The study examines the performance of rainfall thresholds for landslides obtained from different hourly and daily datasets, as well as the use of normalizations for the threshold localization and some preliminary analyses on the impact of antecedent conditions. The manuscript is well written, the study well fits the topics of this journal and is carried out with sound methods and data. To my view, the novelty brought by the study is that it collects from an amount of recent theoretical and smaller-scale studies and collectively examines the practical implications using a large dataset on a wide alpine region. I believe it contributes to our practical knowledge on rainfall thresholds for landslides triggering and therefore deserves publication. Overall, it was a pleasant reading.

I list below a few comments for the author's consideration.

Kind regards,

Francesco Marra

We thank Francesco Marra for the review and the constructive comments. We address here all the points raised in the review.

1. In Section 3.4, did you check the results using absolute quantiles (corresponding to return levels, or probabilities in time) instead to wet quantiles? To my view, the number of wet days contributes generating the local climatology (indeed it does for return levels) and the wet-quantiles somehow forget this. I refer in particular to the discussion in lines 281-291, which I believe would hold more for absolute quantiles. Also, it would be very interesting to see fig 6 with absolute quantiles (return levels) instead of wet-quantiles. I'm not saying this must be included, rather that it should be checked before exclusion (even though I'd be personally interested in seeing the figure anyways)

It is true that return periods and rainfall probabilities should consider both intermittency (frequency of rain) and rain intensity, so the question raised is an important one. We show the results for the normalisation using both "wet" and "absolute" quantiles in Figure 1 here below. Please note that to simplify the plot here, we removed the event properties line, which is present in the original Figure 5 in the manuscript. The "wet quantiles" line matches the "rainfall" line in the original manuscript's Figure 5. We tried to limit the number of normalisation factors considered for simplicity and clarity, and chose wet quantiles because only high absolute quantiles were usable (i.e. > 0 for all susceptible cells, from 0.65 quantile for daily rainfall and 0.9 for hourly rainfall).

As can be seen from Figure 1, using absolute quantiles doesn't seem to differ much from the wet quantiles when using daily rainfall (left side of Review-Figure 1) for the highest quantiles (much worse for lower quantiles). When using hourly rainfall, absolute quantiles do seem to outperform wet quantiles and, in the case of total rainfall, also the normalization using mean annual precipitation. Nevertheless, because the normalization with the MAP seems overall quite robust, and MAP is generally an easily available climatological variable, we still would recommend using MAP as a normalization parameter. We will consider replacing Figure 5 in the manuscript with one that includes the wet quantiles normalization also.

2. I suggest including ID/ED thresholds in the results in Fig. 3. Many readers are familiar with such thresholds and it would be helpful for the quantitative interpretation of the results

We agree that ID/ED are the type of threshold readers are more familiar with, nevertheless the idea of Figure 3 is to be complementary to Figure 2. The optimal ED curve parameters as well as the corresponding TSS are reported in Figure 2 for all the different dataset considered (same cases as in Figure 3, and more). We believe that adding also the ED results in Figure 3 would make it less readable as it would require two extra entries (a and b parameter of the curve in the bottom part of the figure) and a different y axis. In order to facilitate the reader in combining the two figures, we added in the legend of Figure 3 the number matching it with the corresponding case in Figure 2. We will consider adding the TSS values for the ED thresholds in the first 2 panels of Figure 3 in the revised manuscript.

3. Do your archives contain information on the landslides type? Are debris flows included in the database? I would guess that debris flows, generally triggered by short convective events, are more subject to the temporal resolution. If relevant, is there a way to check this from your data? Also, in the discussion (lines 355-356) you recommend "not to extend daily thresholds … into the sub-daily domain" – can this recommendation be made more explicit from the elements in your hands?

The Swiss flood and landslide damage database which we used here contains flood, debris flow, landslide and rockfall events. We used only those classified as "landslides", therefore debris flow shouldn't be included (although the classification is based on the "primary process", so debris flow might have occurred also as secondary process).

The recommendation of not extending daily thresholds to the subdaily domain we report is actually a conclusion from Gariano et al. (2019), as mentioned in line 355, and support our belief that hourly thresholds should not be derived (i.e. extrapolated) from daily resolution. Furthermore, thresholds obtained with daily or hourly data cannot be compared since strong and unrealistic assumptions must be made to allow the comparison (as mentioned in the manuscript in the lines just prior). A weak confirmation of this is that the optimum thresholds when considering the 2003-2005 timeframe and only landslides with known date and timing for mean intensity are 1.34mm/h and 14mm/d (definitely not a 24 factor). Both aspects are not really a conclusion drawn from any analysis carried out here, but rather a general recommendation we believed was important.

