Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-294-AC1, 2020


Natural Hazards and Earth System Sciences
Discussions



# *Interactive comment on* "Exploring the potential relationship between the occurrence of debris flow and landslide" *by* Zhu Liang et al.

**Zhu Liang et al.**

499324556@qq.com

Received and published: 30 November 2020

Dear Editors and Reviewers: Thank you for your letter and for the Reviewers' comments concerning our manuscript entitled "Exploring the potential relationship between the occurrence of debris flow and landslide" (ID: NHESS-294). Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet with approval. Revised portion are marked in red in the paper. The main corrections in the paper and the responds to the Reviewer's comments are as flowing: Reviewer #1: General Comments In this paper, landslide susceptibility analysis and debris flow susceptibility analysis is carried out using Random Forests over the same area and the two resultant hazard maps are compared.



Printer-friendly version

Discussion paper

This is something that is not usually done in ground failure hazard assessments and I find the conclusions from this paper are interesting. The study was designed well and the figures are good, but in some places the work is not explained clearly enough or more information would help the reader to understand. Response:Thank you for your approval of our work. It is true that there is few manuscripts related in ground failure hazard assessments and we will try our best to make the idea more clear.

One thing I would also like to know is this: how have you chosen the training and test data for the study? Are they randomly selected in time and space? Or are historical landslides and debris flows being used to predict the locations of recent landslides and debris flows? Response: In second 3.1, we choose the 5-fold cross validation procedure. The data consists of negative and positive samples. Conditioning that the number of samples are limited and we selected the all the samples in time during 1970∼2010 for modeling. Related information have been added on line 96-97.

I am also confused by the decision made by the authors to convert all their continuous input factors (e.g. aspect) into categorical variables, as Random Forests work well with continuous variables. This is not something I have seen done in other studies using RF for landslide susceptibility mapping. If there is a specific reason that the authors have chosen to do this, it should be explained in Section 2.4. If this is the case, how did the authors choose the number of categories for each input factor? Response: It is true that random Forests work well with continuous variables. We did not convert the continuous input factors into categorical variables. We aim to reclassify variables as the thematic map will become more concise. As for the number of categories, we referred to related references which have been published. S.Chen, Z. Miao, L. Wu and Y. He, "Application of an Incomplete Landslide Inventory and One Class Classifier to Earthquake-Induced Landslide Susceptibility Mapping," inÂăIEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1649-1660, 2020, doi: 10.1109/JSTARS.2020.2985088. Zhu Liang, Wang Changming and Kaleem-Ullah-Jan Khan. Application and comparison of different ensemble learning machines combin-

ing with a novel sampling strategy for shallow landslide susceptibility mapping.ÂăStoch Environ Res Risk AssessÂă(2020c). https://doi.org/10.1007/s00477-020-01893-y

Specific Comments Line 88: I would also like to know if the landslides are mapped as points or polygons in your dataset (I assume points since they are shown as points in Figure 1) and the mapping resolution, or at least the resolution of the google earth images used in the landslide mapping. It would be useful to know what proportion of landslides were mapped using the different methods (i.e. historical records versus google earth image interpretation) and how far back in time your historical records go. Response: We are agreed with the comment. The landslide locations are recorded as a point which are showed in Fig.1. We have provided detail information like time about records.

Line 101: What is meant by "There is no physical relationship between a grid-cell and slope" – do you mean that slope will vary within a grid cell? Response: Landslides are the result of slope processes acting at different spatial and temporal scales that result in geomorphological forms of very different shapes and sizes that are difficult to capture by grid-cells accurately. The geometry of a landslide is better represented by a polygon or a set of polygons in vector format; unless the size of the grid-cell is very small compared to the size of the landslide.We have referred to the original manuscript and make it more clear.

Section 2.4: In this section, I think more justification is needed for the choice of controlling factors. I would also divide the section into "factors used in landslide susceptibility assessment", "factors used in debris flow susceptibility assessments" and "factors used in both". I think you have done this, but I would make it clear at the beginning of each paragraph which input factors you are describing. Response: We are agreed with the comment. We have added related information and divide the sections clearly.

Line 113: It's true that different parameters are used in soil slide and debris flow susceptibility assessment. However, there is also quite a lot of difference between the fac-

tors used by different landslide susceptibility assessments. Response: We are agreed with the new expression which is more accurate and clear.

Line 120: When you say NDVI, are you using pre-event NDVI, as a proxy for land cover type, or post-event NDVI as a direct measurement of vegetation removal caused by the debris flow? Response: NDVI reflects the vegetation conditions in the area and we use the pre-event NDVI as a proxy for land cover type.

Lines 148-153: What are the sources of your datasets? (the geological map, the DEM, the roads, the faults, the rainfall) Response: We are agreed with the comment. We have provided related information on line 150, 151 and 154.

