Landslide topographic signature prediction using segmentation of roughness and Random Forest classification

Mihai Niculiță[§]
Department of Geography
Alexandru Ioan Cuza University of Iași
Caro lI 20A, 700505 Iași, Romania
[§] mihai.niculita@uaic.ro

Abstract—**Landslides creates a typical rough topography which is clearly distinguishable from other types of roughness created by other geomorphological processes or agents. Starting from the idea of geometric signature of topography induced by landslide processes I fed geomorphological objects obtained from segmentation of roughness and their geomorphometry to a machine learning algorithm in order to be able to predict landslide topography presence. The results are good in terms of overall accuracy (0.72 sensitivity and 0.06 false positive rate), but further study is needed to improve the results and to test it in other physiographic settings. This approach of landslide presence prediction is of crucial importance for landslide hazard research, because landslides usually appear in areas previously affected by the process, and this information can be used in landslide susceptibility assessment.**

I. INTRODUCTION

Landslide induced topography exhibit a clear geomorphometric signature [1-3]. Landslide roughness is different from fluvial roughness mainly in scale. The landslide roughness is also defined by more smoothness than pure fluvial roughness. Geomorphometrical objects [4] delineated based on roughness delineate landform facets/segments that exhibit elevation deviation both vertically and horizontally. The target is to use statistical information on the geomorphometry of these segments, in order to predict their landslide/non-landslide status. The main idea is that landslide topography is rough and with curvature different than non-landslide topography, and this morphometric signature can be targeted using high-resolution Digital Elevation Models (DEMs) and Random Forest (RF).

The areas previously affected by landslides are more susceptible to be affected later by landslides [5,6], and a layer with the spatial extension of these landslides can be used in landslide susceptibility assessment [7,8].

Two study areas from Moldavian Plateau were selected for testing the proposed approach (Fig. 1 and 2).

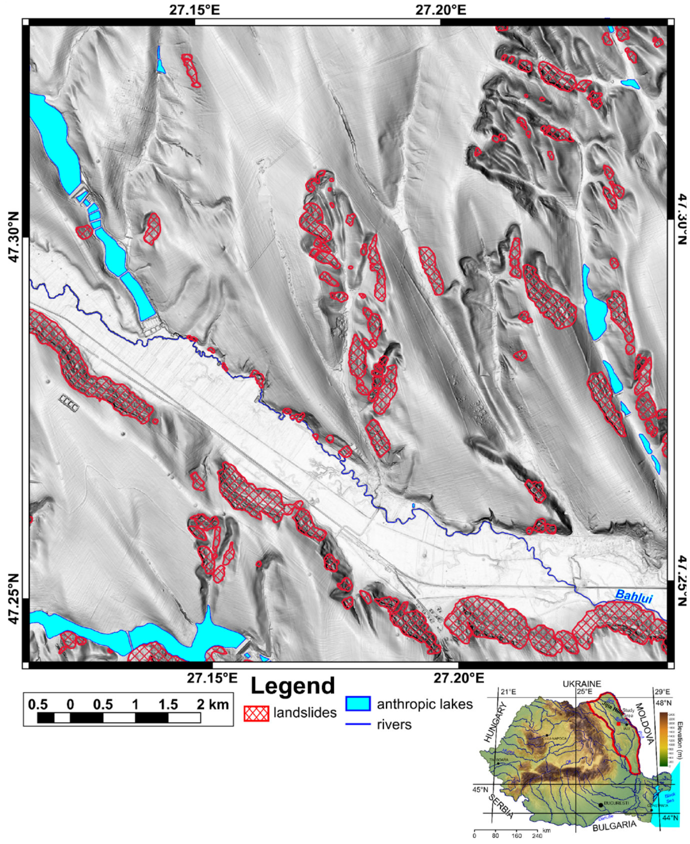

Figure 1. The northern study area and the landslide mapping (a high-resolution version is available at https://doi.org/10.6084/m9.figshare.12226673).

