Sensitivity of land-surface variables to scale in identifying landslide scarps

Flavius Sirbu
Department of Geography
West University of Timisoara
Timisoara, Romania
flavius.sirbu@e-uvt.ro

Lucian Dragut
Department of Geography
West University of Timisoara
Timisoara, Romania
lucian.dragut@fulbrightmail.org

Takashi Oguchi
Center for Spatial Information Science
The University of Tokyo
Tokyo, Japan
oguchi@csis.u-tokyo.ac.jp

Yuichi Hayakawa
Center for Spatial Information Science
The University of Tokyo
Tokyo, Japan
hayakawa@csis.u-tokyo.ac.jp

Mihai Micu
Institute of Geography
Romanian Academy
Bucharest, Romania
mikkutu@yahoo.com

Abstract— **Geomorphological features, in general, and landslides, in particular, come on different sizes and shapes. However, the land-surface variables (LSVs) that are used to describe them are often employed at a unique scale, as resulted from derivation in a standard 3×3 moving window. In this paper we test the hypothesis that identification of landslide scarps improves when individual LSVs are calibrated to scale. For this purpose, we set up an experimental design, which was run in two topographically different locations: Shizuoka Prefecture, Japan and Buzau County, Romania. The experiment includes two steps: i) finding the scale at which each LSV achieves the best prediction of the scarps, and ii) comparing the modeling results of the scaled LSVs against LSVs at the default scale. In the first step, each LSV is up-scaled using focal mean statistics in a steadily increasing moving window, which starts at 3×3 and grows until the respective LSV achieves the maximum degree of fitting with the scarps. The degree of fitting is determined with logistic regression. In the second step, the LSVs at the calibrated scales are used as the input data for a random forest (RF) model. The RF model was ran with the same settings on the default LSVs (i.e. derived in a 3×3 window) as well, for comparison. The results show a consistent increase in the AUC score when LSVs where calibrated to scale, as compared to the modeling with the default LSVs, from 80.31 to 91.54 in the first study area, and from 74.62 to 83.11 in the second one. The results also show that each predictor operates at its specific scale; therefore a single 'optimal' scale does not apply for all LSVs.**

I. INTRODUCTION

The main issue in modeling landslides (identification, mapping, susceptibility and hazard evaluation) is related to their large diversity in shape, frequency and magnitude, due to predisposing (e.g. topographic conditions), preparing (ex. land cover changes) and triggering factors (e.g. precipitation or earthquakes). In order to address this issue, the input data in the models have to be adaptable enough to work on different types of landslides. It is acknowledged that serious obstacles on the way to perform/obtain better models are: the insufficient pre-processing of the input variables in respect to the sampling strategy, the spatial distribution, size and split of the data, and scale of the predictors [1].

The existing work on evaluating the impact of scale of the predictors has been carried out for modeling landslide susceptibility. It has been previously shown that the scale of the analysis should be chosen based on the average size of the landslides [2], while [1] found that scaling the predictors improves

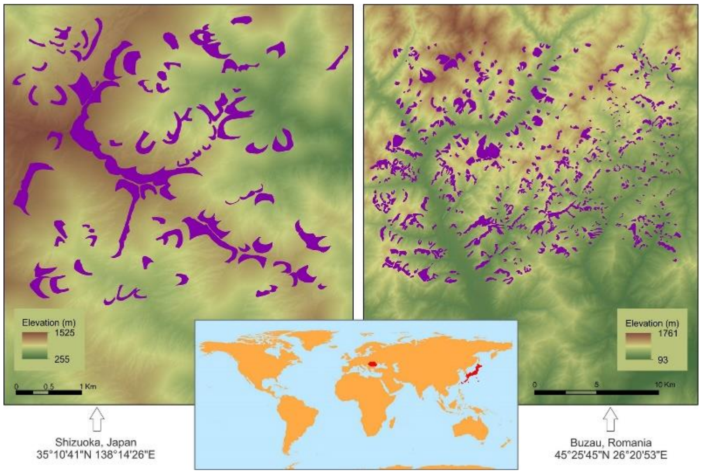

Figure 1. The two landslide scarps inventories and their location

the global models of susceptibility mapping. Reference [3] further showed that predictors at different optimum scales perform better than a single-scale modeling. While the calibration of the predictors to scale is expected to improve landslide identification as well, this has not been proved yet.

In the present study, we test if, and to which degree, calibration to scale of the individual predictors improves the results of identifying landslide scarps procedures.

II. METHODOLOGY

A. Study area

The first study (Figure 1) area is located in the Shizuoka Prefecture, in the SE of Honshu Island, Japan. The existing data consists of a LiDAR – based DEM at 5 m spatial resolution, and an inventory of 117 landslide scarps. The scarps in this area are mostly covered by forest.

