Response to comment on nhess-2022-199

The manuscript presents an interesting application of methods for the definition of empirical rainfall thresholds for landslide occurrence at a national scale. The aim of the paper is clear and the results are also well-presented. Despite some points not very clear, I found the manuscript clear and sufficiently well-organized. From a methodological point of view, I found some problems in the work, which should be addressed before the paper can be reconsidered for publication.

I list in the following some general comments and a few specific technical corrections and other suggestions.

Comment response: Thank you very much for your review, we have tried to make it more clear and correct the problems in the work considering your comments, corrections and suggestions in the new version of the manuscript. We are very grateful and sure that each of your comments contributed to the improvement of our work. Additionally, this document is highly important for the scientific community related to landslides in Peru since this type of work has not been developed in Peru, which, in addition, faces the limited availability of data compared to other countries. Lastly, other investigations also faced similar difficulties (e.g., Kirschbaum et al., 2015; Abraham et al., 2019). In this sense, this study is the first to be carried out on a national scale in Peru and its objective is to support the operational monitoring system of shallow landslides in Peru (https://www.senamhi.gob.pe/?p=monitoreo- silvia), and since our institution (SENAMHI) is responsible for maintaining this system, this work will contribute to giving it scientific validity, understanding its limitations but which will continue to be improved over time.

Response to General comments

The main problem of the work lies in the validation procedure. In particular, the use of only one year of data as validation set is inconvenient. This choice was proved to be not effective cause is too much linked to the variability of the selected year. Indeed, you found that the performances decreased in the validation, "which may be due to the fact that, in the year 2020, there were no extreme rainfall events as in other years, and the number of landslides was lower than in other years". A more reliable procedure would consider a random selection of triggering and non-triggering rainfall conditions in a calibration (e.g. 80% of the total) and a validation set (remaining 20%). You can found examples in: https://doi.org/10.1007/s11069-019-03830-x or https://doi.org/10.5194/hess-25-3267-2021

Comment response: Thank you very much for the observation, this is one of the discussions that we added taking into account your comments and observations. We take this methodology that has already been used in other investigations (e.g., Dikshit et al., 2019; Kirschbaum et al., 2015), however, as we conclude, it did not obtain good results for few data, so we add this discussion so that be taken into account in future research in Peru. We have added your observation to the new version of the manuscript, as you can see below:

"The calibration/validation methodology, based on take one year of observations for validation set, which was used in other research works (e.g., Dikshit et al., 2019;

Kirschbaum et al., 2015), is quite short and there is the risk of overinterpretation. It is therefore highly recommended for future research to expand the dataset and explore other calibration/validation methods, for example, a random selection of the differentiated data set for the calibration and validation (e.g., 70% for calibration and 30% for validation) (Brunetti et al., 2021; Gariano et al., 2020)."

In addition, in our future research we hope to advance in these limitations in Peru, for example, our perspective is to expand the database, for which we are working with INDECI (entity in charge of the attention of the population when landslides occur) for future studies that include greater data extension and include the random selection of the dataset.

--

The use of daily rainfall data is also not the best choice for defining rainfall thresholds, particularly for shallow landslides, given the high uncertainties related to this temporal resolution as highlighted by https://doi.org/10.1007/s1106 9-018-3508-4 https://doi.org/10.1007/s11069-019-03830-x. This should be pointed out and discussed better. I would add that there are currently other satellite-based rainfall products with better temporal resolutions (e.g GPM), which could be employed in such analyses.

Comment response: Thanks a lot for the observation. We agree that more exact thresholds could be defined with sub-daily rainfall data, however we chose to use these daily rainfall data for different reasons, the first is that this work is the first approximation of regional rainfall thresholds in Peru from from which new and better thresholds will be generated, in addition to the fact that we take into account different investigations that developed thresholds from daily rainfall data (e.g. (Berti et al. 2012; Kirschbaum and Stanley 2018; Leonarduzzi and Molnar 2020; Leonarduzzi, Molnar, and McArdell 2017; Monsieurs et al. 2019), in addition to the fact that these thresholds have the objective of improving landslide monitoring services triggered by rainfall that already exists in Peru (https://www.senamhi.gob.pe/?p=monitoreo-silvia) and that our institution, SENAMHI (the hydrometeorological service of Peru), is responsible for monitoring and improving it. Finally, until 2017 we used TRMM data for our hydroclimatic services, however, for a period of time the TRMM data was not maintained, and all our hydrological services that depended on this data had to stop, for this reason, at SENAMHI we choose to generate operational data (PISCO) that takes into account the assisted climatology data (e.g. PISCO monthy mean) but does not depend of external data base.

