We would like to thank the editor David Peres and the two anonymous reviewers for the time taken to handle, read and comment our manuscript. We provide below here our response to the reviewer comments in blue fonts, together with the actions we did on the manuscript to address the comments in *blue italics fonts*.

Response to Reviewer #1

General comment: "The goal of this research is on the characterization of rainfall that lead to the occurrence of debris flows in arid regions. More specifically on a type of debris flows, here called "short-lived debris flows).

Although the paper is well written in general, I have some difficulties at really understanding some parts of the manuscript. The topic is certainly of interest for NHESS and the research gap on studying debris flows in arid environment is well identified. Nevertheless, this research shows scientific and technical weaknesses that I identified a bit everywhere in the manuscript. They are summarized here in bullet points, followed by more specific comments pointing to issues directly in the text (and also the commented PDF)"

We thank the reviewer for the thorough review and for suggesting that the subject of the manuscript is indeed of interest to NHESS and the research gap is well identified.

As the reviewer mentions, it seems some ideas behind our study have been misunderstood. While we found many comments from the reviewers very useful for the improvement of our presentation, we realize that some seemingly important comments are actually related to misunderstandings – we will try here to clarify our perspective and will add the relevant text that is included in the revised manuscript to address these comments and clarify the writing.

We believe our detailed replies will solve these misunderstandings. The edits we propose to the manuscript thanks to the reviewer comments greatly improve its clarity.

Response to major comments:

Comment #1: The definition and identification of debris flows is questionable and I would suspect that in several cases the slope failure processes that are studied are not debris flows. In addition, the justification of defining "short-lived debris flow", i.e. quite a singular terminology, is not backed up by a sound support of the literature.

Thank you for this very important comment. We are aware that the morphological features here examined are somehow different with respect to the typical debris flows observed, for instance, in Alpine regions or other areas with long slopes and rough orography. Indeed, we long debated about the best term to define these processes but we ended up realizing that the only slope failure processes that may form the morphologies we found in the field are debris flows (DF). While also debris avalanches can be considered, as also mentioned by the reviewer in comment #13, all the failures are found along existing channels (established depositional landforms - fans - are well seen) suggesting that it was not the first event at each channel. Following Hungr et al. (2001), we then opted for the term debris flows (see also the new suggested figures 2 and 3 below).

Our attempt to specify their peculiar size with the term short-lived DF was clearly not successful, as both reviewers pointed out. To avoid further confusion, we will change the term throughout the manuscript, including the title, to the more standard debris flows (DFs). We included additional clarifications on this aspect in the text, as follows: *"The DF deposits are located along small ephemeral streams that drain the cliff area above them and present evidence of previous DF activity. This suggests that they were mobilized by flow events and therefore considered as DFs, as suggested by Hungr et al. (2001). The DFs we examine are small-sized, with a maximum runout distance (between the source location and deposit) of a few tens of meters."*

Hungr et al 2001, https://doi.org/10.2113/gseegeosci.7.3.221

Comment #2: The authors use an impressive amount of multi-temporal Lidar derived data to map the debris flows from differences in topography. Although this represents surely one robust way to map mass-movements, there is not really a justification about the use of such a sophisticated approach, especially with respect to the fact that there is no proper analysis/discussion associated with the morphometry of the debris flow processes.

We respectfully disagree with this opinion. A large portion of the work done in this study was the identification of the slope failures, and this is a critical aspect for arid areas where almost no data is available. The studied region is a desert area with almost no human presence with the exception of one main road and few villages and no systematic survey of the area is performed by local authorities. Our approach consists of the minimum we needed to detect these failures and assign them a triggering location in space. Without this long work, no data would be available at all. We would be glad to hear about alternative ways to obtain this information.

This study is part of an ongoing project started at the end of 2019, whose initial focus concerns the detection of past events, both in terms of DF occurrences and corresponding initiating rainfall conditions. The main limiting factor for DF initiation in the study area is rainfall (see also responses below), so we concentrated our efforts on the rain characteristics as presented in this study. Currently and in our future research we are using the LiDAR data together with other methods for developing morphometric analysis aiming to better understand the triggering mechanism, i.e., the failure of the colluvial debris material, and the geomorphic characteristics of the events.

Comment #3: The rainfall analysis of radar-derived data is quite complex to finally say that a few storms have been identified and validated with the use of media/social network information. In other words, the dates of the events could have been found in a easier and more direct way. In addition, overall, the strategy for looking for a potential rainfall candidate should be better defined, especially with respect to the fact the rain gauges are sometimes located on areas quite close to the places of DF occurrence. If a rain gauge station is a few km away from a event, why not using its data instead of information from a radar situated ~70 km that provides data at a 500 m spatial resolution?

We respectfully disagree with this comment. Indeed, we used a complex rainfall analysis in order to point out the triggering storms of past events, and this required long hours of work. Again, we have to recall that we are dealing with areas in which no systematic survey is available, and for which witnesses are only seldom available (in our case 1 only Facebook video). The reviewer suggests that "the dates of the events could have been found in an easier and more direct way": we are keen to hear possible suggestions.

Then on the use of radar instead of rain gauges. Rain gauges are a poor source when it comes to the triggering rainfall conditions of slope failures, especially when rainfall is of convective nature. There is ample evidence in literature showing that rain gauges located only few km away from the triggering location (1-2 km can be enough) systematically underestimate rainfall amounts with respect to the triggering locations (Nikolopoulos et al., 2014; 2015; Marra et al., 2014). This underestimation is, on average, as high as 30% at 5 km distances (~10% at 2 km; Marra et al. 2016), and is related to the sharp spatial scales of convective precipitation – which in the studied area are even sharper than the ones in the mentioned paper (e.g., see Marra and Morin, 2018). Based on this evidence, information at 500 m resolution from a radar situated ~70 km away from the triggering locations is to be considered superior with respect to rain gauges located km away (see also Marra et al., 2017). Some of this literature is already mentioned in the manuscript but not all of them: as many of these papers are co-authored by one of the authors, we tried to limit self-citations to the essential.