The second study area (Figure 1) is located in the Buzau County, Romania. It stretches over two distinct relief and geological units, the Carpathian Mountains and the Subcarpathian hills. The existing data consist of a DEM with a spatial resolution of 4 m, and an inventory of 630 scarps.

The landslide scarps from the two study areas show significant differences in size (Figure 2). The average area of the landslide scarp in the Buzau County study site is 8.64ha and the median size is 4.28ha, while the mean size of the landslide scarps in Shizuoka Prefecture study site is 2.35ha and the median value is 1.44ha

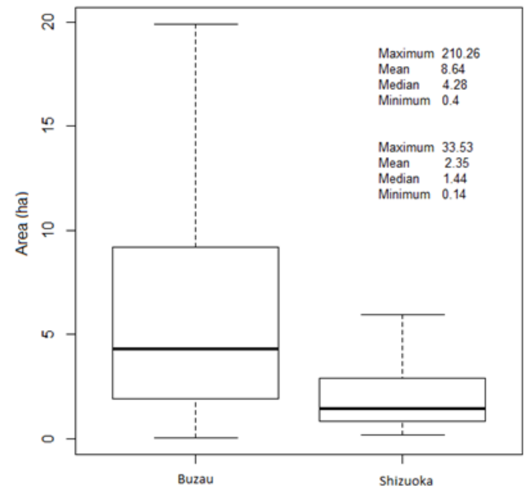

Figure 2. The area of landslide scarps in the two study sites. Boxplots represent 25-75% of values, whiskers represent 10-90% of values and the line within the box represents the median. Extreme values were omitted.

B. Predictors

For this test we use 13 land-surface variables (LSVs) as predictors (TABLE 1), which past studies have found useful in landslide modeling [4], [5] and [6]. The LSVs are extracted from the available DEMs using the System for Automated Geoscientific Analyses – SAGA [7]. The values of these predictors are associated with one randomly sampled point per scarp, as well as an equivalent number of randomly sampled non-scarp points.

C. Scaling the LSVs

In order to find the optimal scale for modeling, each LSV is re-scaled to successively broader representations of topography with focal mean statistics in increasing windows, starting from 3×3 [8]. A simple binary logistic regression is computed [4], where each LSV predicts scarp presence/absence as dependent variable. At each scale, the goodness of fit is evaluated using AUC (area under the curve). The process stops when the first highest AUC value is found.

D. Modelling the landslide scarps

The modeling is performed using the package "randomForest" [9] in the R software [10], with the settings ntree = 501 and mtry = 3 (number of trees and number of candidate variables, respectively). The results of 25 runs are evaluated using the AUC. Random forest (RF) is preferred because it is relatively easy to use, computationally efficient and produces models with a high level of accuracy.

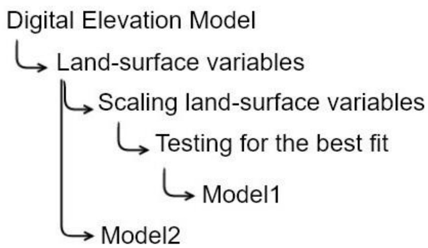

Figure 3. A schematic representation of the methodology

In order to assess the impact of scaling the predictors in modeling, two models are produced: one that uses the scaled LSVs as input data, while the other uses the default LSVs, for comparison (Figure 3). The latter is considered the benchmark in evaluation of the scaling performance.

III. RESULTS

A. Scaling the LSVs

Ten out of the thirteen LSVs were scaled, and three (altitude, total catchment area and topographic wetness index) were only employed in modelling the landslide scarps as they are known to be insensitive to scale.

TABLE 1. CALIBRATED SCALES OF LAND SURFACE VARIABLES

Land-surface variables	**Shizuoka**	**Buzau**
Altitude	**default - 0.642**[a]	**default - 0.599**
Aspect	**default - 0.606**	**5×5 - 0.62** (0.513)
Convergence index	**9×9 - 0.567** (0.542) [b]	**9×9 - 0.614** (0.506)
Curvature plan	**9×9 - 0.567** (0.542)	**5×5 - 0.608** (0.539)
Curvature profile	**9×9 - 0.528** (0.513)	**9×9 - 0.618** (0.513)
Curvature	**7×7 - 0.59 (**0.513)	**9×9 - 0.615** (0.522)
Flow direction	**default - 0.543**	**5×5 - 0.574** (0.545)
Hillshade	**5×5 - 0.541** (0.531)	**5×5 - 0.588** (0.568)
LS factor	**5×5 - 0.6** (0.541)	**default - 0.565**
Slope	**5×5 - 0.663** (0.525)	**5×5 - 0.674** (0.524)
Surface roughness	**default - 0.728**	**5×5 - 0.574** (0.548)
Total catchment area	**default - 0.582**	**default - 0.53**
Topographic wetness index	**default - 0.557**	**default - 0.558**

[a] The table shows the scales (window size to compute focal mean statistics) at which individual LSVs maximize the prediction of landslide scarps and their AUC values. [b] AUC values of the LSVs at default scale are given within the brackets for comparison.

