GEOMORPHOMETRY 2020 PERUGIA, ITALY JUNE 22 - 26 2020

Towards a consistent set of land-surface variables for landslide modelling

Andrei Dornik
Department of Geography, West University of Timisoara, Bd. V. Parvan 4, 300223, Timisoara, Romania

Lucian Drăguț
Department of Geography, West University of Timisoara, Bd. V. Parvan 4, 300223, Timisoara, Romania

Marinela Adriana Chețan
Department of Geography, West University of Timisoara, Bd. V. Parvan 4, 300223, Timisoara, Romania

Takashi Oguchi
Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8568, Japan

Yuichi Hayakawa
Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8568, Japan
Faculty of Environmental Earth Science, Hokkaido University, N10W5 Kita-ku, Sapporo, Hokkaido, 060-0810, Japan

Mihai Micu
Institute of Geography, Romanian Academy, 023993, Bucharest, Romania

Abstract—**This study aims at identifying a set of land-surface variables (LSVs) that enable consistent results in landslide modelling, within the context of automatic landslide mapping. The experiments were conducted in six study areas in Japan, Romania and USA. From an initial set of 24 LSVs, the most consistent predictors of landslide scarps in all study areas were selected through correlation analysis and variable importance investigation: negative topographic openness (radius 1500 m), slope height, and slope gradient. These three LSVs were further employed to model scarps' presence/absence with logistic regression. The results were compared against logistic regression models built on: 1) the best combination of locally fit LSVs (variable number of predictors, ranging from two to five across study areas), which was determined with backward stepwise logistic regression, and 2) a number of six variables reported in literature to best describe terrain properties. The predictive performance of the model built on the three LSVs came close to 1) and exceeded it in two cases, and outperformed 2), except for two cases. We conclude that negative topographic openness, slope height, and slope, which account for scarps shape, position on the slope, and landsliding favorability respectively, have a potential of generalization across landscape conditions in the prediction of scarps presence/absence.**

I. Introduction

Data-driven landslide modelling relies mainly on land-surface variables (LSVs) for automatic mapping, as well as for susceptibility assessment [1]. While LSVs are easy to obtain form Digital Elevation Models (DEMs), a consistent approach in selecting the ones that are the most relevant to landslides is still missing [2]. This lack of consistency makes the results of modelling dependent on the skills and experience of the analyst, thus hampering the comparison between models [3] and preventing their transferability to other areas [4].

Here we report preliminary results of an experiment that aims at finding a set of LSVs capable to help in identifying landslide scarps in various landscape conditions.

II. Study areas

The tests were conducted in six study areas, of different environmental conditions. Three study areas are located in the Buzău county, Romania, at the contact between the Romanian Carpathians and the Subcarpathian Hills, covering 121.7 km^2 (B1), 261.6 km^2 (B2) and 85.1 km^2 (B3) respectively. The annual mean temperature in these areas varies between 4-9°C and the total annual precipitation between 800-1200 mm. These areas are prone to numerous landslides, caused on the one hand by a clay rich substrate and high amount of precipitation, and on the other hand by their location in one of the European seismic hotspots.

Two study areas are located in the southeast of Honshu Island, Shizuoka Prefecture, Japan. The areas are humid and temperate with annual rainfall of approximately 2100 mm and annual mean temperature of 15°C. They are located in a tectonic active zone dominated by medium and high slopes, recording numerous landslides. One area covers 35.9 km^2 (J1) and the second one, 82.5 km^2 (J2).

The sixth study area (U), located in Utah, USA, has a relatively homogeneous lithology dominated by mixed-clastic and limestone deposits and covers an area of 299.2 km^2. The area receives on average 560-600 mm of annual precipitation that falls primarily as snow, while temperatures typically range from -11 to 27°C.

Andrei Dornik, Lucian Dragut, Marinela Adriana Chetan, Takashi Oguchi, Yuichi Hayakawa and Mihai Micu (2020) Towards a consistent set of land-surface variables for landslide modelling: in Massimiliano Alvioli, Ivan Marchesini, Laura Melelli & Peter Guth, eds., Proceedings of the Geomorphometry 2020 Conference, doi:10.30437/GEOMORPHOMETRY2020_44.

