

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/nhess-2021-206-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2021-206

Anonymous Referee #2

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-RC2>, 2021

General Comments

The theme addressed in the manuscript is of interest and relevant within the scope of NHESS, particularly regarding the definition of landslide-triggering rainfall thresholds to be possible to be included in an early warning system for landslides. The manuscript, in my opinion presents some aspects that must be better addressed, modified, or discussed in more detail. Strengths: the methodology /the scientific method. Weaknesses: the analysis of the results /discussion.

Specific comments

Scientific Significance/Originality:

The manuscript presents an interesting approach to evaluate the possible contribution of ANN's to determine rainfall thresholds for landslide occurrence by comparison with rainfall thresholds determined by more frequentist methods. One of the main contributions of the work is the possibility to explore different variables related with the rainfall events that triggered and not triggered landslides to better characterize the rainfall critical conditions responsible for the landslide initiation and improve the predictive performance of the rainfall thresholds and its possible application to a landslide early warning system.

Scientific Quality:

The manuscript reports scientific and technical subjects relevant within the scope of

NHESS, nevertheless, and if I made a correct interpretation of the approach used and results, is my understanding that some aspects must be better addressed, modified, or discussed in more detail.

- The title of the manuscript is suitable to be improved. The early warning component was not effectively explored or sufficiently described in the manuscript and therefore my suggestion is to reformulate or to remove the reference to “can improve landslide early warning” from the title and address it in the conclusions as future work.
- Introduction section: The discussion regarding the different types of rainfall thresholds, explanatory variables, constraints, and strategies to improve their predictive performance available in the literature is very limited and needs to be ameliorated. This, to be perfectly understandable the methodological aspects that the ANN approach intends to fulfil. In addition, from lines 23-40 most of the examples cited by authors regarding the application of ANN are more related with the assessment of landslide susceptibility, which is not the scope of the work, than with rainfall thresholds definition or applications of rainfall thresholds to early warning systems. Consequently, is my understanding that this part should be completely reformulated.
- Lines 38-40. The work objectives could be better defined. Although I recognize that the definition of rainfall thresholds can contribute for the development of landslide early warning thresholds, I think that was not necessarily explored/describe in the present work. To include this topic, it should be clearly presented in the methods and in the results section the link between them.
- On the Data and methods section a brief description about the overall quality/accuracy of the landslide data (particularly about the landslides date of occurrence), and about the completeness of the rainfall database is critical since it can affect the prediction capability of the rainfall thresholds. In addition, authors mentioned that the rainfall data is available until 2018 but landslide inventory include landslides updated for the years 2018-2019. How many landslides fall outside the rainfall data? Moreover, is also mentioned that part of this landslides associated to the FraneItalia database are triggered by anthropogenic causes or earthquakes. A better description of the landslide inventory is needed. In fact, this information is not so much relevant and is unnecessary since the work deals only with rainfall thresholds for landslide initiation. My suggestion is to consider only the landslides that were effectively used for the study. What are the landslide types? Are they shallow or deep-seated? That information should be better addressed and it’s important to understand the critical conditions defined for their occurrence, as largely explored in literature.
- Lines 56-60. In addition, regarding the rainfall data, how was determined the limit of 250 mm for considering an error in the rainfall record? What was the maximum hourly rainfall registered historically? It’s possible to address better what is an evident error /rain gauge malfunction? It’s possible to characterize the fraction of gaps and errors detected in the rainfall records based on visual inspection of the rainfall time series, and the reliability of the procedure? If an error in the rainfall records exist, and if the whole rainfall event surrounding the peak has been removed, what was the criteria to stablish the critical rainfall conditions for triggering the landslide associated to that rainfall event? If exists!
- The application used by Melillo et al (2018) it’s used to derive a set of variables to be used for computing rainfall thresholds for landslide occurrence. These variables are indicated in lines 67-68. Can authors address better the mean intensity and total depth variables and in what differs the total depth from the ID threshold? In addition, regarding those variables and thinking on triggering and non-triggering rainfall events how related with the rainfall events are the critical rainfall conditions (rainfall events that triggered the landslides), they exactly match? I’m asking this, because, if I understand well, the algorithm identified a variable number of rainfall conditions

responsible for the failures.