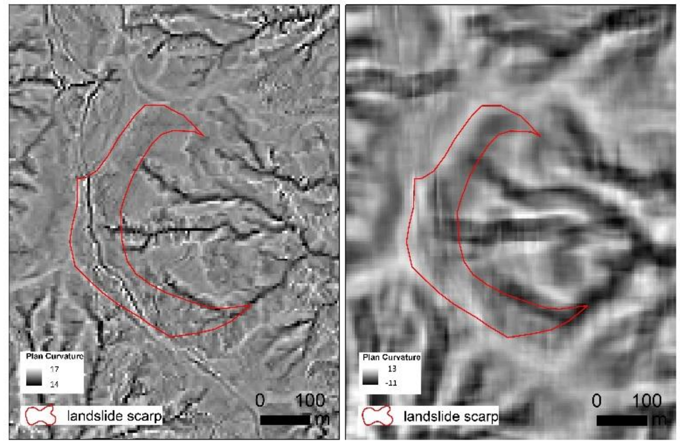

Figure 4. An example of a scaled predictor in relation to a landslide scarp. Left image: plan curvature at the default scale. Right image: plan curvature scaled in a 9×9 window. Note a north-south trench inside the landslide scarp on the left image that is filtered out on the right image.

LSVs calibrated to scale predict better the scarp presence/absence in seven out of ten cases for the first study area, and in nine out of ten cases for the second study area, respectively. The size of the moving window at which LSVs were rescaled vary between 5×5 and 9×9 for both study sites, and four out of ten predictors were found to perform best at the same scale in both sites (TABLE 1). This reinforces the hypothesis that there is no universal scale that works for all the LSVs, but rather each LSV performs best at its own scale, confirming the similar results found by [3].

However, more work is needed to research by how much the scale sensitivity analysis of individual LSVs outperforms the scale sensitivity on global models with all the LSVs at the same scale and if the results will justify the effort of performing another computational step.

Considering that the scales of the LSVs from the two study areas are relatively similar while the size of the landslides scarps differs significantly, it can be suggested that the scale of the LSVs is not straightforwardly related to the size of the landslide scarps. The improvement of modeling seems to be rather due to a better representation of topography through elimination of local noise (micro-topography) (Figure 3).

B. Modelling the landslide scarps

The results show an improvement in the overall AUC score of the modeling, when using the scaled input data, from 80.31 to 91.54 for the first study area and from 74.62 to 83.11 for the second one.

IV. CONCLUSIONS

The results show, in both study areas, that calibration of LSVs to scale improves modeling of landslide scarps. These results confirm the previous findings of [1] and [3] that the scale of the input predictors is an important factor and should be considered when modeling landslides. This work is relevant to increase the accuracy of landslide scarp identification, within the context of automating the inventory mapping.

ACKNOWLEDGMENT

This work was supported by a grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P4-ID-PCE-2016-0222, within PNCDI III.

REFERENCES

[1] Catani, F., Lagomarsino, D., Segoni, S., Tofani, V., 2013. “Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues”, Natural Hazards and Earth System Sciences, 13(11), 2815-2831.

[2] Claessens, L., Heuvelink, G.B.M., Schoorl, J.M., Veldkamp, A., 2005. “DEM resolution effects on shallow landslide hazard and soil redistribution modelling”, Earth Surface Processes and Landforms, 30(4), 461-477

[3] Paudel, U., Oguchi, T. and Hayakawa, Y., 2016. “Multi-Resolution Landslide Susceptibility Analysis Using a DEM and Random Forest”. International Journal of Geosciences, **7**, 726-743.

[4] Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015. “Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling”, Computers and Geosciences, 81, 1-11.

[5] Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., Bălteanu, D., 2016. „Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model”, Geomorphology, 253, 508-523.

[6] Pradhan, A.M.S., Kim, Y.T., 2016. “Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping”, Catena, 140, 125-139.

[7] www.saga-gis.org

[8] Drăguţ, L., Schauppenlehner, T., Muhar, A., Strobl, J., Blaschke, T., 2009. “Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling”, Computers & Geosciences, 35(9), 1875-1883.

[9] Liaw, A. and Wiener, M., 2002. “Classification and Regression by randomForest”. R News 2(3), 18-22.

[10] R Core Team, 2016. “R: A language and environment for statistical computing”. R Foundation for Statistical Computing, Vienna, Austria.