The two areas have a rectangular shape of 100 square km, and are in neighborhood of each other. In the northern study area, the landslide density (landslide surface proportion was computed from

Mihai Niculita (2020) Landslide topographic signature prediction using segmentation of roughness and Random Forest:

in Massimiliano Alvioli, Ivan Marchesini, Laura Melelli & Peter Guth, eds., Proceedings of the Geomorphometry 2020 Conference, doi:10.30437/GEOMORPHOMETRY2020_48.

the landslide inventory of [9]) is 7.53%, while in the southern study area the landslide density is 14,53%. Both relict, old and new landslides are present [10,11], with various typologies: flowslides, translational and rotational slides, and through landslide amalgamation compound and complex landslides appear mainly on the steep hillslopes of cuesta scarps [12,13].

II. Methods

A DEM with a spatial resolution of 5 m was obtained from a 0.5 m spatial resolution bare earth LiDAR through bicubic resampling in SAG GIS [14]. Landslides were delineated using 2D and 3D views of hillshading, slopes and contours. All the areas affected by landslides were delineated without separating different landslide events, that are visible in topography. Relict landslides (see [11] for details on how the relative age was estimated) were not included in the delineation, mainly because their roughness is smoothed by the agricultural works (tillage and terraces).

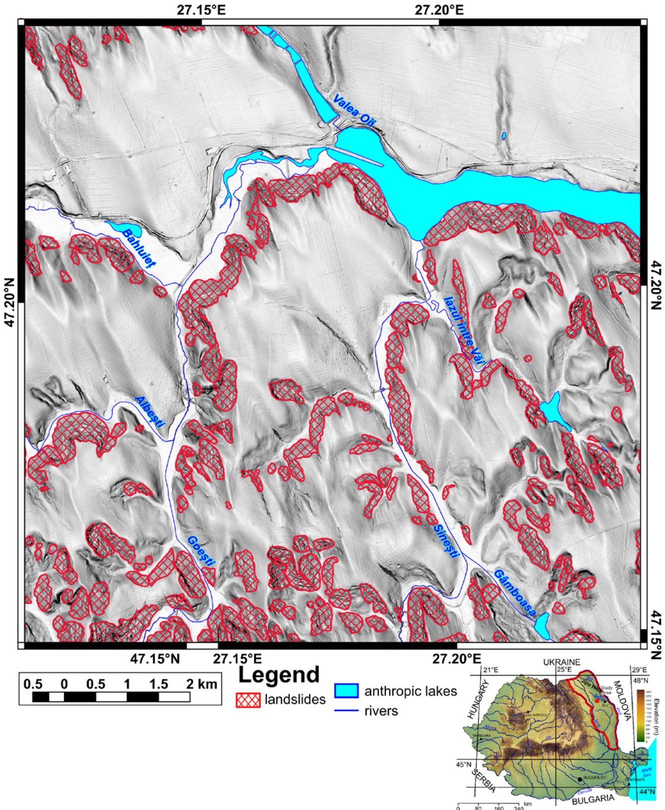

Figure 2. The southern study area and the landslide mapping (a high-resolution version is available at https://doi.org/10.6084/m9.figshare.12226838).

The landforms segments were obtained through watershed segmentation of the vectorial roughness measure [15] computed based on the DEM in SAGA GIS [14]. For every segment the descriptive statistics (minimum, maximum, mean, sum, standard deviation, multiples of 5 percentiles) of the following geomorphometric variables were generated: altitude, vectorial roughness measure [15], real area, terrain ruggedness index [16], texture [17], slope [18] and curvatures (profile, plan, cross-sectional, longitudinal, minimum, maximum) [19].