III. METHODS

A. Data

In the Buzău study areas there are available databases of landslide scarps compiled from different sources, such as archive data, detailed geomorphological field mapping, local authority databases, digital stereographic photo interpretation using color aerial ortho-photographs [5].

For the Japanese study areas we used inventories of landslide scarps provided by the National Research Institute for Earth Science and Disaster Resilience, Japan (NIED) [6-8].

In Utah we used the Landslide Inventory Polygons developed by professional scientists from Utah Department of Natural Resources and Utah Geological Survey. The polygons were built using LiDAR, stereo aerial photography, other data, and field reconnaissance.

In order to avoid differences in modelling results caused by differences in sampling intensity, the existing databases were reduced to contain approximately the same number of scarps. One point was randomly selected within each landslide scarp, and the same number of points was also randomly selected outside scarp polygons as absence data. The databases for each study area, containing between 204 and 220 points, were split into training (70%) and test (30%) subsets.

Besides landslide inventories, Shuttle Radar Topography Mission (SRTM) 1 arc second (ca. 30 m) digital elevation model (DEM) was available at the US Geological Survey website and was downloaded from the USGS Earth Explorer Interface. The DEM resolution is suitable for this analysis, since the average size of scarps is 14.9, 9.96 and 4.18 ha in B1, B2, B3, respectively, 2.26 ha in J1, 2.58 ha in J2 and 3.24 ha in U.

B. Land-surface variables

A number of 24 LSVs, which have been predominantly used in landslide modelling, were derived from DEM. To avoid multicollinearity of variables we conducted a correlation analysis, to retain only the non-correlated terrain variables. The threshold for variable removal was set to 0.6 for correlation coefficient.

Subsequent processing was conducted in order to ensure that all terrain variables have a normal distribution [9].

C. Variable importance analysis

In order to select a generalizable subset of terrain variables for landslide modelling, the uncorrelated variables were subjected to variable importance (VI) analysis using Random Forest (RF) [10]. VI analysis was conducted with randomForest package in R [11] and was based on the mean decrease in accuracy (MDA) algorithm implemented by Liaw and Wiener [12].

D. Landslide modelling

The identified generalizable subset of variables, arising as important predictors in all study areas were used subsequently to conduct landslide modelling (named LSM_VI).

To assess the performance of the identified set of LSVs, we conducted analyses based on the best model specific to each study area, identified with backward stepwise logistic regression (named LSM_best_model).

In addition, we tested the set of generalizable LSVs against six terrain variables (named LSM_tasse), proposed by Lecours et al. [13] to be used in environmental studies, as the six variables capture more than 70% of the topographic structure of an area. These variables are: relative difference to mean elevation value, standard deviation of DEM, easterness, northerness, local mean and slope.

Logistic regression was used for modelling and the results were evaluated both in terms of model fit using training data and model prediction performance using test data. The model fit was evaluated through the Akaike information criterion (AIC), and the prediction performance through overall accuracy and area under the receiver operating characteristic (ROC) curve (AUC) measures.

IV. RESULTS AND DISCUSSION

A. Correlation analysis

Correlation analysis revealed 14 non-correlated LSVs: elevation, slope, profile curvature, plan curvature, convexity, easterness, northerness, mid-slope position, negative topographic openness, positive topographic openness, slope length, slope height, texture, valley depth.

B. Selection of a generalizable set of LSVs

Fig. 1 shows the results of VI analysis in the six study areas. The first variable that showed as important predictor in all study areas is negative topographic openness (radius 1500 m), with MDA values higher than 10%. Slope height stood out also as important predictor in five out of six study areas, with MDA larger than 10%. In B3, slope height recorded a MDA value of 8.7, however being one of the most important five variables. Slope also recorded high MDA values, in U recording the highest value within all tests (38%), and over 10% in B2 and B3. Slope was also an important predictor in B1 with a MDA value of 8.1%. These three variables were thus selected as the generalizable set for LSM_VI model. We interpret negative openness as a generalized concavity that accounts for the shape of the landslide scarps; slope heights would describe the scarp position, and slope gradient is well known as the main topographic predisposing factor for mass movements.

