We provide here our reply to the comments from the Editor and the referees, together with the changes we plan to apply to the manuscript in response. The comments are in black fonts and our reply in blue.

Editor

Dear Authors,
Thank you very much for responding to the reviewers' comments in detail. Following your responses, some other aspects are of concern to the reviewers. I would appreciate it if you could do the necessary minor changes suggested by Reviewer 2. If you could quantify the influence of the uncertainty as indicated by Review 1, that would be greatly appreciated. Following your responses, I will resend the manuscript to the reviewers and ask for another round of recommendations.
Kind regards
Ugur Öztürk
NHESS Topic Editor

We would like to thank the editor for the prompt handling of our manuscript - much appreciated - and for the suggestions.

We are glad to see the positive evaluations from the referees. Despite this, after reading their comments, we realized that our main message was still somehow under-appreciated. We briefly clarify it here before getting into the point-by-point reply.

Our message concerns a theoretical issue that was mostly overlooked: the confusion between *conditional probabilities* and *marginal* (or *unconditional*) *probabilities*. We then provide a practical example, with the goal of making this theoretical issue more tangible. Most of the referee's attention is attracted by this last example, but they fail to focus on the real message - the theoretical one.

With this in mind, we rephrased several parts of the manuscript to better guide the attention of the readers on this aspect - please refer to the response to the reviewers.

We clearly state now that:

"*From the definition of ID thresholds and IDF curves it is clear that the time intervals over which the intensities of rainfall are examined in the two cases are different, although the term used to define them is the same. In fact, duration refers to the total length* ***(or the time to triggering) D*** *of a user-defined rain event, on the one hand, and to a fixed-length temporal running window W, on the other. Once W is chosen in IDF curves, the population of the corresponding intensities is* ***defined \emph{unconditionally} from the beginning of the event or from the landslide triggering moment****. The duration D in ID thresholds, instead, is defined \emph{conditionally} to* ***the beginning of the precipitation events****.* ***The*** *population of ID pairs depends on user-defined choices concerning the identification of the* ***events****.* ***Associating the ID pair of a landslide-triggering event to the IDF curve for the corresponding duration (and thus assigning this ID pair an exceedance probability***

based on the distribution of extreme events of that duration) disregards the conditions by which ID pairs are defined, that is, starting with the beginning of the event.”

And that:

“***The probability of observing an intensity I over the duration D of ID thresholds is a conditional probability, and is conditioned on the fact that the explored time interval (D) starts with the beginning of the precipitation event, however it is defined. Conversely, the probabilities given by IDF curves for a fixed window W are unconditional.*** *It is therefore erroneous to quantify the* ***probability*** *of a given intensity I in the ID space of the ID thresholds using* ***unconditional*** *probabilities from the IW space of the IDF curves.”*

The sentences above are independent of the triggering processes and of the data source used in the example, instead they are general. Several examples that have nothing to do with landslides can be made. For instance, one would not assess the growth of a child by comparing the height of the child with the heights of the general population (unconditional). What one should do is to compare it with the children of the same age (conditional to the age). This clarifies the *conditional* versus *marginal* (or *unconditional*) issue. On top of that, in our case the age of the child (exact window of triggering) is not exactly known, but we only know that the age is “likely” lower or equal than, say, 10 years (where the “likely” refers to our epistemic uncertainty, and this limit refers to the time between the beginning of the precipitation event and either our best estimate of the triggering instant or the end of the precipitation event). This does not mean that conditioning the comparison to 10 years old children is appropriate, since we only know it is lower or equal than that.

Concerning the request from referee #1: the true, unknown, triggering window does not necessarily depend on the dataset used nor on the knowledge of the exact triggering instant. It is always true when the main assumption behind the use of ID thresholds (that is, the DF/landslides are not influenced by the precipitation that occurred before the event begins - i.e., before we start the clock that quantifies D) is met. This is because this assumption implies that the earliest possible start of triggering temporal window W^\dagger, while unknown, is the start of the storm, and the latest possible end of W^\dagger is the triggering instant, which is before the end of the event.