Nikolopoulos et al., 2014, http://doi.org/10.1016/j.geomorph.2014.06.015

Nikolopoulos et al., 2015, http://doi.org/10.1016/j.geomorph.2015.04.028

Marra et al., 2014, http://doi.org/10.1016/j.jhydrol.2014.09.039

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

Marra and Morin, 2018. http://doi.org/10.1016/j.atmosres.2017.09.020

Marra et al., 2017: https://doi.org/10.5194/hess-21-4525-2017

Comment #4: Furthermore, although the authors acknowledge it in several places in the manuscript, little is said about the problems/challenges of using radar data in the identification of rainfall conditions, temporally and spatially. The authors claim that the products have been validated against rain gauge information, but they remain rather vague on the topic. When one knows that the whole analysis is based on four DF events, this has deep implication on the overall robustness of the work and the discussion that comes out from it. In several places, results to contradict each other (see for example comment in the PDF, line 311).

This is a very important point, thanks for bringing it up. Provided that radar data is generally superior to rain gauges for what concerns the triggering rainfall conditions in convective environments (see our response to the above comment), quantitative radar estimates remains far from perfect. We discuss details about this issue in the presentation of radar data (lines 180-187 of the submitted manuscript), and we will now include some additional discussion of these aspects in a new section, together with a caveat for what concerns the sample size of our dataset. We believe the response to the previous comments and the details provided below here will help clearing the doubts and contradictions raised by the reviewer.

The new section of the revised manuscript: *5.3 Limitations of this study: "Our results are based on a relatively small sample of small-sized DFs detected in an arid region of the eastern Mediterranean. Although we extensively explored the region in Figure 1, this inventory cannot be considered complete because DFs could have been missed due to noise in the LiDAR data or other data issues (e.g. see Section 4.3). In addition, weather radar uncertainties may affect the precipitation estimates (e.g., see Marra et al., 2022). One source of uncertainty in particular is worth mentioning: the possible advection of precipitation during its falling from the height of the radar sampling volume (~3 km above the ground level) and the ground could lead to misplacements of the radar estimates of few hundreds of meters; this would typically lead to an underestimation of the rainfall amounts (Marra et al., 2016). In addition, it is important to recall that the adjustment of the radar data is based on few available stations: some level of uncertainty in the quantitative estimates is thus to be expected. While our qualitative results are robust with respect to these sources of error and support our reasoning in terms of process description and understanding, the numbers are subject to residual uncertainty and caution is advised against their direct use in warning systems."*

Comment #5: Concerning the discussion, there is a lack of analysis on the DF processes with respect to the literature. Since arid environments are clearly understudied, one could expect that rainfall conditions are here analysed with respect to what is known from the literature. That would not only allow to better highlight the originality of the work, but also to better identify some problems in the method (errors and uncertainties of the radar product). Without this connection to the international literature, the "arid regions" perspective highlighted in the title is not present, and, consequently, we remain at a very case-study level.

We thank the reviewer for his important comment.

A proposed we have added a new section in the Discussion chapter of the revised manuscript: "*As mentioned in section 4.3, the investigated triggering cells are intense (10-85 mm h-1) and short (20-45 min) (Table 1 and Fig. 6). Both their intensities and durations are lower than the previously suggested threshold for DF triggering in the study area (intensities >30 mm h-1 for duration of one hour or longer, Ben David-Novak et al., 2004). A previous study conducted in the arid slopes of the Grand Canyon, Arizona, cautiously suggested that sustained intensity exceeding 20 mm h-1 and a total rainfall of 25 to 50 mm may be a minimum requirement for DF triggering, without mentioning a minimum duration (Melis et al., 1995). In our study, only two out of the four triggering cells, have maximum intensities exceeding 30 mm h-1, but these intensities were observed only for a short period (<10 min), much shorter than what previously reported (Fig. 6). While for most of our mapped DFs, intensity was indeed >20 mm h-1 the total rainfall for all triggering cells was <25 mm (Fig. 4). Quantitative accuracy of radar data, however, is not perfect and, as mentioned above, possible underestimation cannot be excluded. Considering the limited datasets available for arid regions, it is still impossible to determine a unique threshold for DF* *triggering. More attempts should focus on data collection in these regions and on carefully considering the spatiotemporal distribution of rainfall during the initiating storms. Hints towards the importance of the temporal rain distribution during a storm could already be found in Ben David-Novak et al. (2004) as their data showed that for both studied events the triggering cells reached the area only hours after a significant antecedent precipitation."*

Comment #6: Concerning the literature, one would expect that it is also used to back up the definition of antecedent rainfall conditions. There is also some unclear statements about antecedent and triggering conditions. One could further question the fact that the triggering rainfall, that are also measured here, are not discussed.

Thank you for this comment, which highlights an important weakness of our submission: it was missing a proper definition of triggering and antecedent rainfall. In the study, we make use of three main concepts: (a) the "initiating storm", which is the storm event during which the DF is initiated, (b) the "triggering cell", which is the high-intensity convective cell that occurred during the "initiating storm" and that provided the final trigger, (c) the "antecedent precipitation", which is the rainfall occurred within the "initiating storm" before the "triggering cell" hit.

At this concern, please also note that the submitted manuscript contained a mistake: the initiating storms are indeed separated using 5 dry days (24*5 hours), and not 24 hours.

We included a paragraph at the beginning of section 4 in which these concepts are properly defined, and we then use consistent terminology throughout the revised manuscript: *"We introduce here three concepts that we will use to characterise the properties of the precipitation that led to DF initiation. We define as "initiating storm" the storm event during which a DF is initiated. To this end, storms are defined as wet periods separated by at least 5 days of dry weather (i.e., 120 hours with less than 0.1 mm h-1 in the radar data). We define as "triggering cell" the high-intensity convective cell that likely provided the final trigger to the DF initiation. Last, we define as "antecedent precipitation" the rainfall observed before the triggering cell and during the initiating storm (i.e., between the beginning of the initiating storm and the beginning of the triggering cell)."*

Comment #7: Some parts of the text are quite (too much?) descriptive on the "geological-geomorphological" context in which the DF occur. In addition, such information is not used in the analysis and discussion. For example, when it concerns shallow landslides, one key aspect that explains their occurrence is the availability of colluvium. Without it, rainfall will not have an impact (e.g. Dykes et al., 2002; Parker et al., 2016). Such an aspect on colluvium availability could really explain why in some places DF are not observed despite the presence of potentially "good" rainfall conditions (keeping of course in mind the reliability of the latter). This is something I would really like to be discussed.

