

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

**Anonymous Referee #1**

Received and published: 28 November 2020

### 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. 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.

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? I am also confused by the decision made by the authors to convert

C1

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?

### 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.

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?

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.

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 factors used by different landslide susceptibility assessments.

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?

Lines 148-153: What are the sources of your datasets? (the geological map, the DEM,

C2

the roads, the faults, the rainfall)

Line 149: What is the source of your DEM data?

Line 187: There are several options for optimisation in sci-kit learn. Which one did you use?

Line 193: I think here you mean you are analysing the relative importances of the conditioning factors.

Line 201: Specify here that AUC of 0.5 = No Skill for ROC curve, otherwise people might think the scale is from 0-1

Line 232: When you say “disaster points”, you are referring to debris flows, so I would just say observed debris flows.

Line 257 I don't understand what is meant here by “factor analysis”. What exactly has been done?

Line 260: Please explain how KMO testing works and how to interpret the values

Line 263-265: When you say model 1 and model 2, are these the landslide SZM and debris flow SZM models respectively?

Line 312: You should give examples of studies that use logistic regression and discriminant analysis here to back up your statement

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

C3

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.

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.

Line 343: this is not very clear can you give a more specific example?

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”.

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?

Figure 10: There is no label for the X axis. I assume it should be “importance (%)”

Technical Corrections

Line 52: There is no space between flow”(Varnes,

Line 94: There is no space between 7.1(Fig.4)

Line 148: Do you mean the “raw data” rather than the “row data”?

Line 166: “curves( Green” the space should be before the bracket

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.

Line 189: Are there two spaces between “trees” and “and”?

C4

Line 260: significance (Sig) was defined earlier in the manuscript

Line 337: No space between “respectively(Fig.10)”

Line 286: Zonation not Zoination

---

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