Currently, as SENAMHI we are also focused on the generation of hourly rainfall product (e.g. Huerta et al., 2017 https://doi.org/10.1016/j.dib.2022.108570), but that it be updated in real time for our monitoring services, with which our next investigations regarding thresholds will take into account these hourly data.

--

The whole section 2.4 misses several information and needs a check and a huge review.

Comment response: Thanks a lot for the observation. We rewrite and reorganize the entire section for better understand, as you can see below.

"2.4 Rainfall threshold model

An empirical–statistical approach was used to define rainfall thresholds for landslide-susceptible regions, consisting of the following steps: (1) determination of rainfall events from a historical rainfall series, (2) definition of the variables of rainfall events, (3) define landslides regions from maximum daily rainfall region and GEOGloWS basins for the area studio, (4) threshold estimation for individual rainfall event variables for calibration period based on an objective maximization of predictive performance, (5) threshold estimation for combination of rainfall event variables for calibration period based on an objective maximization of predictive performance, and (6) run thresholds models and get metrics for analysis and discussions 110 (methodology is presented in Figure 2). Below are the details of the method.

The first step was the construction of a historical rainfall series from gridded rainfall data (PISCOpd_Op) for each basin that had a minimum of one landslide event. After obtaining the rainfall series, rainfall events were defined along with a historical series for each selected basin. For this work, we define an independent rainfall event as a series of consecutive rainy days where it has rained above a minimum rainfall threshold (Figure 3). Many authors use minimum thresholds of 1 mm to define rainy days (Dai, 2006; Dai et al., 2007; Han et al., 2016; Leonarduzzi et al., 2017; Shen et al., 2021; Tian et al., 2007; Yong et al., 2010). However, given the great climatological spatial variability in the study area, it was determined that there was not a single minimum threshold for the entire territory, but a minimum threshold was discretized from the bias of PISCOpd_Op for non-rainy days. The PISCOpd_Op bias was determined when rain gauges did not report rain (0 mm), and the discretized minimum threshold (Umin) of rain was defined according to the following Equation 1:

$$U_{min} = \begin{cases} U_0 & \text{if } s \leq U_0 \\ s & \text{if } s > U_0 \end{cases} \quad (1)$$

where s is the average of simple bias when rainfall stations reported a value of 0 rainfall compared with the estimation in PISCOpd_Op. And U0 is the initial minimum rainfall threshold, and it is established as 1 mm for all regions with exception of coastal Pacific regions which is considered 0.5 mm. Once rainfall events were defined, whether they were triggering or non-triggering events was established. A rainfall event is considered a rainfall trigger event if it is associated with a landslide event, i.e., if during the duration of the rainfall event a shallow landslide has occurred.

The second step was to determine analysis variables for each rainfall event, for which the maximum daily intensity Imax (mm/day), the accumulated rainfall E (mm), the duration D (day), and the mean daily intensity Imean = E/D (mm/day) were calculated. Concerning the triggering rain events, two scenarios were considered. For the first scenario (entire event - EE), the properties of the rainfall event (Figure 3) were defined considering the rainfall rate of the landslide occurrence day. The second scenario (antecedent event - AE) defined the properties up to one day before the occurrence, i.e., it did not consider the rainfall rate of the landslide occurrence day. The reason for analyzing the second scenario was to evaluate the level of incidence that is attributed only to antecedent conditions for landslide occurrence, as this allows us to evaluate if it is possible to forecast or warn landslides based only on the antecedent conditions. The temporal evolution of hydrometeorological variables provides an idea of how the critical conditions of the activation of landslides develop (Prenner et al., 2018; Segoni et al., 2018).