- Line 77. The 144 triggering rainfall events are defined for, I suppose, the 144 landslide events /landslide cases included in this study from the FraneItalia database. Please turn clear here and in the data section.
- Line 85: Could authors be more specific regarding the specific objectives of partition of the rainfall events into validation and test dataset? Are the groups similar with respect the distribution of rainfall events characteristics? How the partition considers triggering and non-triggering rainfall events? Please address better this issue.
- Line: 106-107 "Results from ANNs are compared with rainfall duration-depth power-law thresholds derived through the maximization of TSS analysing both triggering and non-triggering events". How was that done – derived by the CTRL-T software? Please turn clear in the methods section.
- Lines 109-115: I acknowledge the examples with other thresholds available in literature, but I think that differences/similarities could be additionally explored. Are thresholds based on the same datasets? For the same periods? For the same landslide types?
- Lines 119-123: This step is not suitable to be included in the methods section? If so, please adjust. In addition, what it's possible to conclude regarding this comparison by the fact that after 5 hours the threshold defined by equation 10 be above the ones defined by equation 7? A figure comparing those two thresholds could help understanding the idea.
- Lines 109 – 133: I strongly believe that aspects described in these paragraphs could be better placed in the methods section to become clearer the author's approach: e.g., the reference to the non-exceedance frequency for triggering events equal to 5%; the list of variable configurations and part of the subsequent descriptions (Lines 125-129).
- In my perspective the fact that the training datasets obtained less TSS than the validations and test datasets need additional discussion. I understand the approach used to preventing overfitting of the thresholds based on the training dataset, but is not supposed to be achieved better results when the predictive capacity is evaluated with the data partition used for training the predictive model? Being, in this sense, the independent validation with the dataset not used for training the model (validation partition) more robust and with general lower predictive results?
- The analysis of results expressed on table 1 (line 140), could be better explored in lines 144-155. Why the analysis is centred on results defined, if I understand well, considering no partition of the rainfall events dataset (TSS all)? Why not considering the more robust evaluation of the predictive performance of the rainfall-thresholds model, this is, with the validation dataset (TSS validation or TSS test), and what explanations could be attributed considering the TSS values obtained for the different variables configuration and for the training and validation datasets.
- A part of the conclusions section is suitable to be moved for the results and discussion section, particularly, the text after line 169 related with the drawbacks of the ANN approach.
- Abstract and conclusions should be adjusted accordingly.

Presentation quality:

Overall, the manuscript is well written. In my opinion, the structure needs some adjustments, by moving some text sections from results to Data and methods and from

Conclusion to Results and discussion. The manuscript presents generally a clear language that is understandable and scientifically precise. I made some small changing suggestions for the title and results analysis, even so, the abstract, the subtitles and the figures and tables captions are in general adequate. With respect to figures and tables they are not in large number, but present generally a reasonable quality and adequate to the purpose of the manuscript. Nevertheless, and in respect to those items described above some comments are made in the "Scientific Quality" section.

Technical comments (e.g., typing errors, format)

Some additionally technical comments are listed below.

Lines 22-23: consider adjusting the use of risk analysis in this sentence, since no exploration of consequences are illustrated by authors. Please see my previous comments about this literature examples using ANN for this manuscript in point 2 of my Scientific Quality comments.

Line 41. Adjust position of citation of figure 1. In the maps its not possible to see the 20 regions of Italy.

In figure 1: i) adjust the Europe map (upper right). Although recognizable it's not properly represented. Include some colour differentiation for the Atlantic Ocean/Mediterranean Sea and Europe countries. Include one or two country/ocean toponyms; ii) Eliminate from the figure the word "Legend" and replace "Rain gauges" by "Rain gauge". The differentiation of the rain gauges 2009-2018 and the landslides are sometimes not easy. Adjust caption: consider changing "and landslides from the..." for "and landslides dataset extracted from the FraneItalia..."

In figure 2 what represents the E-D threshold. "E" is used to describe the cumulation rainfall associated to the rainfall or landslide event? Please turns clear in figure and manuscript text. In addition, what differs from H depth in figure 2b from Total Depth (H) in Line 67?

In Table 1 verify the number of Table. Should be Table 1.

