Nat. Hazards Earth Syst. Sci. Discuss., 1, C912–C928, 2013
www.nat-hazards-earth-syst-sci-discuss.net/1/C912/2013/

Natural Hazards and Earth System Sciences
Discussions

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.

M. C. Mărgărint et al.

adriangrozavu@yahoo.com

Received and published: 17 August 2013

We thank the reviewer for his valuable comments, an answer to all comments are given bellow. Questions and remarks are in italic whereas the answers are in bold.

We also uploaded the pdf version as supplementary material.

1- p. 1752, l. 11: the reference Thiery et al., 2007 "Landslide susceptibility assessment by bivariate methods" deals with the WOFE method not the BLR. The reference was placed at the right position in the text as follows: "Statistical methods include bivariate analysis, like weights of evidence (WOE), which approaches the relations between

the controlling factors individually (Thiery et al., 2007), and multivariate analysis ..."
2- Study area section: To improve the readability of this section, you should make a sub-section for each of the described study areas.

Done

3- The section should propose a more substantial description of the landslides for each study area: type, number, size, etc.... Usually, for this type of paper the authors give these details with descriptive statistics (table or diagram). Please give some photographs of the landslides to illustrate.

We added for each sector a more detailed description, including a table with types, number and state of activity, 4 photos and 1 sketch (Figures 1-5).

4- The methodology section should be divided in three parts, the first one dealing with the description of the methodology, the second one presenting the data and a third one describing the modeling strategy (calibration and validation of the model).

Done

5- Maybe a multicollinearity diagnostic prior to the stepwise LR could be a good opportunity to assess the correlation between the variables.

We considered as less important such a move as long as the stepwise procedure solves the problem of multicollinearity.

6- p.1575, from l. 18 to 24: it still doesn't remain clear how many points were selected per depletion area? Did you select just one point or more?

The depletion areas were identified semi-automatically by using a geomorphometric variable called mass balance index. This parameter was derived in SAGA-GIS using the DEM and vertical distance to channel network as input layers. It was found that values greater than 0.1 of mass balance index correspond largely to landslide depletion areas. Grid points were generated then in the areas with mass balance index values greater than 0.1 and inside landslide polygons. Finally, the resulting point sample was visually inspected and corrected when necessary. These samples contain about 800 – 1000 points. Small landslides often received a single point in the depletion zone, while larger landslides received several points. After the depletion areas were sampled, we generated random samples of similar sizes outside the depletion areas and outside landslide polygons. The article states: "In order to test the predictive potential of the models, 20% of the samples, randomly selected, were used for validation as independent datasets.". The following sentence was inserted into text: "Consequently the training samples represents 80% of the landslide and non-landslide points."

7- For the "0" or "no landslide" sampling, it is usually preferred to use stratified random sampling, or spatially stratified random sampling than classical random sampling in order to avoid potential overfitting problems.

This issue has been addressed previously. We did not however used stratified sampling strategy.

8- You have selected the Jenks method to classify the susceptibility maps. However this method is strongly dependent of the number of selected classes and of the values distribution. Moreover, it is often considered difficult to compare maps classified with this method. Don't you think that using fixed logistic scores or equal interval classification could be better in order to compare the final maps. (This is rather an open question that can be discussed than a major problem).

This is an important issue and as far as we know there is no agreement concerning the best approach. There are several possible ways to separate the susceptibility classes: equal intervals, standard deviation based separations, natural breaks method, quantiles etc. The use of equal intervals has the disadvantage of emphasizing one class relative to others (Ayalew and Yamagishi 2005). The natural breaks algorithm (Jenks, 1977) performs the classification by grouping similar values while maximizing the differences between classes. It gives good results when the LSI histogram shows evident

breaks. Some authors recommend the standard deviation approach as the best choice for class separation (Ayalew and Yamagishi 2005). Though the limits of classes vary slightly from one sector to another when using Jenks method, the differences are insignificant. The high and very high landslide susceptibility classes are delimited by LSI values of 0.49-0.52 and 0.74-0.76 which, in our opinion, allows us to compare the results for the four sectors.

9- p. 158: the paragraph describing the LR model quality assessment has to be developed. Please explain what a pseudo coefficient of determination is, I think few people exactly know what it is. Explain clearly what is a ROC curve and AUC.... what is the real meaning of this test?