This is a general concept and does not depend on our data nor on its uncertainties. Running the requested analyses is therefore irrelevant to our message. Instead, it would increase the focus on a part of the example that is irrelevant to our main message, thereby distracting the attention of the readers from it.

To further improve clarity, we changed the notation for the true unknown triggering window (now $W^\dagger$) to avoid confusion with the assumed triggering window of Section 5 (W*). This came out of a suggestion from reviewer #2 that we believed got to the point of what confused the reviewers of the previous round.

We trust these clarifications explain why some of the aspects raised by the referees are not relevant to our message, and we hope the new version better clarifies these aspects to the readers as well.

Referee #1

I would like to thank the Authors for responding to my comments and clarifying most of my doubts. The manuscript certainly benefited from the review phase; some issues are now better clarified, and the main motivation of the work is more understandable.

Thank you for taking this additional time on our manuscript and providing your feedback.

I've still only one concern, regarding the way the Authors identified and compared the durations D and W. The authors state that "IW* pairs are associated with temporal scales W* that are always smaller than the duration D of ID pairs". Now we know that the triggering instants of the debris flows in the used dataset are not known. Thus, it is clear that the ID pairs, as calculated by the authors, could have included hours that were not related to the triggering: if a debris flow occurred in the middle of one day, the ID pair was defined including around 12 hours – and maybe a few mm of rainfall – which should have been discarded. As a consequence, the calculated ID pairs can be longer (and the intensity lower) than they should be. Therefore, the main reason for the fact that IW* were shorter than ID may lay in the dataset used (with only daily temporal information) and in the method used to calculate D. This has also implications in the discussion on the underestimation of the triggering precipitation (figure 2 and related text).
This issue should be better acknowledged by the Authors (e.g., in lines 130-135), and – on the other hand – should suggest that the results of the real-world example cannot be easily generalized.

Thank you for this comment. The reasoning of the referee, however, is only partially correct. The reviewer concludes that "the main reason for the fact that IW* were shorter than ID may lay in the dataset used (with only daily temporal information) and in the method used to calculate D". This is incorrect. The fact that $W^* \leq D$ does not necessarily depend on the dataset used nor on the knowledge of the exact triggering instant. It is always true when the main assumption behind the use of ID thresholds (that is, the DF/landslides are not influenced by the precipitation that occurred before the event begins - i.e., before we start the clock that quantifies *D*) is met. This is because this assumption implies that the earliest possible start of triggering temporal window W*, while unknown, is the start of the storm, and the latest possible end of W* is the triggering instant, which is before the end of the event.

This being said, we agree on the fact that the results of a real-world example cannot be quantitatively generalized, although the qualitative aspect (direction of the bias) is general. We amended the text in this direction: *"**The quantitative results presented for this study case strictly depend on the data we used (debris flow processes in the Alps, daily resolution on the occurrence of the debris flows, temporal resolution of the radar data, etc.), and on the methods and assumptions we took for designing the experiment (namely, that the triggering window corresponds to the window with the highest severity, i.e., $W^\dagger=W^*$). Nevertheless, they are expected to be qualitatively representative of the general case.**"*

My question is: is the underestimation of the triggering precipitation due to a real difference between IW* and ID pairs, or it is due to the coarse temporal resolution of the dataset (which affects only the calculation of ID)? Perhaps the Authors could try to quantify the influence of the uncertainty in the debris flows triggering instants on the calculation of the ID pairs, before comparing them with the IW* pairs. A way to address this issue can be found in some published works. A few years ago, Peres et al. (2018) [https://doi.org/10.5194/nhess-18-633-2018] proposed a quantitative analysis of the impacts of uncertain knowledge of landslide initiation instants on the assessment of rainfall thresholds, using a synthetic dataset. The authors found that uncertainties in the landslide triggering instants may lead to underestimation of the thresholds and consequently more false positives. More recently, Mondini et al. (2025) [https://doi.org/10.1016/j.scitotenv.2025.179453] characterized the temporal uncertainty of some records included in a landslide catalog before using them to build a prediction tool for rainfall-induced landslides.

