characteristic curve both in model success rate (AUCSR) and model predictive rate (AUCPR). Finally, based on the results of this study, all three models performed consistent with three scenarios means the SUB2 and SUB3 are quite similar and much higher than the contribution of SUB1. And the model ability analysis indicated that RF can obtain higher accuracy following by LR and the lowest is DT.

Keywords: Sample position, Landslide Susceptibility, Logistic regression, Random forest, Decision tree, Viet Nam.