

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/nhess-2021-135-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2021-135

Clàudia Abancó (Referee)

Referee comment on "Evaluating methods for debris-flow prediction based on rainfall in an Alpine catchment" by Jacob Hirschberg et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2021-135-RC2>, 2021

General Comments:

The main question addressed by this paper is the uncertainty on the rainfall thresholds for debris-flow prediction. It presents a study at local scale, but also analyses the implications that using a regional landslide dataset would have had on the final Intensity-Duration (ID) threshold. It deals with several aspects which are well known on the literature by causing uncertainty on the definition of thresholds: such as the statistical techniques used, the size of the data set or the variables included in the ID threshold. The topic is interesting and relevant since, as pointed by the authors, no standardized procedures exist yet to define thresholds and many uncertainties still remain on the data and techniques to be used.

Although the uncertainty in rainfall thresholds definition for landslides and debris flows is a common research topic, this paper clearly shows originality. It deals with some classic topics such as the effects of the database length or the method used to estimate the threshold, but both the methods and the database used are of good quality and original. I find specially interesting the multivariate approach including further rainfall properties (more than the standard ones) and the seasonal proxies.

The paper is very well written, clear and easy to read.

The conclusions are consistent with the evidence and arguments presented. They address the main questions proposed.

The Figures are in general clear, and helpful to follow the paper. I have added some comments in the specific comments for one particular figure, which I find quite difficult to

follow.

As a summary, I really enjoyed reading the paper and I think the authors did a very nice work. However, it needs some revisions before publishing it in NHESS.

Specific comments:

Title: It may be a bit misleading. It does not deal with the limitations of the thresholds but more with the uncertainties on their definition? I would suggest reconsidering it...

L65: I would suggest adding a few references of studies using different MIT, as it is said they range from 10 min to 6 h but no references are given (although they appear later, in 3.3., but I would add them here too)

L82: Although the two methods that are going to be compared are mentioned in the abstract, I would list them here too

L80-85: I miss here stating as an objective (maybe as a secondary one) the analysis of the performance with local vs. Regional dataset, which is stated in the abstract.

L89-98: Cite Figure 1 in this paragraph

L106: Rain gaugeS in plural? If there's more than one, why in the Figure only 1 is shown? Why only data from 1 is used?

L111-113: If the geophones and depth sensors have been removed, which sensors (less maintenance) have been installed?

L115: By including the citation of Badoux in line 113 I think you could delete it in Line 115, it is clear for the Reader that details can be found there

L122: Any reference where snowmelt has been observed (even if not as sole trigger)?

Figure 1: Why only force plate is indicated? I would suggest adding the other sensors (the new ones replacing the geophones)?

L135: 5 mm? This sounds like a very low number...

L136: Does the local rain gauge not have a thermometer? Also, I would suggest moving Temperature, lightning strikes and other parameters to another paragraph, as I understand these are all variables for the Machine learning, but not actually for the main ID thresholds comparison? I was a bit confused reading about rainfall and changing to temperature abruptly, as it's the first time you mention the temperature variable

L160: Reference of TSS?

L169: Include Area Under Curve as clarification of what AUC stands for

L173: I understand that you used the same criteria for both trigg and non trigg rainfalls?

L175: Delete ? before Deganutti

L180 and Figure 2: I think this is more results than methods?

Figure 2 (b): Sensitivity? This word may be confused by SE? Could it be called "Analysis of ID-threshold parameters with changing MIT"?

L184: I would not say that β stabilizes, but reduced the increasing tren dat MIT 3h... (in Fig 2b)

L197: Actually, if it was snowing the data from the rain gauge would not be valid, right? As it is not heated... Have you considered this? If so, maybe you could mention here.

L200: Very interesting selection of parameters!

L208: This last sentence of the paragraph ("Lately, confusion matrix...") is actually a bit confusing to me. You have not used frequentist method, right? You used linear squares and LR&TSS and TSS&TSS methods if I have understood right. Therefore, the sentence is confusing as it seems that you have calculated the confusion matrix and ROC for the frequentist method...

L220: This is also confusing. A record of length 5 years includes 5 annual samples that can include repetition of the same year? Why is the procedure repeated 100 time for each record length? Please clarify

L280: I think it would be good to see the total rainfall amounts at some point in a Figure, as it is stated here that long duration need more rainfall (logic, but still nice to see)

L297: Higher antecedent rainfall amount may lead to higher degree of pore saturation along the entire channel bed, but also, in some cases the antecedent rainfall would mostly contribute to the generation of lateral flow and increase of water table (e.g.:

M.N. Papa, V. Medina, F. Ciervo, A. Bateman 2013, **Derivation of critical rainfall thresholds for shallow landslides as a tool for debris flow early warning Systems**). This could also correlate with the fact that magnitudes are bigger, but I would say that the correlation between the antecedent rainfall and the magnitude it is a tricky point and needs careful evaluation...

L305: TSS&TSS thresholds are lower for short durations (<4.5 h) and higher for long durations- after this I would add (Figure 5e)

L311: However, the biases decrease to _30% already after 6 years or _25 triggering events- For both? Or only for β ? I can't see it that clearly in alpha?

L335: Also, the source of rainfall data is different, right? If I am not wrong the work of Leonarduzzi et al. it was not based only in rain gauge data. Therefore, apart from climatic, topographic and lithologic uncertainties it may be also from the type of rainfall data?

Figure 7:

- The grey dots are very difficult to see, specially over blue, red and green bars. Change colour of bars or make dots bigger
- I find this figure particularly dense and a bit difficult to follow. Some ideas on how it could be made a bit easier to read:
 - I understand that RF_ID+1 is based in one single predictor (the one with best

performance). Why not indicating which one instead of leaving the reader to interpret?

- Same with RF_ID+var and 4 predictors
- Maybe then it would not be necessary to include all the single predictors in the same figure. Either include them in a separate figure or as supplementary material?
- If you think it is relevant to keep the same format, I would suggest indicating the selected predictors for each RF model in some way...