The pseudo coefficients of determination were only mentioned as quality parameters of logistic regression models. They were not actually used to assess the quality of models in our study. Therefore we didn't consider necessary to insist on them. However, the following sentence was added in order to clarify this issue: "Analogous to the determination coefficient used in multiple linear regression, the values of the pseudo-R2s vary between 0 and 1, measuring how well the model is adjusted." The ROC curve and AUC parameter was explained more clearly in the text. The ROC (Receiver operating characteristic) methodology was originally developed in the field of radar signal-detection theory (Peterson and Birdsall 1953). It has been applied and developed in various fields, such as medicine, meteorology etc., including geomorphology and particularly landslide susceptibility assessment (Chauhan et al. 2010; Mancini et al., 2010; Guns and Vanacker, 2012). The ROC curve is a useful tool for assessing the accuracy of predictions issued by binary classifier system. It represents a graphical plot of true positive rate (known also as sensitivity) and false positive rate (known also as 1- specificity). In the context of the current research, the LR classifies the points as landslide points, if the probability value is greater than the specified threshold (0.5) or as non-landslide point, if the probability value is less than 0.5. The group of points representing landslides is the "positive" group, while the one representing non-landslide points is the “negative” group. A true positive prediction is therefore a correct assignment of a point to the landslide group. A false positive prediction is a wrong assignment of a point to the landslide group. A correct assignment to the non-landslide group is called true negative or sensitivity. The number of false positive predictions is equal with 1 minus the number of true negatives. By plotting the fraction of true positives out of the positives (true positive rate) against the fraction of false positives out of the negatives (false positive rate) for all possible values of the threshold parameter (from 0 to 1), it results the ROC curve. The point (0;1), corresponding to the upper left of the plot represents the perfect classification, when all points are correctly classified. The area under the ROC curve (AUC) is an indicator of the LR model quality. For a perfect classification, the AUC is 1. For a random model, the AUC is 0.5. Generally, a good model must have an AUC value greater than 0.7 and an excellent model an AUC value greater than 0.9 (XLSTAT tutorial).

10- The major pitfall of the paper concerns the landslide data used to calibrate and validate the LR model. It is commonly admitted that each landslide type has to be modeled independently as they are controlled by different predisposing factors. For example shallow translational slides are rater influenced by steep slopes and surficial formations, whereas deep seated rotational slides are rater controlled by ground geology/hydrogeology. Moreover, including old deep seated stabilized landslide with present day data can be very critical as they triggered on different environmental conditions... Then the variations observed in the coefficients could not only be explained by the regional setting, but also and especially by the different proportion between the landslides types in each region. This critical aspect and limitation is not discussed in the paper.

Indeed the differentiation of results coefficients are directly related to the two types of landslides (shallow and deep). For our analysis the considered landslide inventory contained only areas with obvious manifestation of landslide processes. These areas may be grafted on large relict and old landslides (glimee and hârtoape), which were

not included in the landslide inventory and, consequently, in the regression equation. We will extend in the discussion section.

11- The results section is too short and lacks of a general synthesis of the results. Some of the figures are cited in the text without any further explanation. You should be more accurate in the results description.

We will deepening this issue at the Results section, including explanations of all figures.

12- The ROC curves of the validation samples have to be presented as well on figure 4 or on an additional figure.

We added a figure with ROC curve for the validation sample.

13- As mentioned before, the discussion is too shallow as it doesn’t discuss any of the limitations of this work and of the quantitative landslide susceptibility in general (e.g. quality of the input data, correlation between the variables, landslide data sampling. . .).

We will add all this issues in the paper.

14- p. 1760, l. 24-27: You state that the relative high coefficients attributed to slope height are “explained by the high relative altitude of landslide depletion area on which the model is based”. Isn’t it that the lithology can be significantly correlated with the altitude in plateau regions with monocline structures? Maybe I’m wrong, but the landslides you describe in the study area section (called hârtoape), seems to be old deep seated landslides, as observed in many other cuesta regions of western Europe (UK, France, Germany, Belgium). This type of landslide can be strongly controlled by the lithology (sliding panels of hard rocks (limestone, sandstone, chalk. . .) on soft rocks (marls, clay, sands. . .). Then the altitude could be considered as a proxy to identify the sensitive lithology, generally hard rocks located at the top of the hill slopes (in absence of more detailed geological maps).

Indeed, deep seated landslides of the plateau region, particularly in Lungani sector are correlated with monoclinic structure and the emergence of hard rock at the top of the landforms. We will detail this in the discussion section.

15- Opening the discussion/conclusion with a reference to other works conducted in Roumania on landslide susceptibility mapping or/and on the possible interest of the local authorities in this work could be interesting.

