Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/nhess-2021-206-RC1, 2021

Comment on nhess-2021-206

Anonymous Referee #1

Referee comment on "Brief communication: Introducing rainfall thresholds for landslide triggering based on artificial neural networks" by Pierpaolo Distefano et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-206-RC1, 2021

Dear Authors, I have read with great attention your brief communication entitled "Rainfall thresholds based on Artificial neural networks can improve landslide early warning". The topic is original and interesting, the research design is robust and innovative, the English is good, the structure is fine. The manuscript surely deserves publication in NHESS. However, before endorsing final publication, I would like to ask you some clarifications and some improvements. I think all the modifications could be considered intermediate between "minor revisions" and "major revisions".

I look forward to receive the revised verison of the paper.

Best regards.

L13: A reference could be useful here.

L16. I find odd not citing Caine (1980) who gave start to the research topic about I-D thresholds.

Fig 1: even zooming the pdf, I cannot distinguish very well red triangles and red points. Could you please change the color of the 2009-2018 rain gauges? Green would be an excellent choice I think.

I recommend to add some details in the methodology description. In particular:

- If possible, I recommend describing the typical landslide typology of your dataset. This is important to understand which is the "target" landslide typology for your model, as different landslide types may be more sensitive to very different rainfall characterisitcs. Since you used FraneItalia, which is basically derived from newspapers, I guess you cannot exactly assess the typology of each landslide of your dataset, therefore your model is aimed to model and predict every landslide tyopology. Is my assumption correct? I don't think that would be wrong, but I think it should be clearly stated in the manuscript.

- Both in CTRL+ and ANN: it is not clear how you relate each landsldie to the triggering rainfall. Do you use the nearest rain gauge? Do youconsider all the sorroundign rain guges? In the secnond case: how you decide which rain gauge is selected to characterize the triggering rainfall?).

- I understand that you train the ANN with 70% of 144 triggering events and 70% of 47398 events. Doesn't it lead to an unbalanced prediction? ANN will be trained to detect non triggering conditions more effectively than triggering conditions.

L86 - please make clear the difference between training and validation dataset. I assume one of them is used for internal verification of the model while the other is used as an independent verification. Could you please make it clearer? I am used to say "calibration", "internal testing" and "independent validation", but since the order of your terms is different I guess some confusion may arise.

L104 - Here you introduce ROC curve, but then you don't use it (and I agree that is not an useful metric for the objective of this work). I think this part can be deleted.

Equations 7,10. These threshlds seem very low. I think in an operational use they would be regularly exceeded, expecially for short durations (think about how many times it rains 5.6mm in one hour). I understand your reasoning about the exponent, which makes the threshold higer for longer durations, but maybe you should state which is the duration range for which the thresholds are valid (e.g. the equation is enpirically defined for duratinos between 10 hours and 100 hours: the rest is an extrapolation where empirical data do not exist). Moreover, you can link this issue with the following discussion (around line 150) about the effectiveness of the ANN: the shortcoming of a power law is that the same equation is assumed valid for all the durations, while ANN could be more flexible).

L127-130 and table 2. Since I = H/D, performances of D-H and D-I should be identical. I think the reason of the differences in Tab 2 is the number of hidden neurons. Results of table 2 are influenced by the reinfall parameters (first column) and by the model metaparameters (e.g. hidden neurons - second colums). This complicates the discussion

and interpretation of the results.

TAb.2 I would add to the table two colums showing TPR and FPR: TSS alone is not very inforative about the effectiveness of the thresholds (e.g. 0.3 could be derived by the couple of values 0.9 and 0.6 or by the couple 0.4 and 0.1). I suggest to keep the table simple and to add TPR and FPR only for the independent verification dataset (the test dataset? See previous comment).

L150 I think this is a good point to add a couple of lines about what I mentioned in one of my previous comments.

L152-155. This point is very important. I-D and E-D thresholds work very well in case of shallow landslides in permeable soil (that's how Caine introduced them back in 1980). Later, researchers tried to estend the applicability of the techniques also to other settings, but the methodology shows evident theoretical and practical limitations (expecially when case studies are tested against a rigourous validation procedure). At present, research focuses on inoovations to increase the effectiveness of the technique proposing enhancements to better adapt to complex case studies. The idea of adding a third variable to the model is one of this innovations and others (e.g. Rosi et al. 2021 - even if they used antecend rainfall as third variable) obtained an increased effectiveness. I suggest adding this reference to your reasoning to better stress the results you obtained.

L156 - I find very interesting your work and I think this use of ANN is very promising. However, I suggest to mention some limitation. For instance, I think ANN would be difficult to operate and this aspect is still open to future research (I think it is why you presented a brief communication instead of a research paper). You add something similar in the conclusion but in my opinion the conclusion sectionshould not contain new concepts and this comment would be better placed at the end of the discussion.