The Random Forest (RF) algorithm [20] implemented in R software [21] as randomForest package [22] was used for fitting the model. Latin hypercube sampling [23,24] was used to select the training dataset with the clhs package [25]. All the segments corresponding to the delineated landslides were selected in the training dataset while the latin hypercube sampling was used to select a double amount of non-landslide segments. The testing was performed on the southern study area dataset (external domain [26]). The RF parameters were tuned in order to derive the best model. An enough number of trees to grow (ntree=100), with a small number of variables randomly sampled as candidates at each split (mtry=3) and a small minimum size of terminal nodes (nodesize=1) are giving the best results. The performance of the models was evaluated in terms of confusion matrix, and not in terms of Out-of-bag (OBB) error. Another important aspect of tuning is the sampling and the class imbalance. The sampling was performed statistically (latin hypercube), but the class proportion is important and used together with priors of the class (classwt=c(0.1,0.9)) parameter of the RF model in order to deal with class imbalance.

III. Results and conclusions

The confusion matrix for the northern study area, where the model was fit and for the southern study area, where the model was evaluated are shown in Table I.

TABLE I. THE CONFUSION MATRIX AND ITS MEASURES FOR THE NORTHERN STUDY AREA (FIRST ROW) AND SOUTHERN STUDY AREA (SECOND ROW)

RF parameters	OBB error	TP	TN	FP	FN	SNS	FPR
10000 segments from which 946 landslides, 100 ntree, 5 mtry, 1 nodesize	5.23	9223	147615	2698	4471	0.67	0.018
		14428	137798	8528	5593	0.72	0.058

Sensitivity and false positive rate are shown in order to assess an overall view of the result, since accuracy is not necessarily the best measure.

The sampling was performed statistically (latin hypercube), and the class imbalance is consistently dealt with this approach: the sampling keeps the proportion of landslide vs. non-landslide segments, and the results are better with RF models that do not assign weights or priors for imbalanced classification. All the geomorphometric variables and their statistics (285) were used in the final model. In this way, all the geomorphometric information is used for the identification of landslide segments. The overall accuracy seems to be good enough, but the spatial results need to be investigated. It seems that especially the basal part of landslides is occupied by false negative segments and the crown by false positive segments (Fig. 3 and 4). False positive segments also appear on some relict landslides and in settlement areas, where there is a certain roughness related to be anthropic modifications of topography. Some unfiltered vegetation, and several gullies also are considered topography related to landslides by the model.

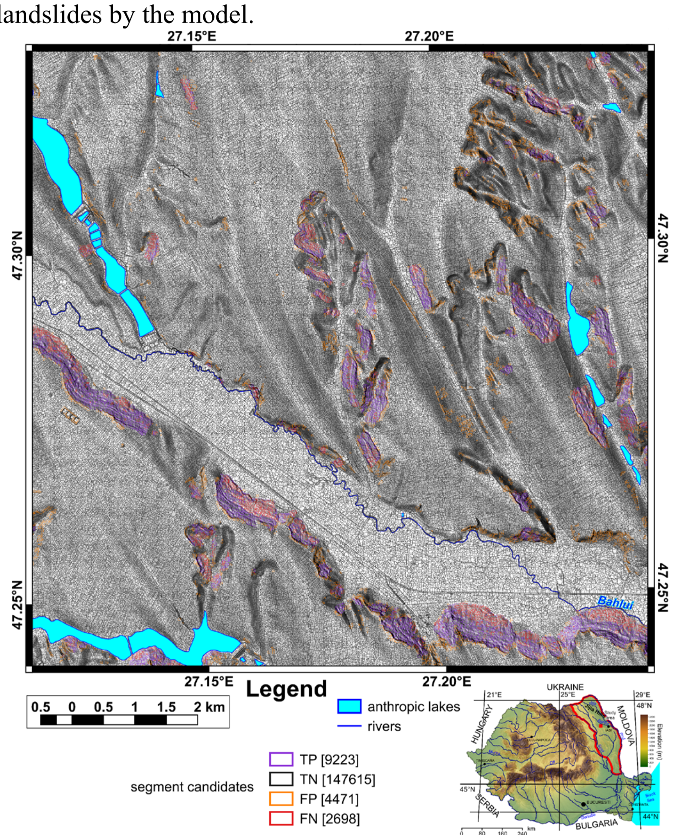

Figure 3. The northern study area and the predicted landslide segments. (a high-resolution version is available at https://doi.org/10.6084/m9.figshare.12226853).

