The manuscript esurf-2021-49 titled "Sediment export in marly badland catchments modulated by frost-cracking intensity, Draix-Bléone Critical Zone Observatory, SE France" presents a really interesting study based on a long-term database collected in a humid badland area located in the SE France. The manuscript presents interesting information and results, and I consider that it could be published after moderate changes. In my opinion, there are some major/moderate issues that should be checked before a new revision in Earth Surface Dynamics.

We thank the reviewer for their constructive comments, which we have addressed as explained below. In the following, the reviewer comments are indicated in black, our response in grey, and the proposed modifications to the manuscript in red in a grey italic paragraph.

1) In general, the manuscript presents new ideas and new data, analysis and results, but I will encourage the authors to highlight the novelty of your manuscript at the end of the Introduction section. In this paragraph, you should also improve the presentation of the general and specific objectives.
2) I think that it is really relevant the long-term dataset that you present in this manuscript. As you briefly mentioned in the manuscript, there are few sites well monitored and with such long-term dataset. In that sense, I think that in the introduction section, it should be remarked, and also in the discussion section (line 270). You can also include other areas/references where this kind of information is recorded (Tabernas in SE Spain, Vallcebre in NE Spain…).

We agree on the necessity to improve the presentation of the general and specific objectives of the study. In that sense, we suggest to modify the introduction as follows:

*Line 84: In this study, the general objective is to develop a similar approach at the catchment scale (0.1-1 km$^2$) in marly badlands. This objective is made attainable by the exceptional long-term dataset available in Draix-Bleone Observatory. At the catchment scale, …*

Moreover, to highlight the challenge of obtaining such a long-term dataset, we propose the following modification :

*Line 272: The Draix-Bléone CZO provides one of the few localities worldwide where such records exist. Similar datasets have been collected in the Araguas and Vallcebre basins of (Northern Spain but cover significantly shorter periods (6-18 months; Regüés et al, 1995; Regüés and Nadal-Romero, 2013).*

3) In the study area, and throughout the manuscript (also in Figure 1) you present two experimental catchments (Laval and Moulin), but in most of the analysis (for example Figures 3 and 4) you only present data and information about the Laval catchment. This issue should be corrected or clarified in the manuscript.

Both the Laval and Moulin catchments are part of the Draix-Bléone observatory and are slightly differently instrumented. The soil temperature dataset that we used is recorded in the Moulin catchment and the climatological station is placed at the outlet of the Laval catchment, while sediment-export data is collected for both catchments. For this reason, it was necessary to present these two experimental catchments in the introduction (Fig. 1). Additionally, because of the availability of two separate datasets, it was possible to analyse the seasonal sediment dynamics in response to climatic events at different scales (Fig. 5). This approach allows us to conclude that these two catchments have the same general dynamics. We chose to work only on the Laval sediment-export data for the correlation with temperature indicators for different reasons:

- Because of its larger size (0.86 km²), the Laval catchment is more representative than the Moulin catchment
- The vegetation cover is more limited on the Laval catchment(32%) compared to the Moulin catchment (46%) and, thus, a larger surface of the Laval catchment corresponds to the environment in which the temperature probes operate (bare marl bedrock ).
- The sediment dynamics are in any case similar for both catchments
- The climatic and soil temperature data is representative of both catchments, since they are very close to each other and share the same lithology
- The uncertainty is smaller for the sediment export form the Laval than for the Moulin 1) because the absolute values are higher and 2) because there are fewer floods where the suspended load data needs to be reconstructed

To clarify our approach, we suggest to add the following paragraph at the end of Section 3.1:

*The soil temperature data has been recorded in the Moulin catchment but is regarded representative also of the adjacent Laval catchment that shares the same lithology. Sediment-yield data is available for both the Moulin and Laval stations (located 100 meters apart at the outlets of both catchments; Fig. 1C) and our analysis of hysteresis cycles shows that both catchments have the same sediment dynamics (see Section 4.1, Figs. 4, 5). Because of its reduced vegetation cover and larger area (0.86 $km^{2)}$, we consider that the Laval catchment is more representative for the analysis of the relationship between sediment yield and temperature indicators (Figs. 3, 4 and 6 to 9) than the much smaller Moulin catchment (0.089 $km^2$). Moreover, the larger sediment export values from the Laval catchment are associated with smaller relative uncertainties. For these reasons, we focus our attention on the sediment-export data from the Laval catchment for our analysis.*

4) Line 115 and also check along the manuscript. You present the results about sediment yield in different units (tonnes, also erosion rates in mm/yr), I think that these results should be presented in specific values (tonnes per Km2 or ha), as the catchment sizes vary. Please, homogenize this information.