We will add all this issues in the paper. We will insert some considerations regarding the limitation of the actual Romanian methodology in landslide susceptibility assessment (from our point of view), like: applying the same weight to the predictors for all administrative units (with neglecting the major geomorphological units), data acquisition at different scales (unrelated with 1:5000 scale, at which it should be realize the final maps), absence of the geomorphometrical parameters, like slope angle, slope aspect, topographical curvatures, distance to drainage network etc. The map carried out by Bălteanu et al, 2010, was realized with an other methodology, at 1:200,000 scale, with a less number of parameters. So, for small administrative units, at large scale, our approach could improve the accuracy of susceptibility maps. We added another 2 references regarding romanian methodology (ChiÅču, 2010; Manea & Surdeanu, 2012).

Technical corrections: 1- p. 1751, l. 11-12: the susceptibility defines the spatial probability of landslide source area, not the occurrence probability (which is the “hazard”).

Modified.

p. 1751, l. 25: please check the sentence (repetition)

We corrected the sentence as follows: "The quantitative methods have developed rapidly during the last two decades due to the growing accessibility of geoinformation tools, . . .”

p. 1755, l. 19 and p. 1756, l. 12: I don’t understand clearly if the term “surface lithology” refers to the outcropping layers or to the superficial depostits/surficial formations.

We replaced “surface lithology” with “lithology”.

p. 1756 l. 8: it is not clear if the aerial images were orthorectified or georeferenced?

We modified the sentence as follows: “The land use layer was created by vectorization of land use polygons on the basis of high resolution 2006 orthophotos, which were georeferenced using the 1:5000 topographic maps.”

p.1756, l. 14: I’m not sure that a higher geological complexity necessarily means that the map is more accurate.

We rephrase as follows: “At this scale, only Helegiu mountainous sector reveals a higher geological complexity.”

p.1758, l. 5: please provide years of publication of the references.

The following references were added in the paper: McFadden, D.: Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics (Edited by P. Zarembka), 105-42. Academic Press, New York, 1973. Cox, D. R. and E. J. Snell.: Analysis of binary data (2nd edition). London: Chapman & Hall, 1989. Nagelkerke, N. J. D.: A note on a general definition of the coefficient of determination. Biometrika, Vol. 78, No. 3: 691-692, 1991.

p. 1758, l. 16-17: please delete the sentence. It was already mentioned in the methodology section.

Done

Figure 1: It is difficult to see the location of the landslides. Can you please increase the contrast between the landslides limits and the hillshade background?

We enhanced the contrast of the figures.

Figure 2c: Please provide the lithology rather than the stratigraphy. The north direction is not indicated on the maps.

We replaced stratigraphy with corresponding lithology and we added the north direction.

Figure 3: The map is still very difficult to read, please select more contrasted colors. Please indicate the north direction.

We enhanced the contrast of colors and added the north direction.

Figure 4: Please add the validation ROC curves

We will add a separate figure with validation ROC curves

Figure 5: The figure might be easier to read with the same y-axis extend on each graph.

We will modify the y-axis.

References

Ayalew, L. and Yamagishi, H.: The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan, Geomorphol, 65, 15–31, 2005. Bălteanu, D., Chendeş, V., Sima, M., and Enciu, P.: A country-wide spatial assessment of landslide susceptibility in Romania, Geomorphology, 124 (3-4), 102–112, 2010. Chauhan, S., Sharma, M., and Arora, M. K.: Landslide susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model, Landslides, 7, 411–423, 2010. ChiÅču, Z.: PredicÅčia spaÅčio-temporală a hazardului la alunecări de teren utilizând tehnici S.I.G. Studiu de caz arealul subcarpatic dintre Valea Prahovei şi Valea IalomiÅčei (in romanian), PhD. Thesis, Bucarest, 2010. Guns, M. and Vanacker, V.: Logistic regression applied to natural hazards: rare event logistic regression with replications, Nat. Hazards Earth Syst. Sci., 12, 1937–1947. Jenks, G. F.: Optimal Data Classification For Choropleth Maps, Univ. of Kansas, 1977. F. Mancini, C. Ceppi, and G. Ritrovato, GIS and statistical analysis for landslide susceptibility mapping in the Daunia area, Italy, Nat. Hazards Earth Syst. Sci., 10, 1851–1864, 2010. Manea, S. and Surdeanu, V.:

Landslides Hazard Assessment in the Upper and Middle Sectors of the Strei Valley, Revista de Geomorfologie, 14, 49–55, 2012. Peterson, W.W. and Birdsall, T. G. 1953 'The theory of signal detectability: Part I. The general theory'. Electronic Defense Group, Technical Report 13, June 1953.