The third step consisted in divide the study area into regions based on clustering techniques (this step is explained in more detail in the section 2.5). Next, GEOGloWS basins were merged with regions in order to determine their spatial correspondence. The fourth and fifth step was to objectively select a rainfall threshold that separates triggering rainfall events from Non triggering rainfall events with the best level of predictive performance. Rainfall thresholds were established by maximizing predictive performance in two ways: the first one only included variables independent of rainfall properties (Imax,E,D, Imean), and the second one determined was through curve-like thresholds that related two properties (Imax −D,E −D, Imean −D) in the form of V = a. D−b, where V represents the variables Imax, E, and Imean; a and b are the scale and shape parameters of the curve (while for logarithmic space, a is the intersection parameter and b denotes the slope of the linear curve). Finally, the sixth step consisted in apply the model to the rainfall events and compare with the observed landslides events and get the predictive performance metrics for each region at calibration and validation periods."

It is not clear how the association between a rainfall event and a landslide is done (Line 120), in order to classify an event as a triggering rainfall event.

Comment response: Thanks for the observation. A rainfall event it considered as triggering event if during the duration of the rainfall event a shallow landslide was occurred. We edit the sentence for better understand, as you can see below.

"A rainfall event is considered a rainfall trigger event if it is associated with a landslide event, i.e., if during the duration of the rainfall event a shallow landslide was occurred."

Moreover, at line 118 it is reported that "For coastal Pacific regions, 0.5 mm was considered the minimum rainfall threshold". What about the other regions?

Comment response: Thanks for the observation. The minimum rainfall threshold considered for other regions is 1 mm, and only for the coastal Pacific region is 0.5 mm. We edit the equation and sentence for better understanding, as you can see below.

"… where s is the average of simple bias when rainfall stations reported a value of 0 rainfall compared with the estimation in PISCOpd_Op. And U0 is the initial minimum rainfall threshold, and it stablished as 1 mm for all region with exception of coastal Pacific regions which is considered 0.5 mm."

At lines 131-136, it is not clear the actual method used to define the thresholds, based both on 1 or 2 variables. How the parameters and the equations were obtained? Before "maximizing predictive performance" a threshold should be calculated using a method. Which method was used? This issue needs to be better explained.

Comment response: Thanks for the observation. The paragraph was edited and corrected for better explain, as you can see below:

"... Rainfall thresholds were established by maximizing TSS predictive performance in two ways: the first way includes every rainfall event property independently (Imax, E, D or Imean), and the second one determined was through curve-like thresholds that related two properties (Imax − D, E − D, Imean − D) in the form of V = a.D^{-b}, where V represents the rainfall properties (Imax, E, and Imean); a and b are the scale and shape parameters of the curve (while for logarithmic space, a is the intersection parameter and b denotes the slope of the linear curve). The approximation of the first form, thresholds based on only one of the rainfall event properties (Imax, E, D or Imean), was estimated whit the minimum radial distance to the perfect classificatory test (TSS=1, with se=1 and 1-sp=0) from the ROC space (e.g., Uwihirwe et al.; Gariano et al.) and the approximation of the second form, curve-like thresholds, was established with the optimization of a and b parameters of the curve model (V = a.D^{-b}) with an initial approximation of the curve based on a=average of the variable V of the triggering rainfall events and b=0. ...."

Moreover, at line 133 is written "variables independent of rainfall properties (Imax,E,D, Imean)"; actually, Imean and D are not independent on each other, being Imean=E/D. Please explain also this point.

Comment response: Thanks for the observation. We refer as variable independent to only one of the rainfall event properties (Imax, E, D, Imean). The sentence was corrected for better understand as you can see below:

"… the first way includes every rainfall event property independently (Imax, E, D or Imean), …"

Finally, I believe that proposing thresholds based only on one parameter (e.g. E, D, Imean, or Imax) is now anachronistic, given the huge literature on rainfall thresholds based on two variables.