IV. Conclusions

The landslide prediction based on roughness proved to be feasible in an object-based approach. The segments that are not predicted (the false negatives) are located mainly inside landslides, where the topography is not rough for various reasons (landslide mass deformation, agricultural works or lake presence). The segments that are predicted as landslides, but in fact are not (false positives), fall in three categories: i) settlements, ii) relict landslides, iii) river and gully incision. While the results can be considered satisfactory for now, future improvements are welcomed.

Acknowledgments

This work was supported by a grant of Ministery of Research and Innovation, CNCS – UEFISCDI, project number PN-III-P1-1.1-PD-2016-0154, within PNCDI III. I am grateful to Prut-Bârlad Water Administration who provided me with the LIDAR data.

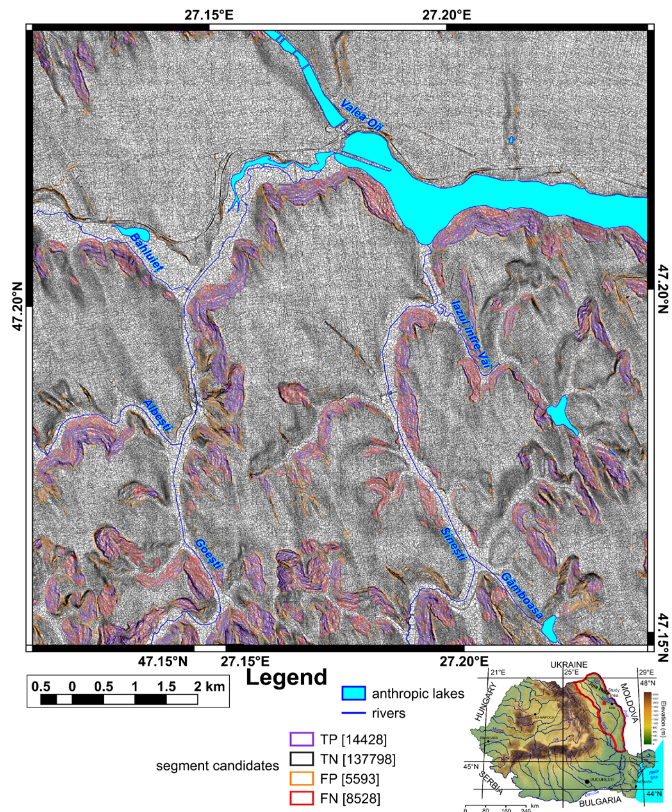

Figure 4. The southern study area and the predicted landslide segments. (a high-resolution version is available at https://doi.org/10.6084/m9.figshare.12226883).

REFERENCES

[1] Pike, R. J. 1988. "The geometric signature: Quantifying landslide-terrain types from digital elevation models". Mathematical Geology 20(5), 491-511.

[2] Marchesini, I., Rossi, M., Mondini, A. C., Santangelo, M., Bucci, F. 2014. "Morphometric signatures of landslides". In: Third Open Source Geospatial Research & Education Symposium, Espoo, Finland, 10-13 June 2014, pp 99-105.

[3] Rossi, M., Mondini, A. C., Marchesini, I., Santangelo, M., Bucci, F., Guzzetti, F. 2013. "Landslide morphometric signature". The 8th IAG/AIG International Conference on Geomorphology, Paris, France, 27-31 August, 2013.

[4] MacMillan, R. A., Shary, P. A. 2009. "Chapter 9 Landforms and Landform Elements in Geomorphometry". In: Geomorphometry - Concepts, Software, Applications. Developments in Soil Science. pp 227-254.

[5] Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Ardizzone, F., Rossi, M. 2016. "Do landslides follow landslides? Insights in path dependency from a multi-temporal landslide inventory". Landslides 14(2), 547-558.