Please also note the supplement to this comment:
http://www.nat-hazards-earth-syst-sci-discuss.net/1/C912/2013/nhessd-1-C912-2013-supplement.pdf

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

Fig. 1. Figure 1. Deep-seated landslide in Transylvanian Plateau, locally named glimee (Căpuşu de Câmpie sector).

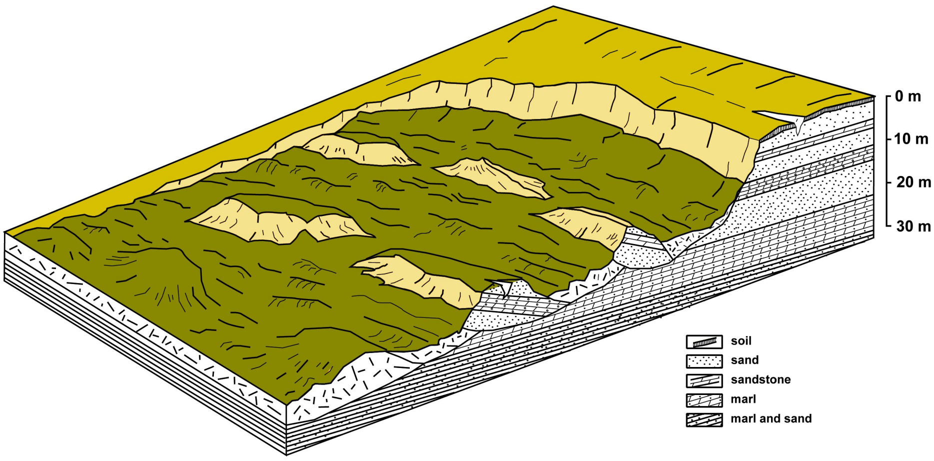

Fig. 2. Figure 2. Block diagram representing the deep seated landslide from Figure 1.

Fig. 3. Figure 3. Deep seated landslide in Moldavian Plateau (Şipote sector).

Fig. 4. Figure 4. Semicircular depression shaped by complex geomorphological processes (hârtop in Molvavian Plateau, Şipote sector).

Fig. 5. Figure 5. Shallow landslide in Şipote sector, detail from Figure 4.

Table 1a. Landslide density for lithology, slope aspect and land use classes.

Lithology		Slope Aspect		Land use	
Capuşu de Câmpie Sector					
Alluvial and colluvial deposits – *Quaternary*	0.196	North	0.710	Built area	0.001
Clays and marls, sand, sandstones, volcanic tuffs – *Sarmatian*	2.526	East	0.696	Arable land	0.200
Clays, sands, volcanic tuffs – *Pannonian*	1.058	South	1.226	Pastures	3.709
		West	1.574	Forest and pastures	4.100
				Forest	0.889
				Waters and wetlands	0.000
Şipote Sector					
Gravels, sands –*Quaternary*	0.472	North	1.638	Built area	0.084
Marls, clays, sandstones and sand complexes – *Sarmatian*	1.123	East	0.781	Arable land	0.268
		South	0.494	Pastures	2.744
		West	1.275	Forest	2.604
				Waters and wetlands	0.005
				Degraded land	5.053

Fig. 6. Table 1a. Landslide density for lithology, slope aspect and land use classes.

Table 1b. Landslide density for lithology, slope aspect and land use classes.

Lithology		**Slope Aspect**		**Land use**	
Lungani sector					
Gravels, sands (fluvial terraces) – *Quaternary (pleistocene)*	0.065	North	1.138	Built area	0.321
Sands, clays – *Quaternary (holocene)*	0.328	East	0.861	Arable land	0.498
Marls, clays, sandstones and sand complexes – *Sarmatian*	1.200	South	0.466	Pastures	1.911
		West	1.618	Forest	1.950
				Waters and wetlands	0.003
Helegiu sector					
Gravels, sands –*Quaternary*	0.279	North	1.138	Built area	0.021
Sandstones, volcanic tuffs – *Tortonian*	1.048	East	0.736	Arable land	0.065
Sandstones, marls, gypsum – *Helvetian*	1.161	South	0.852	Arable land and pastures	1.462
Sandstones, clays – *Volhinian*	1.189	West	1.316	Pastures	1.998
Marls, clays, salt – *Badenian*	0.902			Forest	0.221
Sandstones, menilite, dysodilic shales – *Latorfian-chattian*	1.196			Waters and wetlands	0.000
Argilaceous shales, clays, sandstones – *Priabonian*	0.814			Degraded land	3.268
Calcareous sandstones, marls, conglomerates – *Lutetian*	1.888				

Fig. 7. Table 1b. Landslide density for lithology, slope aspect and land use classes.