Comment response: Thanks for the observation. We agree that there is a large amount of literature on thresholds based on two variables, although there is also literature that evaluates one-parameter variables and/or how they impact when combined with other variables (Hirschberg et al. 2021; Leonarduzzi et al. 2017 ; Uwihirwe, Hrachowitz, and Bogaard 2020), in this sense, our approach, being a novel work in Peru, is to provide variables that could be beneficial for certain regions and in future research combine or improve them in greater detail. Additionally, this paper is highly important for the scientific community related to landslides in Peru since this type of work has not been developed in Peru, which, in addition, faces the limited availability of data compared to other countries.

Regarding the thresholds based on two variables, actually there is no need to calculate both E-D and Imean-D thresholds, given that they are analytically equivalent, being Imean=E/D. I can't figure out how different results are obtained for the two types of thresholds (I-D and E-D); they should have the same performaces).

Comment response: Thanks for observation. We have taken into account your comment. According to the way we have defined the variables for a dataset, Imean, that is affected by D, does not have the same distribution as E. For example, two events with the same E (e.g. E=10), can have different D (e.g. D equal to 2 and 4 days), therefore, the Imean of both resulting events are different (Imean equal to 5 and 2.5 respectively), so the threshold could not be defined as the division of both. The next Fig. X1 shows what is mentioned for an example dataset, where it is observed that E and Imean have different density distributions and therefore their predictive potentials also change (i.e., the thresholds do not have the same Imean relationship =E/D).

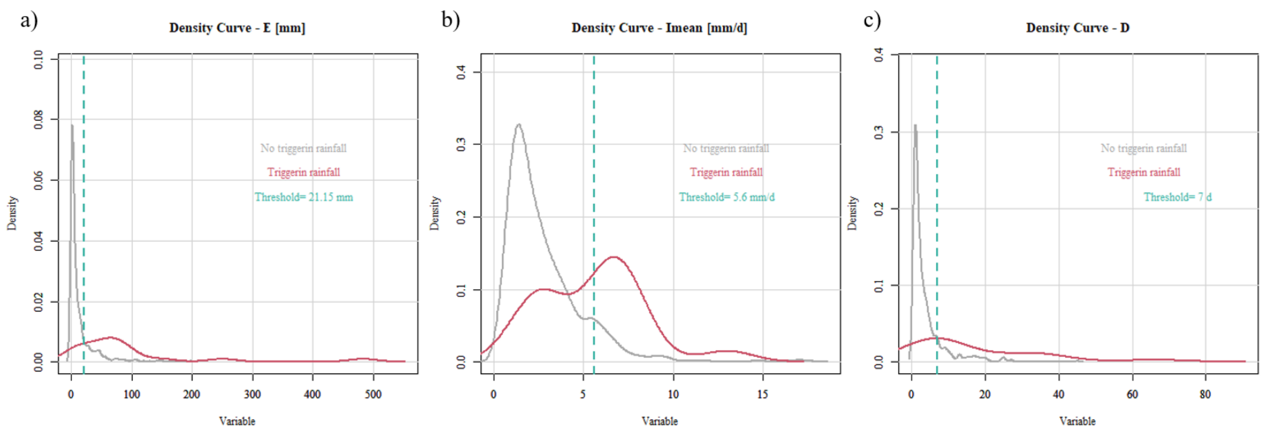

Fig. X1: Density plot of the variables E (a), Imean (b) and D (c) for the same data set, where it is observed that the distributions of the variables E and Imean are different.

--

Line 170: actually, a threshold is represented by a point in the ROC space (the point is the TRP, FPR couple), so I believe that the area under curve is only a quadrangle. Please explain better this point. Being the thresholds represented only by one point in the ROC space, I would suggest using the distance of this point from the perfect classificatory point (upper left corner of the space, TPR=1, FPR=0) instead of the area under curve. You can find more details in https://doi.org/10.1016/j.geomorph.2014.10.019

Comment response: Thanks a lot for the suggest and clarification of the topics. We use this method to estimate the thresholds of one variable, and we edit and correct the paragraph and explanation of the calibration methods for better understanding, as you can see below:

"For thresholds based on rainfall event properties independently (Imax, E, D or Imean), the overall impression of the predictive power was estimated whit the so-called receiver operating characteristic (ROC) curve (Fawcett, 2006), from which the minimum radial distance to the perfect classificatory test (TSS=1, with se=1 and 1-sp=0) was used to select the individual variable threshold (e.g., Uwihirwe et al.; Gariano et al.) …"

Lines 179-182: actually, more simple and useful definitions are: TPR = TP/(TP + FN); FPR= FP/(FP + TN). I would suggest using these definitions instead of mentioning sensitivity and specificity.