Line 149: What is the source of your DEM data? Response: We provide related link where we download the DEM data on line 150.

Line 187: There are several options for optimisation in sci-kit learn. Which one did you use? Response: Cross-Validation were applied in our work (line 187).

Line 193: I think here you mean you are analysing the relative importances of the conditioning factors. Response: Yes, what you think is correct and we have explain it correctly on line 193.

Line 201: Specify here that AUC of 0.5 = No Skill for ROC curve, otherwise people might think the scale is from 0-1. Response: We are agreed with the comment. We have explain the AUC value more clearly on line 201.

Line 232: When you say “disaster points”, you are referring to debris flows, so I would just say observed debris flows. Response: Disaster points in our work referred to all landslide locations not just to debris flows.

Line 257 I don’t understand what is meant here by “factor analysis”. What exactly has been done? Response: Factor analysis is a method which is usually used for dimensionality reduction and exploring the major factors. We have added related information on line 257-259.

Line 260: Please explain how KMO testing works and how to interpret the values. Response: We have added related information on line 261-262.

Line 263-265: When you say model 1 and model 2, are these the landslide SZM and debris flow SZM models respectively? Response: 36 watershed units with distribution of high or very high-grade slope units were taken as model 1 and the left 8 watershed units as model 2, which has been explained on line 263-265. The models established is to explore the major conditioning factors for analyzing the reason why some high or very high-grade susceptibility watershed units are covered with low susceptibility slope units.

Line 312: You should give examples of studies that use logistic regression and discriminant analysis here to back up your statement Response: We are agreed with the comment. We have added related references here.

Line 314: Random Forests have been applied to landslide susceptibility in several previously published works, which should be referenced here. Some examples: Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z. and Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena, 151,pp.147-160. Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V., 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Natural Hazards and Earth System Sciences, 13(11), p.2815. Zhang, K., Wu, X., Niu, R., Yang, K. and Zhao, L., 2017. The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. Environmental Earth Sciences, 76(11), pp.1-20. Response: We are agreed with the comment. We have added related references here.

Section 5.2: If I understand correctly, what you are saying here is that landslide susceptibility maps should not be used in debris flow hazard assessment and vice versa. This seems to me to be an important conclusion from this paper and should be stated

more clearly. Response: Yes, it is correct that different kinds of landslides should be evaluated respectively conditioning that conditioning factors and scale varies. We add related information on line 338-339.

Line 343: this is not very clear can you give a more specific example? Response: We adjusted the language order and enhanced the before-and-after logic.

Table 1: The layout of this table is a bit strange, having a single row with so much information in it. I also think the parameters may not make any sense to someone who has not used the sci-kit learn package for example “max_features, sqrt”. Response: We are agreed with the comment. The application of machine learning usually involves several hyper-parameter needed to be tuned before modeling. Different machine learning methods need different optimum parameters or different optimisation technique will generate different values of parameters which are not easy to implement and even means nothing to someone who are not skilled at modeling. The main aim of our work is to explore the relationship between soil slides and debris flow by mapping respectively not for the comparison of different methods. Therefore, it will be easier to be understood if Table 1 is removed.

Table 1: You have two models here: one for landslides and one for debris flows. Did the optimisation technique you used yield the same optimum parameters for both models? Response: The optimum parameters are not the same because the number of samples of landslide and debris flow are different. On the other hand, the conditioning factors for landslide and debris flow are also different. We should make it clear by listing the parameters respectively.

Figure 10: There is no label for the X axis. I assume it should be “importance (%)” Response: It will be more clear when the fig. is labeled with the X axis and we have add related information. Technical Corrections

Line 52: There is no space between flow”(Varnes, Response:We have make it correct.

Line 94: There is no space between 7.1(Fig.4) Response:We have make it correct.

Line 148: Do you mean the “raw data” rather than the “row data”? Response: It should be raw data. Line 166: “curves( Green” the space should be before the bracket Response: We have make it correct.

Lines 170-172: Do these need to be separately numbered equations? Also in “Sensitivity” and “Accuracy” some of the word is in italics and some is not. Response: Accuracy, Sensitivity and specificity are three similar indexes for evaluating the performance of model and we have referred to some other papers that the equations are listed in together. TP, TN, FN and FP should be in italics and we have made them correct.

Line 189: Are there two spaces between “trees” and “and”? Response: We have make it correct.

Line 260: significance (Sig) was defined earlier in the manuscript Response: We have make it correct.

Line 337: No space between “respectively(Fig.10)” Response: We have make it correct.

Line 286: Zonation not Zoination Response: We have make the word correct.

We appreciate for Editors and Reviews’ warm work earnestly, and hope that the correction will meet with approval. Once again, thank you very much for your comments and suggestions. With best regard, Yours sincerely, Zhu Liang Jilin University

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-294, 2020.