

## ***Interactive comment on “Assessing the spatial variability of weights of landslide causal factors in different regions from Romania using logistic regression” by M. C. Mărgărint et al.***

**Anonymous Referee #1**

Received and published: 28 June 2013

1) General comments The paper investigates the spatial variation of predictor variables and their coefficients in modelling landslide susceptibility in four different regions in Romania. It presents new and interesting data and uses a commonly applied methodology (logistic regression) to model landslide susceptibility. Compared to the version available for the quick review the paper was significantly improved, however there are still some major limitations in the paper. The data was newly modelled using a train and test dataset. Unfortunately, in the validation step the AUC values were only calculated for the train dataset. Only in table 3 true positives and true negatives are mentioned for both the train and test datasets. The AUC values should also be calculated for the test

C394

dataset, since otherwise the high AUC values for the train dataset might indicate just an overfit of the model. The results section is relatively short and statements in the discussion section are not always supported by the results (e.g. "landslide susceptibility in all sectors is generally explained by the slope angle, land use and slope height above the channel network", this is not true for Helegiu sector where slope height is not significant at all!) The discussion seems to be a bit too simple and should be more detailed and of course supported by the results. The paper would benefit of a table showing the landslide densities for each land use and lithology class for each of the four sectors. These results could then also be more clearly discussed. Beside this, nothing is said about the correlation between the predictor variables. There might be correlations e.g. between lithology and slope or land use and slope. This might explain some of your results, e.g. why slope angle is less important in Helegiu sector. In the study area section it is mentioned that different landslide types are dominant in specific sectors. They conclude that "it can be stated that the weights assigned to causal factors by means of logistic regression are capable to reveal some important regional characteristics for landslide manifestations". This statement does not get fully clear reading the paper. For me it seems that this might be also linked to the different landslide types. But this issue is not addressed in the discussion section at all. It still gets not clear for all regions which landslide types do occur. Please provide a table on no. of landslides with regard to the landslide types and state of activity for each sector. So far you have used all landslides, but there might be a huge difference if only shallow landslides are modelled or also the large landslides like glimee and hartoaape. Regarding the glimee in the first manuscript it was stated that these are generally stabilized at present (which would mean, that modelling landslide susceptibility with present day data is very critical). In the current manuscript this part was removed and it does not get clear why? Regarding the hartoaape it is stated that these are slide amphitheatres, which are semicircular depressions, shaped through successive landslide and/or erosion processes.... It does not get clear if such features can also be formed just by erosion processes how the authors decide whether it is a landslide or not. For both processes a sketch and photo

C395

would be very helpful to understand these special landslide types. Compared to the first version a whole paragraph on landslide susceptibility assessments in Romania in the past was deleted. From my perspective this was a very interesting paragraph and should be integrated again. If possible the authors should also discuss their approach and its benefits compared to the other approaches applied in Romania, at least the ones which were also carried out in their study area like Balteanu et al. 2010. The paper could really benefit from this. Based on these comments I suggest that the paper is reconsidered for publication after major revision.

2) Specific comments + Throughout the paper there is some sort of mixture of the terms weights and coefficients and causal factors and predictors, respectively. When referring more strictly to the statistical approach the authors use coefficients and predictors, in other cases weights and causal factors. I would prefer coefficients and predictors. Especially causal factors is misleading since most predictors have no direct cause-effect relationship towards landsliding. + The abstract should be updated after all comments are addressed and the paper improved. + Introduction: You should add a short definition of landslides.

p.1751 l.11: ...defined as spatial occurrence probability...

p.1752 l.29: what is the meaning of "roughly the same predictors"?

p.1753 l.12-13: Please check the sentence. Furthermore, you should refer in the discussion sections also to your previous study and give some statements if the results are similar or different and if different, why?

p.1753 l.21-22: please explain shortly what relatively high relief fragmentation means, and what does low relief fragmentation mean.

p.1754 l.20ff: nothing is mentioned about the elevation in this sector. Do you have some numbers on the amount of deforestation? Was everything transformed into pastures?

p.1755 l.18: Have you transformed the aspect layer into northness and eastness?

C396

Otherwise there is always a problem with aspect since values close to 0 and close to 360 are indicating North!

p.1757 l.18ff: It does not clear how the landslide points in the depletion areas were sampled. Have you used just one point per landslide or more? Furthermore, you write that the same random sample size were generated outside the landslide depletion areas. Does this mean that non-landslide points can also be located in the landslide accumulation areas? This should not be the case! Furthermore, you should clearly state that the model was trained with the train dataset of 80% of the landslide and non-landslide points.

p.1757 l.26ff: Please check sentence

p.1758 l.3ff: Most of these belongs to the introduction section. Here you should only describe the method you applied. Please describe the ROC Curve and AUC parameter more detailed as well as the way you prepared table 3.

p. 1758 l.16-17: This is partly a repetition.

p. 1759 l. 1-3: This belongs to the methods section.

p.1759 l.20: slope aspect might was removed because it was not transformed (see above).

p.1759 l.22: mean curvature probably was not selected since in 3 sectors plan and profile curvature was selected

p. 1760 l.5ff: Why is it obvious that many landslides occurred prior to land use change? Nothing is stated on landslide age, maybe you should add something in the study area section. What are the consequences for your results and the applicability of your map if the landslides are older than the information on land use in your dataset?

p.1760 l. 20: What is the meaning of unproductive land class?

Fig 1: Figure Caption: please add "(in red)" after distribution

C397

Fig. 3: The two red classes are difficult to be differentiated. You might add a yellow colour before the orange one and delete one red colour.

Fig. 4: Please add the ROC Curves and AUC values also for the test datasets

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 1, 1749, 2013.