I acknowledge that this would need some additional work (perhaps too much for a brief communication?), but this issue came out only after the first round of review, during which it was clear that the triggering instants of the debris flows were not known. The best option would be to carry out these analyses to remove any doubt. If the Editor considers this to be too much work for a brief communication, I believe it is at least necessary to discuss the uncertainties related to the data in greater detail

As mentioned in our reply to the comment above, the fact that $W^\dagger≤D$ (where $W^\dagger$ is the true, unknown, triggering window) does not necessarily depend on the dataset used nor on the knowledge of the exact triggering instant. It is always true when the main assumption behind the use of ID thresholds (that is, the DF/landslides are not influenced by the precipitation that occurred before the event begins - i.e., before we start the clock that quantifies *D*) is met. This is because this assumption implies that the earliest possible start of triggering temporal window W^\dagger, while unknown, is the start of the storm, and the latest possible end of W^\dagger is the triggering instant, which is before the end of the event.

This is a general concept and does not depend on our data nor on its uncertainties. Running the requested analyses is therefore irrelevant to our message. Additionally, we are afraid it would distract the readers from the main idea behind this study, that is the confusion between *conditional* and *marginal* (or *unconditional*) *probabilities* applied to the case of ID thresholds and IDF curves.

To better deliver our message, we included several edits to the main text, as follows:

Section 2: ***"It follows that the ID space is the space of the precipitation intensities I observed over a D that begins with the start of the precipitation event, however this is defined."***

Section 3: "***It follows that IDF curves provide the annual exceedance probability of an intensity I over a temporal window of length W.***"

Section 4: ***"****From the definition of ID thresholds and IDF curves it is clear that the time intervals over which the intensities of rainfall are examined in the two cases are different, although the term used to define them is the same. In fact, duration refers to the total length* ***(or the time***

to triggering) D *of a user-defined rain event, on the one hand, and to a fixed-length temporal running window W, on the other. Once W is chosen in IDF curves, the population of the corresponding intensities is* ***defined \emph{unconditionally} from the beginning of the event or from the landslide triggering moment****. The duration D in ID thresholds, instead, is defined \emph{conditionally} to* ***the beginning of the precipitation events****.* ***The*** *population of ID pairs depends on user-defined choices concerning the identification of the* ***events****.* ***Associating the ID pair of a landslide-triggering event to the IDF curve for the corresponding duration (and thus assigning this ID pair an exceedance probability based on the distribution of extreme events of that duration) disregards the conditions by which ID pairs are defined, that is, starting with the beginning of the event.*** *[...]* ***The probability of observing an intensity I over the duration D of ID thresholds is a conditional probability, and is conditioned on the fact that the explored time interval (D) starts with the beginning of the precipitation event, however it is defined. Conversely, the probabilities given by IDF curves for a fixed window W are unconditional.*** *It is therefore erroneous to quantify the* ***probability*** *of a given intensity I in the ID space of the ID thresholds using* ***unconditional*** *probabilities from the IW space of the IDF curves****"***

Section 5: *"****These quantitative results strictly depend on the data we used (debris flow processes in the Alps, daily resolution on the occurrence of the debris flows, temporal resolution of the radar data, etc.), on the methods and assumptions we took for designing the experiment (namely, the triggering window W^* corresponds to the window with the highest severity). Nevertheless, they are expected to be qualitative representative of the general case.****"*

Referee #2

I have reviewed the manuscript entitled "Brief communication: Threshold and probability. The conceptual difference between ID thresholds for landslide initiation and IDF curves" as an additional reviewer for this round. This paper examines the conceptual differences between the ID thresholds and IDF curves in relation to debris-flow and landslide triggering. In Sections 1–4, the authors clearly and concisely describe the essential issues regarding rainfall data processing for ID thresholds and IDF curves. In Section 5, based on a well-constrained debris-flow dataset, the authors convincingly demonstrate how these processing differences can affect our understanding of rainfall conditions that trigger debris flows. The results are robust and highly meaningful, and will likely stimulate further analyses of rainfall-induced landslide and debris-flow initiation processes. Although the presented framework requires independent IDF curves, it is widely applicable across diverse environmental settings and effectively removes the dependence on local conditions and subjective interpretation. Overall, this paper represents a necessary and timely contribution to our community and could serve as a benchmark study. It is a promising and well-written manuscript that I believe will reach a wide audience. I have a few comments and suggestions that may help improve the clarity and overall flow of the paper. I hope my feedback will assist the authors in further refining this excellent work.