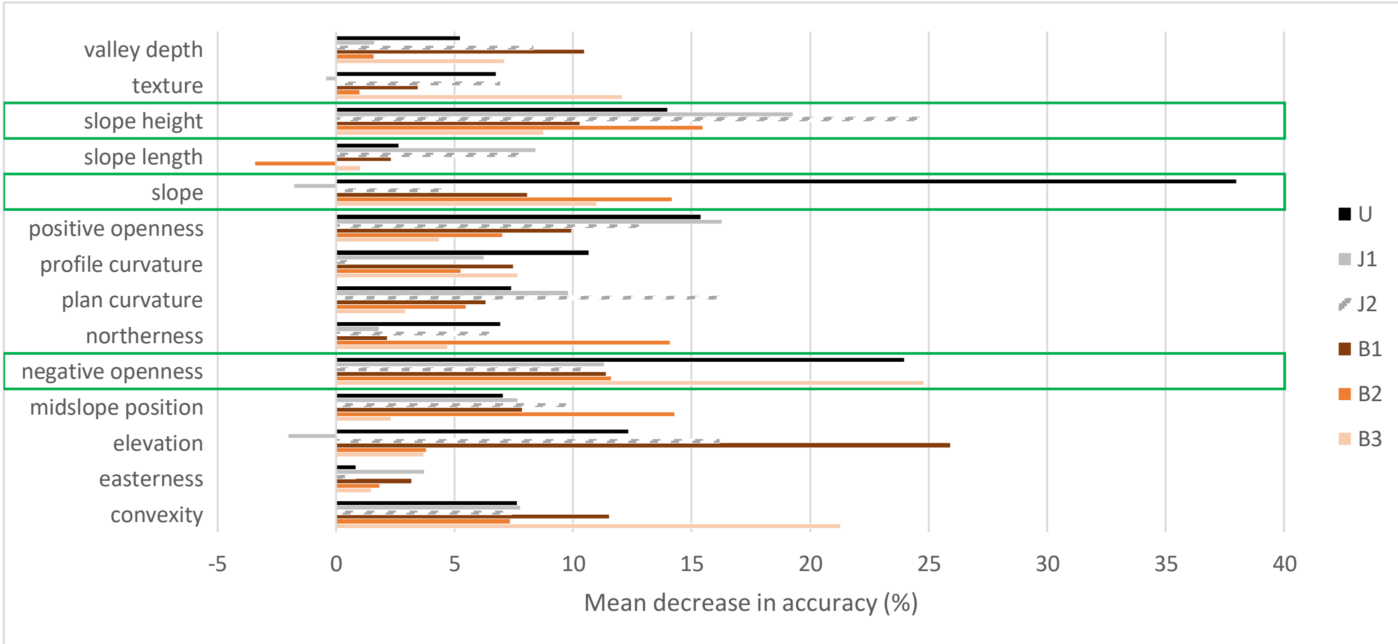

Figure 1. Variable importance expressed as mean decrease in accuracy in the six study areas.

Other variables were more location specific, highlighted as important predictors only in one to three study areas. For example, convexity, elevation and valley depth recorded MDA values higher than 10% in B1, midslope position and northerness in B2, positive openness in J1, elevation and plan curvature in J2, and profile curvature in U. Other terrain variables like easterness, slope length and texture were among the least important variables in all study areas.

C. Models evaluation

In terms of model fit measured by AIC, our proposed model was relatively similar to the best model, losing between only 7.4 in study area U and 28.8 in B3. The other study areas recorded AIC value larger by less than 20. Comparing our proposed model with LSM_tasse, in five out of six cases LSM_VI model was better with AIC value lower by 2.6, 4.6, 9.1, 13.7 and 14.2. In study area B2, LSM_tasse model was better than our model, by 12.4 (Table 1).

Overall accuracy of LSM_VI recorded values lower than the best model in four out of six cases, however by only 1-7%. In study areas J1 and B3 our proposed model performed better than the best model, by 2 and 12% respectively. In four out of six cases our proposed model overperformed LSM_tasse, recording overall accuracies higher by 3, 5, 11 and 15%. In study area U the models had the same accuracy, 70%. Only in one case, B1, LSM_tasse performed better than LSM_VI, by 12%. The absolute values of overall accuracy for LSM_best_model range between 52 and 74%, for LSM_VI between 56-70%, and for LSM_tasse between 53-70% (Table 1).