4. The title focuses on the temporal resolution aspect while the paper provides quite a lot of additional information. Perhaps you can consider expanding it

We are considering several options to improve the title, as this was brought up by both referees. Working title: "Deriving rainfall thresholds for landsliding at the regional scale: daily and hourly resolutions, normalization and antecedent rainfall"

5. Lines49-54: More details on the point (b) (i.e. poor matching of landslide in space) should be provided in the introduction as this is a crucial to the findings. There are few lines afterwards but I think some (more) lines are needed in the introduction as well

In the introduction, we list 3 different undesired consequences of choosing higher temporal resolution rainfall in the context of landslide forecasting with rainfall thresholds. We later expand on the issue of rain gauge density in section 3.3.

6. Line 58: what do "(analysis steps)" refer to?

It refers to the two aspects mentioned right after (normalization and use of antecedent wetness). We removed "(analysis steps)" to simplify

7. Line 97: could you provide more details on the optimization (what was optimized, how, why)?

We modified the sentence to: "We choose 24 hours for daily rainfall data and 6 hours for hourly rainfall data. The hourly interstorm period of 6 hours separating events is selected as the one leading to the best performances (highest True Skill Statistic, see methodology explained hereafter), within a range of 2-12 hours, which is the amount of dry hours expected to separate individual storms. This is longer than the requirement of statistical independence between events, which Gaal et al. showed to be at least 2 hours. This difference reflects the role that antecedent rain plays in landslide generation."

8. Line 181-182: "Averaging…" this sentence was not clear to me. Also what do you mean by "trends"?

The idea is that because we're considering an entire (heterogeneous) country, there are differences which don't allow to see on an event scale the signal in the antecedent wetness (which is the reason for example why the MAP normalization improves the results). As an example of this, following the approach developed in Chleborad 2003 for the Seattle area using triggering and antecedent rainfall (see also Review-Figure 2), the data across the whole of Switzerland don't show any sort of clustering or threshold behaviour as it does on the local scale (mentioned in lines 299-301 of the manuscript).

We modified the sentence to: "Averaging the antecedent rainfall allows us to see general tendencies (i.e. whether antecedent wetness is generally higher or lower), not focusing on every individual event." Reading this and the comments from Ben Mirus, we also decided to expand the explanation of the methodology for studying the information content of antecedent rainfall.

9. Line 195: it looks like RDI retains good predictive power because of the stations density, is this correct? Are there other reasons?

In general, the good performances of RDI are due to its high quality (station density, interpolation method), which definitely play a role also here. But in this case we're also suggesting that the signal at the daily scale might be stronger than expected. In fact, we're removing the advantages of the daily resolution we mention here (longer record, more landslide events), but still the difference in performances compared to hourly data is not dramatic. Nevertheless, we cannot quantify this effect exactly, and separate how much of the performances at the daily scale are due to the higher station density or better interpolation (higher data quality).

10. Line 215: perhaps I did not understand: why does the sparseness of the points in the figure imply lower robustness? It this not just a consequence of the data sample?

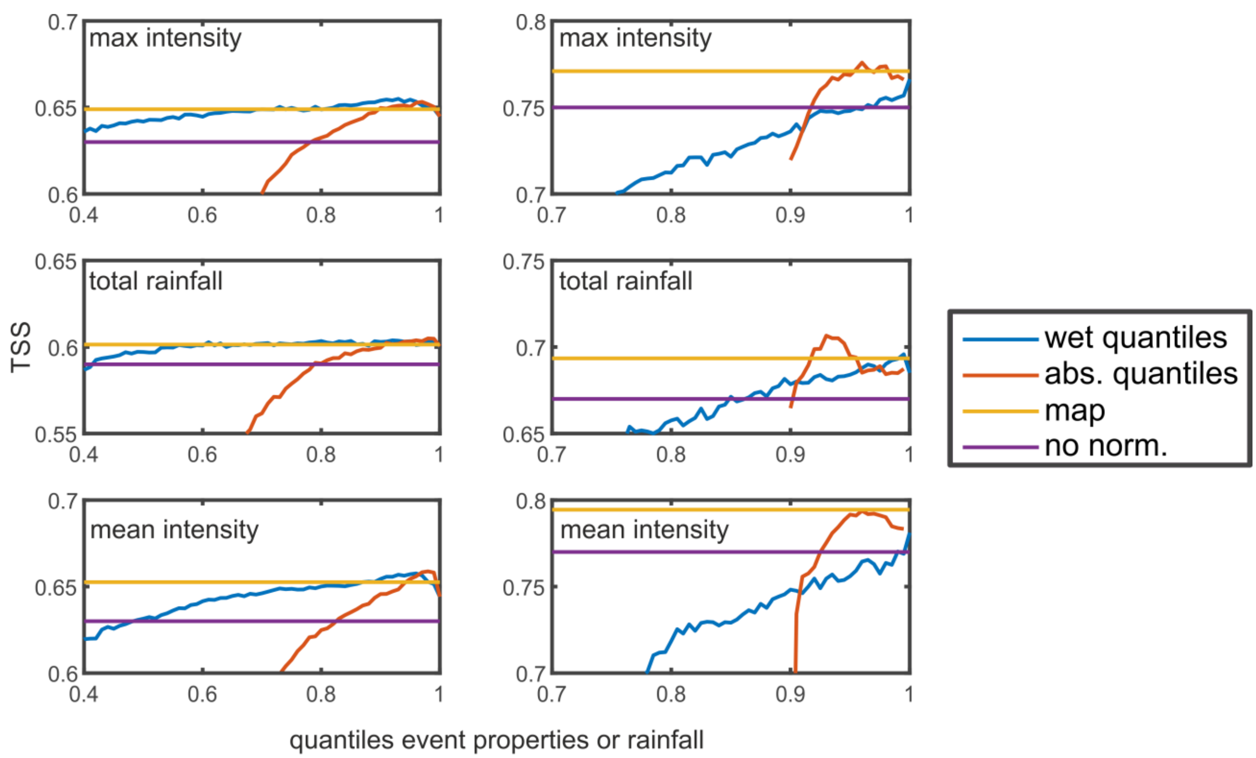
The sparseness is definitely a consequence of the sample size. It just visually shows how, when the size is so small, a few "wrong" points could affect the results and change the thresholds. That the robustness depends on the sample size (especially with uncertain data) is a pretty obvious concept, but we thought visualizing the effect could help the reader grasp the issue and its consequences.