Sincerely,
Haruka Tsunetaka

Thank you for spending time and effort on our manuscript. We are glad to read such a positive evaluation. We did our best to address your concerns and suggestions, as detailed below.

In general, we realise that several of the raised points concern section 5, that is a simple example we propose to make the theoretical reasoning in section 4 a bit more tangible. It is important to clarify this to avoid the readers putting too much weight on the technical details of this section instead of our main message. For this reason, we now included a new paragraph at the beginning of section 5: ***"The inconsistency highlighted in the previous section lies on a theoretical level, and holds for all the instances in which conditional and unconditional probabilities are confused. To make this theoretical issue more tangible, we provide here a real-world example. It should not be interpreted as an approach to solve the issue above, which cannot be solved. The results we will present cannot be quantitatively exported to other cases, but its qualitative implications in terms of direction of the biases are expected to hold."***

1. Differences between debris flows and landslides

My first concern relates to the differences between debris flows and landslides in the triggering mechanisms implicitly evaluated by ID thresholds. As described in L23-24, ID thresholds for landslide triggering commonly evaluate whether slope is activated "trigger" and "cause" prepared by rainfall input (Bogaad and Greco, 2018). However, some researchers argue that ID thresholds for debris-flow triggering reflect various processes, such as changes in sediment availability (e.g., Pastorello et al., 2018; Tsunetaka et al., 2021a), whether debris flows reach to the monitoring station (e.g., Bel et al., 2017), and changes in sediment composition (e.g., Guo et al., 2016; Tsunetaka et al., 2021b). These differences, which ID pairs may evaluate different initiation mechanisms between landslides and debris flows, should be well considered through the manuscript.

In my view, these differences relate to only interpretation of real world example (i.e., Section 5). Thus, my recommendation is that, either deleting or moving the related sentence (L23-24) and paragraph (L78-87) to Section 5, or providing a more precise explanation of the above differences. By doing so, the explanations provided in Sections 1–4 would more clearly highlight that they describe generalized issues concerning the ID–IDF relationship, which apply universally, regardless of whether the triggering process regarding debris flows or landslides.

Thank you for this consideration. The reviewer correctly evaluates that the reasoning about trigger and cause may only relate to the example in section 5. To better clarify this aspect, we amended a paragraph in section 4, that now reads: *"These false alarms add to the* ***biases related to uncertain knowledge of the triggering moment and epistemic uncertainty on the triggering processes and to the*** *ones caused by systematic sampling of rainfall away from the location of the triggering landslide and by the use of coarse temporal resolution rainfall data."*

Concerning the sentence in the introduction, we'd like to highlight that lines 23-24 do not mention the "trigger-cause" concept. Actually, they provide a caveat on the fact that the effects of precipitation are mediated by physical processes which then lead to the possibility of triggering. For this reason, we believe that the statement fits where it currently is (introduction).

2. Difference in the definition of W* between Sections 4 and 5
In the previous review round, both reviewers raised several concerns regarding the analysis presented in Section 5. In my view, these concerns mainly stemmed from the difference in the definition of W* between Sections 4 (L72: true, unknown triggering interval) and 5 (L105: time interval during which the most severe intensity was observed). In the revised manuscript, the authors have addressed this issue by clearly describing how W* is defined. However, I am still concerned that this difference may cause readers to misunderstand how the authors distinguish between theoretical facts and their interpretations throughout the paper. Indeed, it appears that the authors themselves may still be somewhat uncertain about this distinction (L139–140). It might be clearer to readers if, in Section 5, all results were consistently described in terms of W corresponding to max(Tw), with the subsequent discussion and interpretation developed under the assumption that such W is approximately equivalent to W*.

Thank you for this comment, we think it goes right to the point. And, even more, thanks for the suggestion, which we applied.