We thank the reviewer for this important comment, also raised by reviewer #2. The issue of colluvium and sediment availability is crucial as it is one of the principal three factors needed for DF triggering (together with slope and water). Our field observations, however, show that sediment availability is hardly a limiting factor in most of the slopes of the study region. In fact, we deem that slope steepness is more of a limiting factor than sediment availability.

To make this aspect clearer, we have included additional information on the description of the study area: *"The ~40 km long studied escarpment can be generally divided into three parts: lower soft sediments, middle colluvium and upper cliffs (Fig. 2). In many places along the lower parts of the escarpment (altitude over 300 m below the mean sea level), the carbonate rock slopes are covered by lacustrine and fluvial sediments (Fig. 2c), deposited during high lake stands in the Pliocene-Pleistocene periods (Bartov et al., 2002, 2007; Sneh, 1979; Begin et al., 1980). Many of these exposures are soft and/or cohesionless, mostly composed of carbonate pebbles. The upper third of the escarpment, usually the steepest, is rocky and composed of hard carbonate rocks (mostly dolomites). The middle part of the escarpment is mostly covered by colluvium composed of fragments originated from the rock mass above (Fig. 2b). While the colluvium coverage may change from place to place along the escarpment, it is still abundant and does not represent a limiting factor for DF triggering in the study area. The colluvium thickness is changing laterally with an observed maximum value of a few meters. At the surface, the colluvial material is usually grain supported with increasing amounts of fine particles at depth of a few tens of centimetres. The colluvium fine fraction (<2 mm) is dominated by crashed dolomites and some quartz, calcite and phyllosilicates probably from eolian source. This fine material becomes muddy and unstable once exposed to water. The lacustrine and fluvial sediments together with the colluvial deposits constitute the typical source material of DFs in the study area (Ben David-Novak et al., 2004)."*

Furthermore, we rephrased part of the discussion in section 5.2 to: *"Given the vast sediment availability, the rarity of DFs in the area is usually explained by the dry weather…"*

<u>Response to specific comments:</u>

Comment #8: From the title, we expect several things such as "short-lived" debris flows and antecedent rainfall be a focus of the introduction. However, antecedent is mentioned only once, while "short-lived" is not.

Comment #9: From the title, we expect the focus on the arid regions, however, most of the issues related to the study of debris flows in arid regions that are mentioned in the introduction are illustrated with the description of the study area. I would have expected an introduction that better highlights the challenges/novelties/needs to study debris flows in these arid regions.

Thank you for these comments, we will reply together to these two as they share many elements. As mentioned above, we replaced the term short-lived debris flows with debris flows and added a paragraph to define antecedent precipitation in Section 4. In the introduction of the submitted manuscript we described previous studies conducted in arid and semi-arid regions (lines 29-51). We now further develop the subsequent paragraph to emphasize the gap and needs in studying debris flows in arid regions and the specific case of antecedent precipitation:

"… They suggest that minimum conditions for DF triggering consist of rainfall intensity exceeding 30 mm h-1 for duration of at least one hour, although the typical lifetime of convective cells in the area is shorter (around 20-40 min according to Belachsen et al. (2017)). Moreover, the

characteristics of the rainfall occurred before the triggering cells were not considered in previous studies (e.g., Ben David-Novak et al., 2004).

The possible importance of antecedent precipitation on landslide triggering was extensively studied for non-arid environments (e.g. Glade et al. 2000; Aleotti 2004; Guzzetti et al. 2008; Frattini et al. 2009; Kim et al. 2021). In tropical areas (Brand, 1992) and slopes covered by grains having large inter-particle void space (Corominas and Moya, 1999), it was suggested that antecedent rainfall is less important than in other environments, possibly because of the high permeability of the local soils that reduces the potential of failure (Rahardjo et al., 2001). Conversely, to the best of our knowledge, the effect of antecedent rainfall on DF triggering in arid regions was not yet explored, probably due to the lack of DF observations and of adequate rainfall data."

Comment #10: Lines 53-56 explain the objective of the research. Two key methodological aspects are highlighted: high-resolution topographic models and high-resolution radar rainfall estimates. It is somehow surprising that none of these "technical aspects" are mentioned in the introduction.

It is true that these methodological aspects are not mentioned in the introduction, but it is also true that our study do not aim at advancing them. We personally think including introduction text on LiDAR and radar methods would be out of the scope of our introduction, but we are open to this possibility in case this is deemed important.

Comment #11: Section 3.1 on mapping methods: I have some difficulties at understanding why there is a focus on the use of multi-temporal Lidar-derived DSM to map the debris flows. Lidar data acquired here 4 times (every other year over the period 2013-2019) represent a great source of data for the characterization of the processes. However, such data are not useful to get the timing information of the debris flows initiation that would allow the rainfall characteristics be analysed. The only way to get the exact timing of a debris flow, assuming that at least an one-day accuracy is needed for such a rainfall analysis, is certainly through direct field observation/media/social network (as explained in lines 110-111). Therefore, also the use of orthophotos is not really appropriate here.

Note however, that the processes that are being studied always need a minimum of "geomorphologic" characterisation (size, shape, mobility) and as such, combining very-high resolution orthophotos and Lidar-derived topographic data, is a great plus to achieve this.