In terms of AUC, the best model performed better than our proposed model only in two cases (J2 and B1), by 0.01 and 0.1, in other two cases the models performing identical (B2 and B3), with an AUC of 0.65. In J1 and U, LSM_VI performed better than the best model, by 0.09 and 0.03. Compared to LSM_tasse, our proposed model recorded higher AUC values in five out of six cases, with differences ranging between 0.01 and 0.16. Only in study area B1, LSM_tasse recorded an AUC higher than LSM_VI, by 0.12. The absolute values of AUC ranged between 0.63-0.82 for LSM_best_model, 0.63-0.83 for LSM_VI and 0.63-0.81 for LSM_tasse (Table 1).

The results should be interpreted also regarding the number of variables, being well known that between two similar models, it should be preferred the simpler one. While the LSM_VI used three terrain variables and LSM_tasse used six variables, the best model used various numbers of variables, ranging from 2 in J1 to 5 in B2.

Table 1. Models fit and prediction performance assessment.

Study area		AIC			Overall accuracy			AUC		
	Var. no. best model	LSM_best_ model	LSM_VI	LSM_ tasse	LSM_best_ model	LSM_VI	LSM_ tasse	LSM_best_ model	LSM_VI	LSM_ tasse
J1	2	186.70	197.30	206.40	0.52	0.64	0.53	0.63	0.72	0.64
J2	4	178.60	188.10	202.30	0.74	0.70	0.55	0.82	0.81	0.65
B1	3	167.80	183.90	188.50	0.63	0.56	0.68	0.73	0.63	0.75
B2	5	168.30	185.30	172.90	0.64	0.59	0.56	0.65	0.65	0.63
B3	4	158.70	187.50	201.20	0.62	0.64	0.59	0.65	0.65	0.64
U	4	147.80	155.20	157.80	0.71	0.70	0.70	0.80	0.83	0.81

V. CONCLUSIONS

We found three LSVs with the potential of describing satisfactorily landform scarps in various landscape conditions. Negative topographic openness, slope height, and slope account for scarps shape, position on the slope, and landslide favorability respectively.

The logistic regression model based on these three LSVs produced results comparable to models built on locally calibrated LSVs, as well as to a model built on a double number of LSVs.

REFERENCES

[1] Reichenbach, P., et al., *A review of statistically-based landslide susceptibility models.* Earth-Science Reviews, 2018. **180**: p. 60-91.

[2] Budimir, M., P. Atkinson, and H. Lewis, *A systematic review of landslide probability mapping using logistic regression.* Landslides, 2015. **12**(3): p. 419-436.

[3] Guzzetti, F., et al., *Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy.* Geomorphology, 1999. **31**(1-4): p. 181-216.

[4] Cama, M., et al., *Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy).* Geomorphology, 2017. **288**: p. 52-65.

[5] Sîrbu, F., et al., *Scaling land-surface variables for landslide detection.* Progress in Earth and Planetary Science, 2019. **6**(1): p. 44.

[6] Shimizu, F., et al., *Landslide maps series 14 "Shizuoka".* Technical Note of the National Research Institute for Earth Science and Disaster Resilience, p 221, 2002.

[7] Uchiyama, S., et al., *The NIED landslide mapping team.* 2012.

[8] Oyagi, N., S. Uchiyama, and M. Ogura, *Explanations of landslide distribution maps.* Technical note of the National Research Institute for Earth Science and Disaster Resilience, 2015. **vol 394, pp 391-314**.

[9] Csillik, O., I. Evans, and L. Drăguț, *Transformation (normalization) of slope gradient and surface curvatures, automated for statistical analyses from DEMs.* Geomorphology, 2015. **232**: p. 65-77.

[10] Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.

[11] R Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* 2015: p. Vienna, Austria, http://www.R-project.org/.

[12] Liaw, A. and M. Wiener, *Classification and regression by randomForest.* R news, 2002. **2**(3): p. 18-22.

[13] Lecours, V., et al., *Towards a framework for terrain attribute selection in environmental studies.* Environmental Modelling & Software, 2017. **89**: p. 19-30.