We modified the symbol for the true, unknown triggering window in section 4, which now is symbolized with a dagger instead of an asterisk. In addition, we edited the related text in section 5 as follows: "*Here we will provide an example in which the triggering time interval is assumed to be the time interval during which the most severe intensity was observed****, meaning that we assume $W^\dagger=W^*$.***"

Furthermore, in the discussion of Figure 1b, we added: ***"Here, the scaling of IW* pairs and ID pairs is similar because our definition of W^* follows what is done in IDF curves, but the intensities associated with the true $W^\dagger$ may scale differently from the IDF curves, as they may be conceptually different.***"

We think the clarity is now much improved thanks to this suggestion.

Some readers may also wonder why the authors did not apply the same framework to a broader dataset that includes landslides or other regions. However, I recognize that the authors have used a well-constrained, high-quality debris-flow dataset, and that extending the same analysis to other phenomena or regions would be extremely challenging due to the inherent difficulties in identifying a sufficient number of ID pairs and determining reliable W* values. Therefore, I consider this dataset and the associated analysis to be particularly valuable. That said, these practical and methodological challenges may not be readily apparent to some readers, so providing a brief clarification in the text could further help convey the value and uniqueness of this dataset.
For landslides, it is generally impossible to predict where they will occur in advance or to identify the exact time of initiation. The recurrence interval of landslides in a given region typically spans several decades to centuries, making it difficult to obtain a sufficient number of ID pairs at the regional scale. Consequently, most ID thresholds for landslides have been derived at the national scale. Preparing a landslide dataset suitable for an analysis such as that presented in Section 5 is therefore extremely difficult.

For debris flows, regional ID thresholds are often derived from ID pairs based on observed occurrences within the same or nearby catchments. However, in many cases, “occurrence” refers to the arrival of debris flow at the observation point rather than the initiation of motion. As mentioned in Comment 1 and the references therein, this implies that the threshold inherently includes the processes of debris-flow development and runout, which depend not only on rainfall but also on sediment availability, distribution, and composition. Hence, the strict identification of W^* is practically difficult.

Thank you for these considerations. The reviewer agrees with our choices in terms of dataset and methods. The only request is to better explain the importance of the dataset following the concerns of the previous reviewers. Because it appears both reviewers are now satisfied, we believe no action is required in this instance.

The paragraph in Section 4 (L78-87) already mentions, at least in part, that W^* is practically indeterminable. As mentioned earlier, I suggest moving that paragraph to Section 5 and expanding on it there. My recommendation is to strengthen the explanation of the practical difficulties in determining W^* and in obtaining numerous ID pairs from real-world data. The authors could clarify that the analysis in Section 5 essentially deals with another metric (such as W corresponding to max(Tw)) but that, in this study area and dataset, assuming W = W^* is reasonably valid, citing adequate references to support this rationale. This approach would clearly separate Sections 1–4 as describing theoretical principles and Section 5 as demonstrating an empirical application and interpretation based on real data.

Thank you for the suggestion. As mentioned above, we modified the symbol for the true, unknown triggering window in section 4, which now is symbolized with a dagger instead of an asterisk. In addition, we edited the related text in section 5 as follows: “*Here we will provide an example in which the triggering time interval is assumed to be the time interval during which the most severe intensity was observed****, meaning that we assume*** ***$W^\dagger=W^*$.***”

Furthermore, in the discussion of Figure 1b, we added: ***“Here, the scaling of IW* pairs and ID pairs is similar because our definition of W^* follows what is done in IDF curves, but the intensities associated with the true $W^\dagger$ may scale differently from the IDF curves, as they may be conceptually different.***”

I also agree with the discussion in L123-134. However, as noted in Comment 1, it is important to emphasize that the key finding, that W^* ranges from 30 minutes to 6 hours, was derived specifically for debris flows. Whether landslides exhibit a similar pattern remains unknown. If the debris flows analyzed here are sourced primarily from channel-bed sediment, this relatively broad time interval may reflect temporal variations in sediment availability within the catchment (e.g., Tsunetaka et al., 2021a).

It is true that these results pertain to debris flows but, as mentioned above, the qualitative results hold irrespective of the processes. This simple real-world example is intended to make the theory more tangible. As mentioned in our responses to the previous review round, we don’t think we should dive into discussions about the triggering processes as this is way out of our scope. Actually, we probably should not, because we don’t have sufficient elements for that type of reasoning.