Thank you for pointing out this important aspect. We understand the confusion regarding the use of multiple DSMs and orthophotos for this study. We need to clarify this point before describing the changes in the manuscript. This study started at the end of 2019, and aims at better understanding the major mass wasting processes along the western Dead Sea escarpment. The used LiDAR scans however were not part of this specific project, and were produced within a different project dealing with sinkholes evolution. Nevertheless, the scans ended up being helpful in identifying mass wasting processes and providing constraints on their temporal occurrence. We use the LiDAR to detect the events and narrowing the time window for the search of the storm that initiated the DF.

To make these aspects clearer, we rewrote the relevant parts of the manuscript, as follows.

The beginning of section 3.1 Mapping methods: *"Aiming to detect modern, natural changes resulting from mass-wasting processes along the studied escarpment, we compared aerial photos and high-resolution digital surface models (DSM) that were available for the years 2013-2019. During that period, no additional instrumentation was installed in the study area to identify triggered DFs. We therefore used the available orthophotos and DSMs to map new DF deposits and to minimize the time interval of triggering."*

The beginning of section 3.2 Identified debris flows and field observations: *"We identified 43 DF deposits occurred between the years 2013-2019. Two additional deposits were classified as rockfalls and removed from the analyses. The DF deposits are located along small ephemeral streams that drain the cliff area above them and present evidence of previous DF activity. This suggests that they were mobilized by flow events and therefore considered as DFs, as suggested by Hungr et al. (2001). The DFs we examine are small-sized, with a maximum runout distance (between the source location and deposit) of a few tens of meters. The areal extent of each DF deposit ranges between 10 and 1000 m2 with an average value of 150 m2 and elongated shapes of ~30 m in length and ~5 m in width. Although during the studied period (2013-2019), only small-size and short runout distance DFs could be clearly mapped, longer runout distance DFs from past events are observed along the studied escarpment (Fig. 3)."*

The beginning of section 4.2 Identification of the most likely initiating storms: *"Since the study area was documented by airborne LiDAR roughly every second year, we search for all the potential initiating storms that could have triggered the DFs during the time intervals between two subsequent scans (Fig. 5)."*

Comment #12: Section 3.2:

- *Here reference is made to landslide mapping; which sound different from that of debris flows. Rock fall identification is also mentioned. This is confusing to refer to slope processes that are not the focus of the study.*

Thanks for this comment. We rephrased this section to avoid confusion of the reader: *"We identified 43 DF deposits occurred between the years 2013-2019. Two additional deposits were classified as rockfalls and removed from the analyses. The DF deposits are located along small ephemeral streams that drain the cliff area above them and present evidence of previous DF activity. This suggests that they were mobilized by flow events and therefore considered as DFs, as suggested by Hungr et al. (2001)."*

Hungr et al 2001, https://doi.org/10.2113/gseegeosci.7.3.221

- *This is only here that a definition of short-lived debris flows is provided. This is a terminology that is barely used in the literature and one would welcome more insight on the reason why the authors pay a focus on this process differentiation from "normal" debris flows (see also my comment on the introduction).*

Thanks for pointing out these aspects. As mentioned above (see response to comment #1), we now use the standard term debris flows (DFs) to avoid confusions.

Comment #13: Figure 2 is the only visual information that allows to see what a short-lived debris is. And here I must admit that I question the processes that are analysed. To me some of the features look more like debris avalanches. This is the reason why more illustrations (as stated earlier) could be needed. My doubts about the characterization of the processes are further confirmed with the description provided in lines 133-142.

We refer here to the response to comment #1. In addition, following this and other comments we added to the revised manuscript two additional figures (Figures 2 and 3, see below) that will give the reader another visual description of the studied DFs.

Comment #14: Line 122. Larger debris flows (than the short-lived ones) are not included in the analysis while their occurrence is said to be possible. I am ok with that. However, the authors say that the scars left by these DF could have been blurred by subsequent road construction and floods. Hence my question. If large features can disappear from the landscape, what about the short-lived DF? How reliable is the inventory?

Thanks for this important comment. Indeed, large features can disappear from the landscape, and the same is even more true for smaller DFs such as the ones we examine here. Nevertheless, it should be pointed out that the mapped debris flows are usually located in the middle of the slopes far from any anthropogenic intervention. Despite this, our inventory cannot be considered as complete (just check, for instance the additional DFs discussed in section 4.3 of the revised manuscript). Indeed, our inventory is not considered complete in our analyses: we describe observations concerning observed events which are not influenced in any way by the fact that other events can be not in the inventory.

We added a paragraph on this aspect in the discussion: *"Our results are based on a relatively small sample of small-sized DFs detected in an arid region of the eastern Mediterranean. Although we extensively explored the region in Figure 1, this inventory cannot be considered complete because DFs could have been missed due to noise in the LiDAR data or other data issues (e.g. see Section 4.3)."*

Comment #15: Section 4.1. Rain gauge data are used. However this is not mentioned in the objectives where only the focus on radar-derived rainfall is made.

Rain gauge data are used to adjust weather radar data and evaluate the completeness of the record. We included a reference to this aspect in the objectives: *"In this paper we aim at improving our understanding of the critical conditions for DF triggering in arid areas by combining high-resolution topography models, field surveys, and an advanced archive of high-resolution gauge-adjusted radar rainfall estimates which comprises both triggering and non-triggering events."*, although weather radar estimates are almost always gauge-adjusted so that the sentence could be redundant.

Comment #16: Figure 5. shows that the highest intensities are not necessarily over the group 3 of SLDFs. However, SLDF do not seem to have occurred in the other areas. Figure 6 confirms this.

This is exactly our point! We discuss this aspect in lines 258-262 of the first submission, and we suggest possible explanations. Overall, we suggest that the antecedent rainfall is the limiting factor and high intensity rain is not enough to trigger the mapped debris flows.

Comment #17: Line 255-256: for the first time in the manuscript, a definition of antecedent rainfall is provided. One would expect something definition according to what is usually adopted in the literature on DF so that better justification/comparison/discussion is carried out. Line 294, antecedent rainfall is defined as a continuous rain period that ends up at the moment when the potential highest rainfall intensity is seen. Here also, one would need reference to the literature (see for example: . Bogaard, T.A., Greco, R., 2016.)

