Comment on nhess-2021-299

Anonymous Referee #3

Dear authors, I have read and carefully evaluated your manuscript "Assessing the importance of feature selection in Landslide Susceptibility for Belluno province (Veneto Region, NE Italy)". I found that the idea of the work is original and the outcomes are interesting, thus I recommend publication after a round of minor revisions. Please consider my comments hereafter.

Dear Reviewer, thank you for reviewing the manuscript. We have marked the changes using track changes manuscript (Line numbers are referred to track changed file).

GENERAL COMMENTS

- English is understandable but a further revision from the most experienced coauthors would be useful to improve it.

Thank you for your comment, we have improved the English language throughout the manuscript.

- The idea behind this work is interesting and original. The analysis performed is incomplete. The comparison between different models and different parameters' configurations is performed only in terms of performance metrics and visual inspections. In my opinion this is not enough to characterize the differences between the maps, a more thorough comparison is needed. I strongly recommend following the approach of Xiao et al. 2020, which is quite simple and could be done in a few minutes, adding a nice part in the discussion: the authors could make a comparison between the "full" and "reduced" version of each model by making the arithmetical difference of the susceptibility values calculated by the 14-factors-version of the model and the 9-factors-version of the same model. The same comparison could be repeated for the three models. The calculated difference would be useful to check if the differences are randomly distributed or if they consist in systematic errors driven by the spatial distribution of some variable (maybe one of the discarded ones? That would be even more interesting). That would add a nice value to the discussion and interpretation of the differences encountered.

We agree with your comments and therefore we added the above mentioned analysis in the discussion section.

- It is not clear how you actually implement the models. Do you use Matlab, R, or other software?

We used a Python environment and used machine learning modules from Scikit-Learn to perform the susceptibility modelling.

SPECIFIC COMMENTS

L21= which threshold? This is not clear

We have added the threshold (0.30 for statistical and 0.03 for ML algorithms).

L22-24= I suggest being more precise and accurate. I would name clearly which are the variables discarded and the most important ones.

We have added the list of the discarded variables. Please see lines 26-28.

L26-30= I suggest deleting. The first sentence is a repetition. The second one is not fundamental.

Based on your suggestion we have deleted the repeated sentences.

L34= A reference is needed here about impacts of landslides, e.g. Froude et al., 2018

We have added the suggested reference Froude et al., 2018.

L47-48= what do you mean with “the significance of landslide studies”? I suggest rephrasing.

Based on your suggestion, we have rephrased the sentence.

L51= reference needed (e.g. Reichenbach et al., 2018)

We have added the suggested reference Reichenbach et al., 2018.

L62= data driven is a broader “family” of approaches, it does not apply to FR alone.

We agree with you and rephrased the sentence.

Section 2.1= I would move the figure of the study area here.

We have moved the figure of study area in section 2.1.

L165 (and elsewhere)= please use a more specific term instead of “precipitation”: which variable are you using? Mean annual precipitation? Mean monthly rainfall?

Thank you pointing it out, we have used Mean monthly rainfall and mentioned it through out the manuscript.

Table 1: I have some concerns about some factors. Elevation: I think at this latitude it is important because it largely influences climate, including the amount, intensity and distribution of rainfall. Slope: I would use the range 0°-86°. Aspect: motivation is quite confused. TWI: please change “measures” with “estimates”. TRI: please, rephrase the last part of the description (fluctuation, moves): it seems that the topographic surface moves. 7 and 8: please add the values of the ranges. 10: which feature of the rainfall regime are you considering? Mean monthly rainfall? 11: I don’t know the study cited here and I couldn’t find it in the reference list; I suggest to cite a paper specifically focused on the importance of lithology in LSM (e.g. Segoni et al., 2020). 11 and 13: I would include all the classes in the third column and not “etc.”.

Thank you for the detailed comment, we have modified the table 1 based on your suggestions.

L306 and 310: here I understand that you use natural breaks (l306) […] which is better than natural breaks (l310). Please, revise.

We have revised the sentences.

Fig 6 and 10: it would be nice to mark somehow the discarded factors. E.g. by coloring them in a different way or by adding a horizontal red line at 0.3 to show the cut-off value.

Thank you for suggestion, we have improved the figure 6 and 10.

L402: 0.910

We have changed 0.91 to 0.910

L408: after this discussion, it would be interesting to show the comparison I suggested in my general comment. You can show the three maps FREBF14 - FREBF9, XGB14-XG9 and RF14-RF9. And briefly

describe the spatial patterns of the differences shown: if you remove 5 factors, are the differences random or do they follow a systematic pattern?

Thank you very valuable suggestion, with the analysis we were able to compare the resultant maps and it has an added value to the manuscript. We refer you to the end of the discussion section and figure 12.

L460: instead of making reference to two figures that are not here, you could provide another figure, or a table, listing the factors ranked by importance. It would be meaningful to observe if the ranks are similar and if the discarded variables are similar.

Thank you for the suggestion, we have inserted figure 13 listing the factors ranked by importance for all the three models.

REFERENCES

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. Natural Hazards and Earth System Sciences, 18(8), 2161-2181.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. Earth-science reviews, 180, 60-91.

Segoni, S., Pappafico, G., Luti, T., & Catani, F. (2020). Landslide susceptibility assessment in complex geological settings: Sensitivity to geological information and insights on its parameterization. Landslides, 17(10), 2443-2453.

Xiao, T., Segoni, S., Chen, L., Yin, K., & Casagli, N. (2020). A step beyond landslide susceptibility maps: a simple method to investigate and explain the different outcomes obtained by different approaches. Landslides, 17(3), 627-640.