3. Scaling limitation of IDF curves
The authors, for convenience, have estimated the return periods of very short-duration rainfall (less than 15 minutes) based on existing IDF curves. I am concerned that the validated lower limit of the existing IDF scaling (Borga et al., 2005) may be around 15-minute rainfall durations. The estimation of return periods for such short-duration rainfall involves high uncertainty. In fact, in Figure 1b, there appear to be at least two unrealistic data points plotted at less than 1 hour on the x-axis and around 200 mm h^{-1} on the y-axis. Although I understand that there is currently no practical alternative approach, a brief mention of this limitation in the main text would further improve the clarity of the manuscript. I also believe that this concern is independently addressed by the results presented in Figure 3, which show the decorrelation time of rainfall. I was quite impressed by how closely this figure conceptually aligns with Figure 2. If space permits, adding a more detailed explanation and discussion of Figure 3 would make the manuscript even more refined and insightful.

This is a great point to address. The “unrealistic” values you mention concern an event that occurred over the night between July 16 and 17, 2007 in Passeier Valley. The reason we are familiar with this event is that the radar showed extremely large values, exceeding 100 mm over less than one hour, with no rain gauge data provided by the regional authorities to support such numbers. Certainly an “unrealistic” value, especially coming from the radar. However, radar maps suggested that the St. Leonard in Passeier station was directly hit by the event - a rare occurrence (Lengfeld et al., 2020). Further investigations revealed that the St. Leonard in Passeier data was removed during quality control due to “unrealistic” high values. Once collected, the “unrealistic” values from the rain gauge fully confirmed the “unrealistic” values from the radar. These numbers are real.

The work by Lengfeld et al. (2020) shows that the German rain gauge network, one of the densest in the world, captures less than 20% of the hourly extreme events. Together with the systematic sampling bias related to the analysis of landslide and debris flow triggering discussed in Marra et al. (2016) this explains why some very real values may appear as “unrealistic”.

Lengfeld et al., 2020, ERL, https://doi.org/10.1088/1748-9326/ab98b4

Marra et al., 2016, JoH, http://doi.org/10.1016/j.jhydrol.2015.10.010

4. Comparison of slopes regarding ID, IW*, and IDF
The discussion comparing the slopes of ID, IW*, and IDF relationships may need to be moderated, as the current analysis does not provide sufficient evidence to draw a definitive conclusion. Because the D values of the ID pairs are relatively large, the data points in this dataset appear only within the range of approximately 1 to 48 hours on the x-axis in Figure 1b. The scaling for durations shorter than 1 hour is essentially extrapolated.

Thank you for this suggestion, we amended the text to make it softer, as follows: *"These latter threshold* ***seem to better*** *align with the regional scaling of extreme rainfall (dashed-dotted lines in the background),* ***suggesting that*** *the apparent difference in the power-law scaling of ID thresholds and IDF curves discussed by \citet{Bogaard2018} can be attributed to methodological issues in the choice of rain duration, often made regardless of the physical processes responsible for debris flow or landslide occurrence."*

Considering this, when focusing on the range between 1 and 48 hours in Figure 1b, the differences in slope among the ID threshold, IW* threshold, and IDF scaling appear to be nearly equivalent. Therefore, the overall difference in slope might simply reflect the data limitation that there are few ID pairs with small D values. For landslides, triggering rainfall events with D < 1 hour are extremely rare. However, for debris flows, such short-duration triggering events have been reported in various catchments (e.g., Abancó et al., 2016; Bel et al., 2017; Tsunetaka et al., 2021a).

Thank you for raising this consideration, on which we fully agree. We added a sentence to this section to better highlight this aspect: "***Here, the scaling of IW* pairs and ID pairs is similar because our definition of W^* follows what is done in IDF curves, but the intensities associated with the true $W^\dagger$ may scale differently from the IDF curves, as they may be conceptually different."***

Line by line comments
Title: Since the case study focuses specifically on debris flows, it might be helpful to include the term "debris flow" in the title to clearly indicate the study target.