Thanks for pointing this out. As mentioned above, we are now using a rigorous definition of the concept of antecedent precipitation (see response to comment #6) and we rephrase the revised introduction to better present the knowledge gap concerning antecedent precipitation.

<u>Response to specific comments in the pdf:</u>

Line 7: this term is never used in the main text

We have updated to "*heavy*".

Line 12: that occurred

We have updated the text.

Line 29: why this highlight on temperate regions? There are other climatic regions than temperate and arid where DF are common

We have removed this sentence.

Line 36: if the region is limited in human settlements, one could assume that the DF impacts are potentially limited. I know want the authors want to stress here. However, I think that the sentence could be better formulated.

We have removed “a few” from the sentence.

Line 43: not clear

We rephrased it to “the future occurrence”

Line 46: from what is explained in the former sentences, drier (arid) conditions lead to more DF. From that, I would argue that the most limiting factor is more on the supply of material rather than on the precipitation.

About the sediment availability, please refer to our response to the comments above. The former sentences, however, do not suggest what the reviewer mentions (i.e. that drier conditions lead to more DF) so we don’t know how to further address this comment.

Line 75: having a sub-section for highlighting one single paragraph is not necessary I think.

We have merged the two subsections.

Line 84: capital letters?

We have rephrased it to *“systems of tropical origin, termed Tropical Plumes, …”* to make it clearer.

Figure 1: this map shows the general location of the fours study sites. I think that a zoom on these sites would also be welcome in order to better understand the type of processes that are studied as well as their close environments.

Following the reviewer’s comments we have added two figures (2 and 3 in the revised manuscript) showing examples of the studied escarpment and DFs.

Line 208: how (what criteria?) do you discard 6 potential candidates for rain storm?

Thanks for this comment. The criterion was explained in the sentence but in an unclear way. We rephrased it to make it clearer: “*Nevertheless, only seven of these eleven storms satisfied the* *conditions over all the mapped DFs of a group of interest (marked by red in Table 1 and blue rectangles in Fig. 5). Therefore, only seven storms can be considered as potential initiating storms.*"

Line 210: from figure 3, I assume that this common storm is for groups 1 and 4? If so, that contradicts the definition of storm in the region that usually have a cell size < 8km (see lines 150-154). This needs clarifications.

The reviewer is confused here. The mentioned lines explain exactly that storms may contain several convective cells and that convective cells have a scale of <8 km: *"...a specific storm, that may last for a few days, could represent the trigger of several groups. However, since the distances between groups observed in the same 2-year time intervals is always greater than 8 km, and the typical scale of convective cells in the region is smaller (Belachsen et al., 2017; Marra and Morin, 2018), these groups were likely triggered by different convective cells. As we will see, this is possibly the case of our groups one and four, which occurred in the northern and southern parts of the study area, respectively (Fig. 1)."*

Table 1: what is the meaning of this? (REFERS TO "FID")

FID is the object ID of a shapefile in ArcGIS Desktop. In our case it is the identification numbers of the DFs. We added this explanation to the caption of Table 1.

Table 1: explain the color differentiation in the caption.

Thanks, we included these details in the revised caption.

Line 225: in line 207, you say 8 potential storms. Here you say 9. This needs clarification.

Thank you for spotting this typo. We have corrected it.

Line 234: so, one social media report allows to outperform this rather complex radar analysis?

Yes, we were lucky enough to have one case with direct witnesses reporting on social media. One out of 43.

Line 236: here is a example of detailed information whose the usefulness could be questioned.

We have removed this sentence.

Figure 4: these Fid numbers are not explained. What do they represent?

These are the identification numbers of the DFs. In the revised caption we defined it.

Figure 4: i do not understand how to read this. Two meanings for one axes?

Yes, indeed the axes values represent there two meanings, both, the rain intensity and the total cumulative rain during the initiating storm. If this is deemed confusing we can include a secondary axis on the right part of the panels: it would have the same numerical values and different units (as it is now).

Line 276: this paragraph is something that one would expect in the introduction.

This paragraph has moved to the Introduction.

Line 280: this is not something that is specifically demonstrated here. In this research you show that antecedent rainfall could play a role. But, you do not explore in detail the behavior of these rains that add something new to what is already known from the general literature.

As there is no such thing as "general literature" on DF triggering in arid areas, but only a couple of papers, we put our discussion in perspective to those papers. Moreover, we think the discussion section is exactly the point in which we should discuss hypotheses that are not necessarily proven in the study and propose possible new ideas.

Figure 7: 7b is not indicated on the figure

Thanks for pointing this out. We have added the labels to the subpanels.

Line 311: if I understand it correctly, the 2014 rain, that is higher in max intensity and in antecedent conditions triggers less and small DF that a "smaller rain" in 2015. Also, if larger DF are associated with the 2015 rain, why don't we also observe smaller features? what does this imply? also with respect to the reliability of the rainfall analysis.

Clearly, uncertainty in the rainfall analysis could be a reason for this. But there is a point the reviewer is missing. Although we would like to have clear-cut thresholds to separate triggering and non-triggering rainfall, reality is far more complicated than our models, and a number of aspects may influence the triggering, aspects that may not have been addressed in this study - or even in no study at all since to our knowledge there is no rainfall threshold able to separate with 100% accuracy triggering and non-triggering events. In this study we created a relatively large dataset of DF in an arid region and we show that intensity alone (basically the only threshold used so far in literature for these areas) is not enough. We also show that antecedent precipitation could represent an important element for DF triggering in these regions. In one case a previous storm showed both larger intensities and antecedent precipitation but did not trigger. To our view, if this 'contradiction' means anything about the rainfall analysis, it means that it is robust enough not to be confused by hard thresholds on precipitation alone.

Line 315: I find it a bit strange to have such a focus on early warning system here while it is never mentioned earlier in the manuscript. I encourage the authors to better focus their discussion on the actual outcomes of their research.
Thank you for the suggestion, we updated the section title to: "*Implications for debris flow occurrence in the region*"

Line 341: but you mostly discuss the antecendent rainfall, not the triggering ones.