Comment response: Thanks for the suggest. We use these definitions as we review that they were also used in many other current publications, but we have edited and added these citations for better understanding, as you can see:

"Some of the most common measures for landslide forecasting are the sensitivity (se = TP/(TP + FN)), specificity (sp = 1 − FP/(FP + TN)) and true skill statistic (TSS = se + sp −1) (e.g., Staley et al., 2013; Gariano et al., 2015; Leonarduzzi et al., 2017; Mirus et al., 2018; Leonarduzzi and Molnar, 2020; Hirschberg et al., 2021).

… The benefit of using the specificity over the false positive rate (FPR=FP/(FP+TN)) is that in a perfect model TSS, sensitivity and specificity all equal 1 (Hirschberg et al., 2021)."

Passing to Section 3, regarding the regionalization, it is not clear how many empirical points are employed for calculating the thresholds in each of the 11 regions. Please add this information and discuss possible limitations in case of thresholds based on too few points.

Comment response: Thanks for the observation. We add this information, and add a discuss on the new version of the manuscript, as you can see below:

"Hirschberg et al. (2021) found that 25 events are enough to limit the uncertainties in the ID threshold parameters to ±30% in his study, based on this, it is observed that there are several regions (Andes 3, 5, 6 and Amazon 1, Amazon 3 and Pacific 2) that do not exceed that quantity, so these regions have a greater source of uncertainty due to the quantity of the data. A summary of the number of shallow landslide events used for the research and the thresholds with best performances per region is presented in Table

3.”

Table 3. Number of SL events and best thresholds for one and two variables for each region (Th: threshold, SL: number of landslides per region, Cal: Calibration, Val: Validation)

Region	SL total	SL Cal	SL Val	Best Th - 1 variable	TSS	Best Th - 2 variables	TSS
Pacific 1	46	43	3	I_{max}	0.68	$I_{max} - D$	0.71
Pacific 2	27	20	7	I_{mean}	0.61	$I_{mean} - D$	0.61
Andes 1	34	28	6	I_{mean}	0.43	$I_{mean} - D$	0.44
Andes 2	98	83	15	E and I_{mean}	0.58	$I_{max} - D$	0.64
Andes 3	17	10	7	I_{max}	0.92	$I_{max} - D$	0.91
Andes 4	65	54	11	E	0.51	$I_{mean} - D$	0.52
Andes 5	14	7	7	E	0.67	$I_{mean} - D$ and $E - D$	0.66
Andes 6	4	3	1	D	0.68	$E - D$	0.65
Amazon 1	6	6	-	I_{mean}	0.74	$I_{mean} - D$	0.77
Amazon 2	54	41	13	E	0.57	$E - D$	0.58
Amazon 3	12	10	2	E	0.68	$I_{mean} - D$ and $I_{max} - D$	0.73

Figure 6. Please note that the thresholds should have duration ranges based on the minimum and maximum durations of the triggering events. Theoretically, you can’t draw a threshold in a duration value when you don’t have a triggering event. This allow also avoiding having very low values of thresholds at long durations (see thresholds for Andes 4, 5, 6). Moreover, I would suggest correcting all the equations replacing Y and X with Imean and D, and replacing the “^” with a proper superscript.

Comment response: Thanks for the observation. The figure was edited taking in account your suggestions on the new version of the manuscript as you can see below:

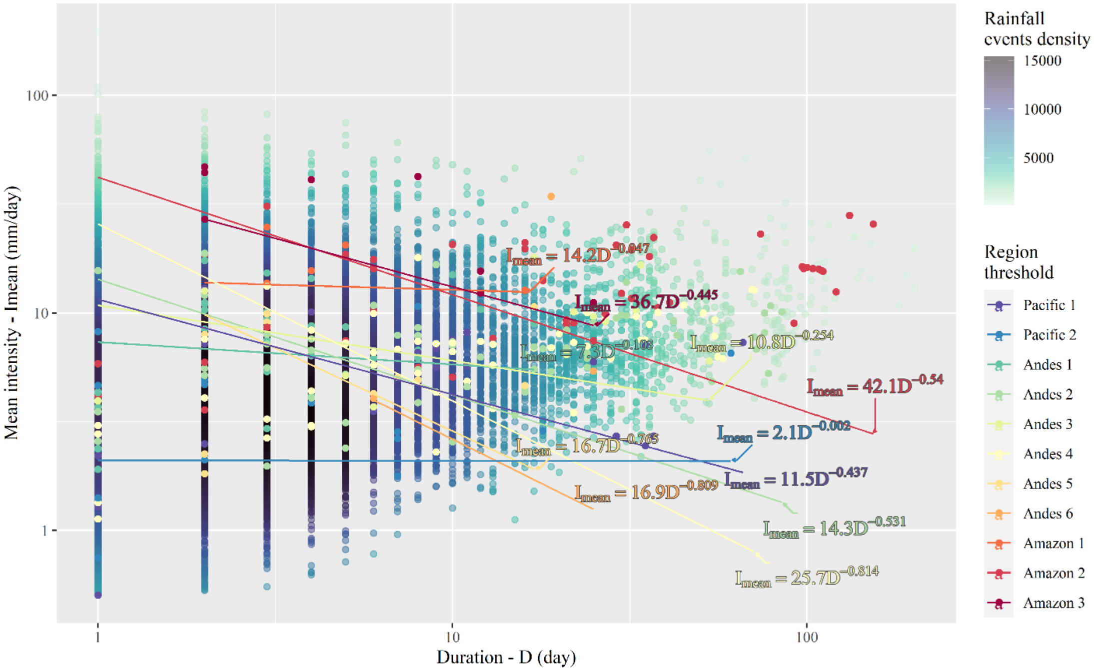

Figure 7. Is there some physical explanation for the variation of the values of the 1-variable thresholds? In some cases, I see differences that seem not related to morphology or other environmental factors.

Comment response: Thanks for the observation. We have taken into account your observations and recommendations and have included them in the discussions of the new version of the manuscript, as you can see below:

“Regarding the variability of the thresholds, we can explain it mainly to the rainfall climatology in Peru. It can be seen that the magnitudes have a relationship with respect to the spatial distribution of rainfall in Peru, that is, low thresholds related to rainfall of lesser magnitude in the arid zones in the western part of Peru (Pacific region), thresholds intermediates related to the increase in the magnitude of rainfall in the middle part or mountainous region (Andes region) and the highest thresholds related to wet regions (Amazon region). However, the Andes 1, Andes 3 and Andes 6 regions do not have this relationship, so this discussion is not conclusive and is considered to be related to limited data, so it is suggested that this variability be discussed in future research that include more shallow landslides events data.”

Response to technical corrections and suggestions

Abstract: I would use the present tense in the abstract

Comment response: Thanks for the suggest. It was edited on the new version of the manuscript, as you can see:

“Abstract. The objective of this work is to generate and evaluate regional rainfall thresholds obtain from a combination of high-resolution gridded precipitation data (PISCOpd_Op), developed by the National Service of Meteorology and Hydrology of Peru (SENAMHI), and information from observed shallow landslide events. The landslide data were associated with rainfall data, determining triggering and non-triggering rainfall events with rainfall properties from which rainfall thresholds are determined. The validation of the performance of the thresholds is carried out with events that occurred during 2020 and focus on evaluating the operability of these thresholds in landslide warning systems in Peru. The thresholds are determined for 11 rainfall regions. The method of determining the thresholds is based on an empirical–statistical approach, and the predictive performance of the thresholds is evaluated whit the “true skill statistics” (TSS). The best predictive performance is the mean daily intensity-duration (Imean − D) threshold curve, follow by accumulated rainfall E. This work is the first attempt to estimate regional thresholds on a country scale in order to better understand landslides in Peru, and the results obtained reveal the potential of using thresholds in the monitoring and forecasting of shallow landslides caused by intense rainfall and in supporting the actions of disaster risk management.”