11. Line 323-329: I'd argue that Marra 2019 do not claim/confirm that higher resolutions are superior as no evaluation of the predictive performance was done. Rather, systematic differences are highlighted, with consequences for the physical interpretation of the triggering amounts and the quantitative comparison of thresholds and threshold parameters obtained from different datasets

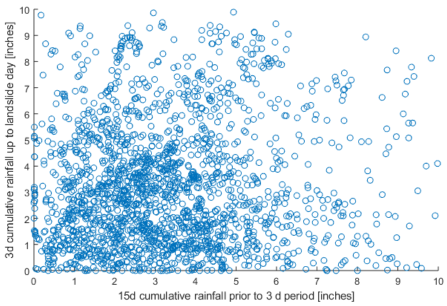
Changed to: Previous studies (e.g. Marra 2019 and Garaiano et al., 2019) have focused on the effect of temporal resolution, and showed that using lower temporal resolutions leads to the underestimation of the thresholds. From a theoretical point of view, we argue that hourly rainfall data are superior to daily data as they can capture the short convective events lasting few hours which are known to trigger landslides and which get averaged out in the daily sum. Also in the work presented here, when we consider the exact same time period and landslide events, we see that performances at the hourly temporal resolution are superior to those at the daily resolution, especially for high quality datasets (RHIR).

12. Fig. 2 and 3 took me some time to understand. I could not find suggestions on how to make them more immediately understandable, but I feel it is something to communicate within the review

Hopefully the improvement in linking the two mentioned Figures in point 2 above will help with the understanding. We also added a reference in the titles of the subplots in Figure 3 to the comparisons in Figure 2.



Review-Figure 1 . True Skill Statistic values for the best threshold for the different normalisations, for the daily (RDI, left) and hourly (RHIR, right) rainfall data. On top for maximum rainfall, in the middle for total rain, and the bottom for mean intensity. For the normalisation by wet and absolute quantiles of rainfall, the TSS is computed for each 0.01 quantile value (x axis). For the normalisation by mean annual precipitation (map) and the TSS value of the variable without normalisation (no normalisation), the constant value of the TSS is indicated as a straight line across all x values.



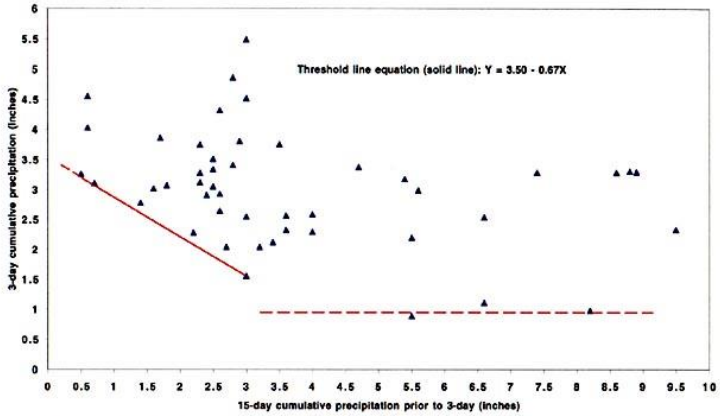


Figure 2a. Preliminary graph showing estimates of 3-day and prior 15-day cumulative precipitation associated with historical landslides that were part of events with 3 or more landslides in a 3-day period, in Seattle (filled triangles). The solid red line is a lower-bound threshold (visually identified) for the initiation of landslides when the 15-day cumulative is 3.0 inches or less. The dashed horizontal line is a lower-bound threshold, that was tentatively proposed, for conditions of 15-day antecedent precipitation exceeding 3 inches. (Chleborad, 2000).

Review-Figure 2. Left, the cumulative 3 days rainfall up to landslide days and the corresponding prior 15-days cumulative rainfall, following the approach in Chleborad (2003). Right, the original figure from Chleborad (2003) for Seattle Area. Rainfall amounts are reported in inches to facilitate comparison.