Thank you for this comment. We have changed it to: "*investigate the rainfall conditions leading to the initiation of DFs*"

Response to Reviewer #2

General comment: This manuscript aims at analysing the rainfall conditions leading to debris flows in an arid area in the Dead Sea region. To do so, Lidar-based DSM and media posts are used to gather a debris flow inventory. Rain-gage and radar data are employed to characterise the rainfall conditions leading to debris flow initiation. Generally, I appreciate the work. The manuscript is interesting and fits well with the journal topics. I feel, however, that the manuscript needs some major improvement for its publication.

We thank the reviewer for the thorough review.

Response to major comments:

Comment #1: The definition of antecedent rainfall and triggering rainfall should be clarified in the manuscript. In the current version of the manuscript, two different definitions of antecedent rainfall are given. The first definition is provided in Lines 255-256, where antecedent rainfall is defined as "the time passed between the beginning of the storm and the time defined as the offset of the triggering convective cell". The second is given in lines 293-294, where antecedent rainfall is defined as "the rain accumulated on the deposit pixel during the period starting on a 24 h break in rain until the specific measured intensity". This is rather confusing. Usually, the term antecedent rainfall refers to the rainfall that falls in an area during a given period of time before the beginning of the triggering rainfall episode. From the two definitions provided in the manuscript, it is not clear which is the distinction between triggering rainfall and antecedent rainfall.

Thank you for this comment, which highlights an important weakness of our submission: the proper definition of triggering and antecedent rainfall. In the study, we make use of three main concepts: (a) the "initiating storm", which is the storm event during which the DF is initiated, (b) the "triggering cell", which is the high-intensity convective cell that occurred during the "initiating storm" and that provided the final trigger, (c) the "antecedent precipitation", which is the rainfall occurred within the "initiating storm" before the "triggering cell" hit.

Please also note that the manuscript contained a mistake: the initiating storms are indeed separated using 5 dry days (24*5 hours), and not 24 hours. Hence the doubts about the double definition.

We included a paragraph at the beginning of section 4 in which these concepts are properly defined, and we then use consistent terminology throughout the revised manuscript: *"We introduce here three concepts that we will use to characterise the properties of the precipitation that led to DF initiation. We define as "initiating storm" the storm event during which a DF is initiated. To this end, storms are defined as wet periods separated by at least 5 days of dry weather (i.e., 120 hours with less than 0.1 mm h-1 in the radar data). We define as "triggering cell" the high-intensity convective cell that likely provided the final trigger to the DF initiation. Last, we define as "antecedent precipitation" the rainfall observed before the triggering cell and during the initiating storm (i.e., between the beginning of the initiating storm and the beginning of the triggering cell)."*

Comment #2: I suggest modifying the structure of the manuscript. As it is now, the results of the analysis of the triggering rainfall conditions are only presented very briefly in section 4.3 (only *one paragraph is used). The results of the analysis of the antecedent rainfall are scattered throughout the manuscript. Part of them are given in section 4.3, where the triggering rain should be characterised. Some others are presented as part of the discussion (section 5.1). This is surprising given the title. I think that including a results section explaining both the analysis of the triggering rainfall and the antecedent rainfall conditions before the discussion would help make your point clearer.*

We reorganized part of the manuscript according both reviewers' suggestions. Specifically, we have moved a section from the discussion (5.1) to the introduction (1), and moved the paragraphs of section 5.1 about antecedent precipitation (which indeed include some results) in the results section (4.3).

Comment #3: In the discussion, the fact that not all intense rainfall events can trigger debris flows is solely attributed to the role of antecedent precipitation. However, other factors such as sediment availability also play an essential role in debris flow initiation. This fact has not been considered. It should be at least mentioned in the discussion and the abstract.

We entirely agree with this reasoning and with the fact that in arid regions sediment availability is often critical. Our field observations, however, suggest that colluvial sediments on the steep slopes in the study area are vastly available and do not represent an important limiting factor. The small size of the steep slope areas is to be considered more of a limiting factor than sediment availability. Just as an example it is important to note that Figure 9 is obtained only focusing on the locations where DF occurred. This implies that the conditions on slope and sediment availability were certainly met, at least for all the storms occurred before the triggering.

To better explain this aspect, we edited this part of the discussion to: *"Given the vast sediment availability, the rarity of DFs in the area is usually explained by the dry weather (i.e., low number of storms) and the small areal extent of the susceptible steep slopes: only convective cells hitting the small susceptible area can trigger a DF."*

Comment #4: [Section 2.1: Geography and geological settings] The authors provide an extensive description of the geological setting of the studied area in which the debris flows occur. However, more relevant information to understand the materials involved in the debris flows is to be provided. Which is the granulometry of the materials involved in the debris flows? Is the availability of sediment homogeneous over the study area? These two factors are relevant as the soil properties might influence the velocity at which pore pressures are dissipated. Sediment availability plays an important role in debris flow initiation.

Thank you for this comment. As suggested above, we provided new text to explain the importance of the sediment availability. Nevertheless, we also recall here that sediment availability is not regarded as a critical limiting factor in the study area, as also confirmed by our field surveys.

The new text of the revised manuscript: *"The middle part of the escarpment is mostly covered by colluvium composed of fragments originated from the rock mass above (Fig. 2b). While the colluvium coverage may change from place to place along the escarpment, it is still abundant and does not represent a limiting factor for DF triggering in the study area." (from Section 2).*

Following this comment, we also added some quantitative information about the grain-size analysis of the source material and a better description of the DF deposits:

In section 2: *"The colluvium thickness is changing laterally with an observed maximum value of a few meters. At the surface, the colluvial material is usually grain supported with increasing amounts of fine particles at depth of a few tens of centimetres. The colluvium fine fraction (<2 mm) is dominated by crashed dolomites and some quartz, calcite and phyllosilicates probably from eolian source. This fine material becomes muddy and unstable once exposed to water."*

In section 3.2: *"For most of the mapped DFs (N=37), the deposits consist of angular dolomite and limestone fragments, unsorted grains with a maximum size of a few tens of centimetres. This deposit composition reflects the colluvium composition above it which mostly composed of the upper escarpment cliffs and some aeolian fine particles. Hence, the source for the debris material is the talus at the base of theses cliffs (Fig. 2b). Differently, in fewer cases (N=6) located at the bottom of the lower step on the escarpment, the deposits consisted of rounded fragments, apparently derived from the nearby cliff of lacustrine and conglomerate sediments (Fig. 2c). These same two sources of debris materials were also reported by a previous study in the region (Ben David-Novak et al., 2004)."*

In addition, we added a figure with a representative example of the slope over the study area, highlighting the contributing slope areas (Fig. 2).