We respectfully disagree with this suggestion. The study uses debris flows as an example but it could have used synthetic data instead because it ultimately focuses on a conceptual problem that is not necessarily related to the triggering processes.

L27-29: Readers who are less familiar with rainfall thresholds may wonder why the parameter E is sometimes used. A brief explanation of its meaning and rationale, supported by an appropriate reference, would help clarify this point.

Thank you, we specified: "*the total precipitation depth (for which the symbol E is typically adopted)*"

L50: IDF are -> IDF curves are

Corrected, thank you

L61: an user defined -> a user defined

Corrected, thank you

L78-79: Please consider softening the tone of the explanation slightly to make it more balanced and accessible to a broader readership.

Thank you for the suggestion, we rephrased the sentence to: *"It is therefore erroneous to quantify the **probability** of a **triggering event of** intensity I in the ID space of the ID thresholds using **unconditional** probabilities from the IW space of the IDF curves".*

L89-90: It would be helpful to clarify whether these debris flows were initiated by landslides or if they mainly resulted from bulking and entrainment of unconsolidated channel-bed material. A brief explanation would improve the reader's understanding.

Thanks for the suggestion. We included this information in the text, as follows: *"**We discuss** the potential effects of the inconsistency highlighted above using data from 12 storms that triggered 133 debris flows in the eastern Italian Alps during 2005-2014. They constitute $\sim$40\% of all the debris flows recorded in the area in this period \citep{Nikolopoulos2014} **and resulted from bulking and entrainment of unconsolidated channel-bed material.**"*

L98: It is not entirely clear whether these represent the triggering locations or the observation locations. Please clarify how the triggering location was defined in this dataset.

The mentioned sentence clearly states that these are the locations of the debris flow triggering: "For each debris flow, we extracted the precipitation time series observed by the radar over the triggering locations." To avoid possible confusion on this, we included additional information about the spatial accuracy of the database we used, as follows: "***The triggering location of the debris flows in the database is provided with an uncertainty which is much smaller than the pixel size of the radar data (approximately one order of magnitude smaller, Marra et al., 2014)**."*

Figure 1: To further aid interpretation, you might consider adding summary scatter plots for all 133 events alongside Figure 1a: specifically, $W*$ vs. D and $I*$ vs. I. Including such plots could make the relationships easier to grasp, particularly for readers who are less familiar with rainfall threshold analyses.

Thank you for this suggestion. We prefer to keep the figure as is to avoid overemphasizing the study case and distract the readers from the main message.

References
Abancó, C., Hürlimann, M., Moya, J., & Berenguer, M. (2016). Critical rainfall conditions for the initiation of torrential flows. Results from the Rebaixader catchment (Central Pyrenees). Journal of hydrology, 541, 218-229.
Bel, C., Liébault, F., Navratil, O., Eckert, N., Bellot, H., Fontaine, F., & Laigle, D. (2017).

Rainfall control of debris-flow triggering in the Réal Torrent, Southern French Prealps. Geomorphology, 291, 17-32.
Guo, X., Cui, P., Li, Y., Fan, J., Yan, Y., & Ge, Y. (2016). Temporal differentiation of rainfall thresholds for debris flows in Wenchuan earthquake-affected areas. Environmental Earth Sciences, 75(2), 109.
Pastorello, R., Hürlimann, M., & D'Agostino, V. (2018). Correlation between the rainfall, sediment recharge, and triggering of torrential flows in the Rebaixader catchment (Pyrenees, Spain). Landslides, 15(10), 1921-1934.
Tsunetaka, H., Hotta, N., Imaizumi, F., Hayakawa, Y. S., & Masui, T. (2021a). Variation in rainfall patterns triggering debris flow in the initiation zone of the Ichino-sawa torrent, Ohya landslide, Japan. Geomorphology, 375, 107529.
Tsunetaka, H., Shinohara, Y., Hotta, N., Gomez, C., & Sakai, Y. (2021b). Multi-decadal changes in the relationships between rainfall characteristics and debris-flow occurrences in response to gully evolution after the 1990–1995 Mount Unzen eruptions. Earth Surface Processes and Landforms, 46(11), 2141-2162.