[6] Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Ardizzone, F., Rossi, M. 2017. "Characterization and quantification of path dependency in landslide susceptibility". Geomorphology 292, 16-24.

[7] Samia, J., Temme, A., Bregt, A. K., Wallinga, J., Stuiver, J., Guzzetti, F., Ardizzone, F., Rossi, M. 2018. "Implementing landslide path dependency in landslide susceptibility modelling". Landslides 15(11), 2129-2144.

[8] Necula, N., Niculiță, M. 2017. "Landslide reactivation susceptibility modeling in Iași Municipality". Revista de geomorfologie 19, 101-117.

[9] Niculiță, M., Mărgărint, M. C. 2014. "Landslide inventory for Moldavian Plateau, Romania". In: International Conference Analysis and Management of Changing Risks for Natural Hazards Padua, Italy, 18-19 November 2014, AP3-1-9.

[10] Niculiță, M., Mărgărint, M. C., Cristea, A. I. 2019. "Using archaeological and geomorphological evidence for the establishment of a relative chronology and evolution pattern for Holocene landslides". PLoS One 14(12), e0227335.

[11] Niculiţă, M., Mărgărint, M. C., Santangelo, M. 2016. "Archaeological evidence for Holocene landslide activity in the Eastern Carpathian lowland". Quaternary International 415, 175-189.

[12] Mărgărint, M. C., Niculiţă, M. 2017. "Landslide Type and Pattern in Moldavian Plateau, NE Romania". In: Rădoane M, Vespremeanu-Stroe A (eds) Landform Dynamics and Evolution in Romania. Springer Geography, vol Springer Geography. Springer, Cham, pp 271-304.

[13] Niculiţă, M. 2011. "A landform classification schema for structural landforms of the Moldavian platform (Romania) ". In: Hengl T, Evans IS, Wilson JP, M. G (eds) Geomorphometry 2011, Redlands, USA, September 7-11, 2011. pp 129–132.

[14] Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J. 2015. "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4". Geoscientific Model Development 8(7), 1991-2007.

[15] Sappington, J. M., Longshore, K. M., Thompson, D. B. 2007. "Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert". Journal of Wildlife Management 71(5), 1419-1426.

[16] Riley, S. J., DeGloria, S. D., Elliot, R. 1999. "A terrain ruggedness index that quantifies topographic heterogenity". Intermountain Journal of Sciences 5(1-4), 23-27.

[17] Iwahashi, J., Pike, R. J. 2007. "Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature". Geomorphology 86(3-4), 409-440.

[18] Evans, I. S. 1979. "An integrated system of terrain analysis and slope mapping". Final report on grant DA-ERO-591-73-G0040. University of Durham, Durham, UK.

[19] Wood, J. 1996. "The geomorphological characterisation of digital elevation models. PhD Thesis, University of Leicester, Leicester, UK.

[20] Breiman, L. 2001. "Random Forests". Machine Learning 45(1), 5-32.

[21] R Team (2019) "R: A language and environment for statistical computing". version 3.6.2 edn. R Foundation for Statistical Computing, Vienna, Austria

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

[23] Minasny, B., McBratney, A. B. 2006. "A conditioned Latin hypercube method for sampling in the presence of ancillary information". Computers & Geosciences 32(9), 1378-1388.

[24] Roudier, P., Beaudette, D. E., Hewitt, A. E. 2012. "A conditioned Latin hypercube sampling algorithm incorporating operational constraints". In: Minasny B, Malone BP, McBratney AB (eds) Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping, Sydney, Australia, 2012, CRC Press, pp 1-6.

[25] Roudier, P., Brugnard, C., Beaudette, D., Louis, B. 2019. "clhs: Conditioned Latin Hypercube Sampling. version 0.7-2 edn. R package.

[26] Brenning, A. 2005. "Spatial prediction models for landslide hazards: review, comparison and evaluation". Natural Hazards and Earth System Science 5(6), 853-862.