<u>Response to specific comments:</u>

Comment #5: [Section 3: Debris flow detection and characterisation] The occurrence of past debris flows is mapped mainly using Lidar-derived DSM. However, due to the relatively low frequency in which lidar surveys have been conducted, this product does not allow determining the triggering times. Media reports are useful but tend to include only events close to urban areas that impact the population. Perhaps the use of satellite images can improve the debris flow inventory in terms of the number of events and determination of the triggering time. Has this possibility been considered?

Yes, this is accurate. We tried to use satellite images for this purpose. Unfortunately, we couldn't find high-resolution (<10 m/pixel) satellite images of the study area for the relevant period. We did find that the Sentinel-2 mission has the relevant area, however, the resolution was too poor to observe the relatively small-size DFs. Since we started this project in 2019, we are now following each event more closely and have connections with the nature reserves in the area that report us about relevant events if occurred. In addition, soon we will install several time-lapse cameras and it seems that also the resolution of the satellite images getting better.

Comment #6: [Section 4: Rainfall data] From the text in section 4.1 it seems that rain-gage measurements are only used to adjust the radar rainfall estimates. However, line 199, states that rain gages are used to fill in the gaps in the rainfall time series at the locations of debris flows. In such cases, the rainfall accumulations at the rain-gage sites could be very different from the rainfall falling at the catchments where debris flows are triggered. Have you adjusted the rainfall measurements in some way?

It seems the reviewer misread the sentence as the radar data was not filled with rain gauge data. The sentence actually says that: *"In order to make sure that we did not miss a critical event, we compared the radar data with rain gauge stations."*. As mentioned in the submitted manuscript, we found that only two events in the whole period 2013-2019 were missing.

Comment #7: [Line 11]: The term short-lived debris flow is not widely used. It might be a good idea to provide the definition earlier in the text. Currently, the term is not defined until line 119.

Thanks for this comment. Our attempt to specify the peculiar size of our DFs with the term short-lived DF was clearly not successful, as both reviewers seem to dislike it. To avoid further confusion, we have changed the term throughout the manuscript, including the title, to debris flows (DFs).

We have included additional clarifications behind our choice in the text, as follows: *"The DF deposits are located along small ephemeral streams that drain the cliff area above them and present evidence of previous DF activity. This suggests that they were mobilized by flow events and therefore considered as DFs, as suggested by Hungr et al. (2001). The DFs we examine are small-sized, with a maximum runout distance (between the source location and deposit) of a few tens of meters."*

Comment #8: [Lines 201-206] How do you define antecedent rainfall? What criteria are used here to distinguish between antecedent rainfall and triggering rainfall?

Thank you for this important comment. We kindly refer to the reply to comment #1.

Comment #9: Fig 5, line 265: typo: red triangles.

Thanks, corrected.

Comment #10: Fig 6 The caption should be improved. What do the black dots represent?

Thanks for pointing this out. The black dots represent the locations of the rain gauges.

The caption is improved to: *"The cliff-top (dashed line) together with the Dead Sea shoreline mark the narrow band of steep escarpment where debris flows may potentially be triggered. Black dots show the location of the rain gauges. (a) Antecedent precipitation – map of the total rainfall during the March 2014 event until the onset of the trigging cell. (b) The total precipitation observed during the triggering cell (10:30-11:05). (c) Peak intensity of the triggering cell (5 min time interval). (d) Peak intensity of the triggering cell (30 min time interval)."*

Comment #11: [Section 4.3: Characterisation of the triggering rainfall] Reading the sub-section title, I expect to find an extensive description of the triggering rainfall events. However, this topic is addressed only in the first paragraph. Lines 253-261 focus on the antecedent rainfall.

We have adapted the section title to better represent the content: *"Characterization of the rainfall leading to debris flows"*

Comment #12: Line 316-327 I wonder which is the sediment availability in the different catchments where debris flows have been reported. Assuming that you had a complete inventory, it could be the case that some rainstorms did not trigger any debris flow simply because there was not enough prepared sediment in the catchments (see: Bovis, M. J., & Jakob, M. (1999). The role of debris supply conditions in predicting debris flow activity. Earth Surface Processes and Landforms, 24, 1039–1054).

Thanks for this comment. The issue of colluvium and sediment availability is crucial as it is one of the three factors needed for DF triggering (together with slope and water). At this concern, we need to point out two aspects. Our field observations, however, show that sediment availability is hardly a limiting factor in most of the slopes of the study region. In fact, we deem that slope steepness is more of a limiting factor than sediment availability.