I would use rainfall instead of precipitation everywhere in the text.

Comment response: Thanks for the suggest. We use rainfall instead of precipitation on the new version of the manuscript.

Line 24: “Terrain saturation is the original cause of landslide occurrence”. Actually, this depends on the type of landslides.

Comment response: Thanks for the observation. This sentence it was removed for better understand according the comment.

Line 33: perhaps the correct reference is Segoni et al 2018 (already mentioned), not Segoni et al 2014

Comment response: Thanks for the observation. The correct reference is Segoni et al., 2018, and it was corrected in the new version of the manuscript.

Line 36-39: The sentence “For example, global thresholds have been developed based on antecedent precipitation indices (Caine, 1980; Guzzetti et al., 2008; Kirschbaum and Stanley, 2018), and national thresholds have been established under an empirical–statistical approach (Leonarduzzi et al., 2017; Peruccacci et al., 2017a; Uwihirwe et al., 2020).” is not correct. Actually, both the mentioned thresholds based on antecedent precipitation and the cited national thresholds are established using and empirical-statistical approach. Please review and correct.

Comment response: Thanks for the observation. The text has been rephrased in order to clarify the main idea, as you can see below.

“For example, there is been developed empirical–statistical approach to the estimation of global thresholds (Caine, 1980; Guzzetti et al., 2008; Kirschbaum and Stanley, 2018), and national thresholds (Leonarduzzi et al., 2017; Peruccacci et al., 2017a; Uwihirwe et al., 2020).”

Line 38: Note that there are two references to the work “Peruccacci et al. (2017)” a and b, which are actually the same.

Comment response: Thanks for the observation. The reference has been corrected.

Line 47: I would suggest using “slope” instead of “hillside”

Comment response: Thanks for the observation. The text has been changed, as you can see below.

“Thresholds can be set for different spatial scales depending on the extent of the analysis, and these can be categorized into six classes: global, national, regional, basin, local, and slope scales. …”

Line 51: in relation to environmental subdivisions within a national territory, please consider also the work of Peruccacci et al. (2017) – already mentioned – which present several morphological, geological, meteorological, climatic subdivision of the Italian territory with the aim of defining rainfall thresholds.

Comment response: Thanks for the suggestion. The reference was added, as you can see below.

“…, as well as environmental subdivision within a national territory based on erodibility and climatology represented by the maximum daily intensity of a rainfall event (Leonarduzzi et al., 2017) or on topography, lithology, land-use, land cover, climate, and meteorology (Peruccacci et al., 2017).”

Caption of figure 2. Delete “Methodology six steps”

Comment response: Thanks for the observation. The caption has been corrected.

Line 101: I suppose you wanted to write “is shown in Figure 2”.

Comment response: Thanks for the observation. It is Map Figure (Fig. 2), and it was edited on the new manuscript.

Line 185-186: please check syntax and grammar.

Comment response: Thanks a lot for the observation. I checked the syntax and grammar and simplified the paragraph, as you can see below.

"For thresholds based on independent variables (Imax, E, D, Imean), the overall impression of the predictive power was estimated whit the so-called receiver operating characteristic (ROC) curve (Fawcett, 2006), from which the minimum radial distance to the perfect classificatory test (TSS=1, with se=1 and 1-sp=0) was used to select the individual variable threshold (e.g., Uwihirwe et al.; Gariano et al.) while for the threshold curve (Imax−D, E−D, Imean−D) the scale parameter a and the shape parameter b are simultaneously tuned to maximize the the true skill statistics (TSS) (e.g., Leonarduzzi et al.; Hirschberg et al.). This maximization was carried out automatically using the shuffled complex evolutionary algorithm (SCEA-UA) (Duan et al., 1993), considering the TSS as the objective function. The methodology was applied for each region within the analysis area, finding different thresholds for each of them."

Line 196: actually, TSS varies between -1 and 1, as you correctly mentioned some lines above.

Comment response: Thanks a lot for the observation. It was edited in the manuscript.

Table 3: I would suggest using always the same number of decimal places.

Comment response: Thanks a lot for the suggestion. It was edited in the manuscript using 2 decimal places.