To make this aspect clearer, we have included additional information on the description of the study area: *"The ~40 km long studied escarpment can be generally divided into three parts: lower soft sediments, middle colluvium and upper cliffs (Fig. 2). In many places along the lower parts of the escarpment (altitude over 300 m below the mean sea level), the carbonate rock slopes are covered by lacustrine and fluvial sediments (Fig. 2c), deposited during high lake stands in the Pliocene-Pleistocene periods (Bartov et al., 2002, 2007; Sneh, 1979; Begin et al., 1980). Many of these exposures are soft and/or cohesionless, mostly composed of carbonate pebbles. The upper third of the escarpment, usually the steepest, is rocky and composed of hard carbonate rocks (mostly dolomites). The middle part of the escarpment is mostly covered by colluvium composed of fragments originated from the rock mass above (Fig. 2b). While the colluvium coverage may change from place to place along the escarpment, it is still abundant and does not represent a limiting factor for DF triggering in the study area. The colluvium thickness is changing laterally with an observed maximum value of a few meters. At the surface, the colluvial material is usually grain supported with increasing amounts of fine particles at depth of a few tens of centimetres. The colluvium fine fraction (<2 mm) is dominated by crashed dolomites and some quartz, calcite and phyllosilicates probably from eolian source. This fine material becomes muddy and unstable once exposed to water. The lacustrine and fluvial sediments together with the colluvial deposits constitute the typical source material of DFs in the study area (Ben David-Novak et al., 2004)."*

In addition, we added to the revised manuscript two additional figures 2 and 3 (see below) that will give the reader another visual description of the studied DFs and sediment availability on the studied slopes.

We have also rephrased part of the discussion in section 5.2 to: *"Given the vast sediment availability, the rarity of DFs in the area is usually explained by the dry weather…"*

It should also be recalled that, despite our efforts, our inventory cannot in any way be considered as complete. For instance, section 5.1 of the submitted manuscript, which will now be moved earlier in the Results section, provides as example of DFs that were only identified at a later stage in the study based on the results of the study.

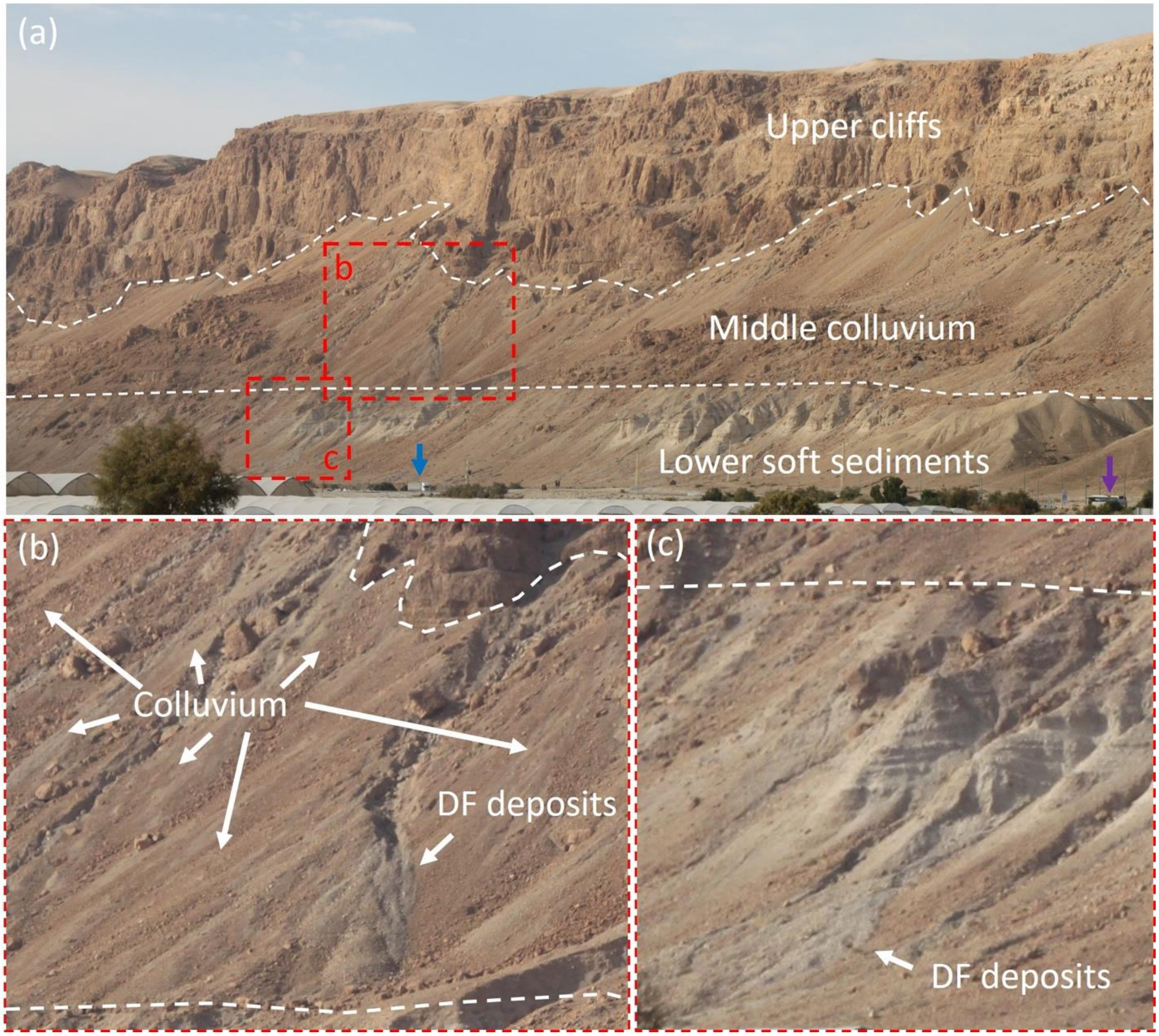

Figure 2. The studied escarpment and the observed DF deposits. (a) The northern part of the study area where group 1 DFs were observed. The escarpment is generally divided into three parts: the upper cliffs, middle colluvium, and lower soft sediments. A truck and a bus are marked for scale by blue and purple arrows, respectively. The extent of the zooming-in photos are marked by red rectangles. (b) A section of the middle colluvium part of the escarpment. Colluvium cover (some are pointed by white arrows) with some large boulders is cut by small ephemeral streams that ends with DF deposits. The source material is usually clearly seen in the colluvium above these deposits. (c) A DF lobe at the end of a short ephemeral stream. The deposits source from the light-color lacustrine sediments.

Figure 3. An example of DF deposits triggered prior to 2013. A distal lobe (white arrow), two levees (blue arrows) and a wide and shallow channel in-between them at the end of ephemeral stream drained the cliff area.