To take this comment into account, we suggest to modify the text (lines 115-118) as follows:

*The Draix catchments record some of the highest specific sediment yields observed worldwide: average annual sediment yields for the Laval and Moulin catchments are around 12,000 and 570 tonnes, equating to specific sediment yields of around 14,000 and 5,700 tonnes/$km^2$/y , respectively. Considering only the unvegetated parts of the catchments as contributing to the sediment yield and the measured sediment density of 1700 kg/$m^3$, the average erosion rate is around 8 mm/yr for both catchments (Mathys, 2006).*

5) It is also not clear how do you record the data in the field (around line 129). You explain that you record data during events but, do you have a continuous temporal record? It is true that the maximum sediment export values are recorded during flood events, but it would be necessary to understand how have you measured the data during all the study period. In that sense, it is always an interesting information to indicate which % of the sediment have been exported during the flood events, or even

which % of the sediment have been exported during the maximum events… in that sense the high variability presented in Figure 3 could be better understood.

6) Information provided in lines 128-130 is already a result. In that sense, I suggest two options: (i) to remove from this section and move to the results section; (ii) or to add some reference where this information is already presented.

We do not have continuous temporal data on sediment export. The sediment trap stores bedload continuously but topographic surveys are performed after each major flood event. Suspended-sediment concentration is measured by turbidimeters only during flood events. However, sediment export both as suspended load and as bedload is considered negligible during normal flow in the study area. To support this hypothesis, we computed the annual sediment export from the discharge and the concentration dataset between 2014 and 2018. Making the hypothesis of an inter-event concentration between 0.1 g/L and 1 g/L (which is already a very high value for low-flow), we find that only 0.12% to 1.2% of the total annual sediment export is missing by neglecting the inter-event concentration as we did in the manuscript .

We therefore consider that the total sediment export is equivalent to the total sediment export of the flood events reasonable. We added this information in the manuscript as follows:

*Line 126: Bedload volumes are measured after each flood by topographic surveys of a sediment trap located upstream of the station. Bedload volume is then converted into mass using a density of 1700 kg/m$^3$, constrained by measurements in the sediment trap (Mathys, 2006). The raw data we use is therefore a series of event-scale sediment yield. An analysis of inter-event sediment export shows that flood export represents more than 99% of the total annual sediment export. Thus, sediment export both as suspended load and as bedload is considered negligible during low flow and we define the total sediment export as the monthly or yearly sum of the suspended load and bedload contributions during floods. For a few flood events, the suspended-load data is missing. In such cases, we reconstructed the event-scale suspended sediment yield based on the average proportions of suspended load and bedload, computed from multiple complete years of total load records.*

7) Temperature depths and temporal range. Lines 134-140. I think you should better justify the selection of the different depths why (1, 6, 12 and 24). Do you have any reference to add to justify this selection? Or previous knowledge about it in the study area? I'm also surprise about the time period selection, as you limited the information from mid-October to the end of March. I understand your decision and your justification, but it is true that in humid badland mountain areas there can be low temperatures (below 0ºC) also in April or beginning of October.

Temperature probes were implanted in 2005 and the depths were selected to span the depth of the loose regolith (Maquaire (2002)).

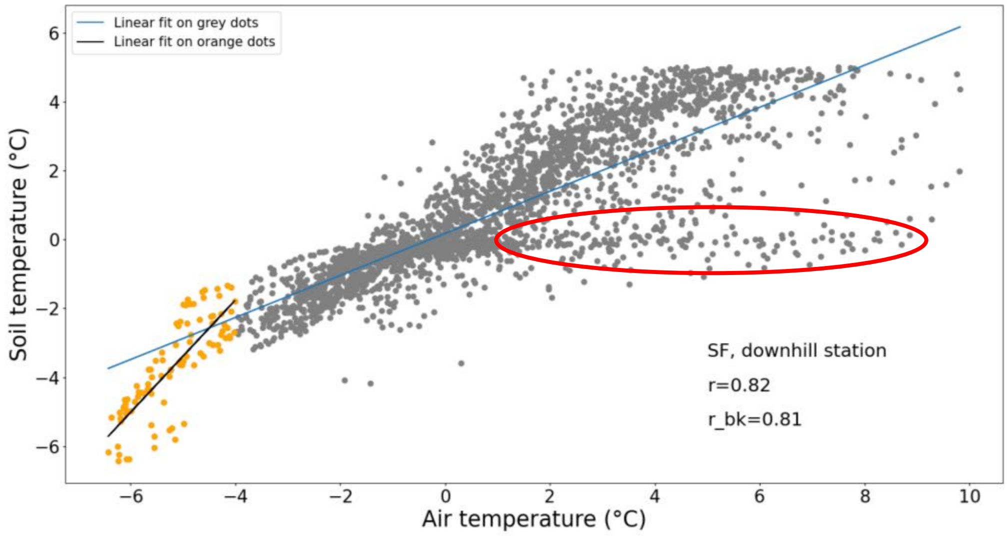
In order to justify our choice of time-span, we have now computed two indicators for the spring-summer-autumn period (April 1$^{st}$ to October 17$^{th}$): the time spent below 0°C, and the time spent in the frost cracking window ( -3° - -8°C).We show that these are both negligible with respect to the values for these parameters during winter. To take these comments into account we suggest to modify the text as follows:

*Line 138: At each location, four probes are available to measure soil temperature at depths of 1, 6, 12 and 24 cm, respectively, which span the ranges of depths that have been reported for the weathered regolith in this area (Maquaire, 2002).*

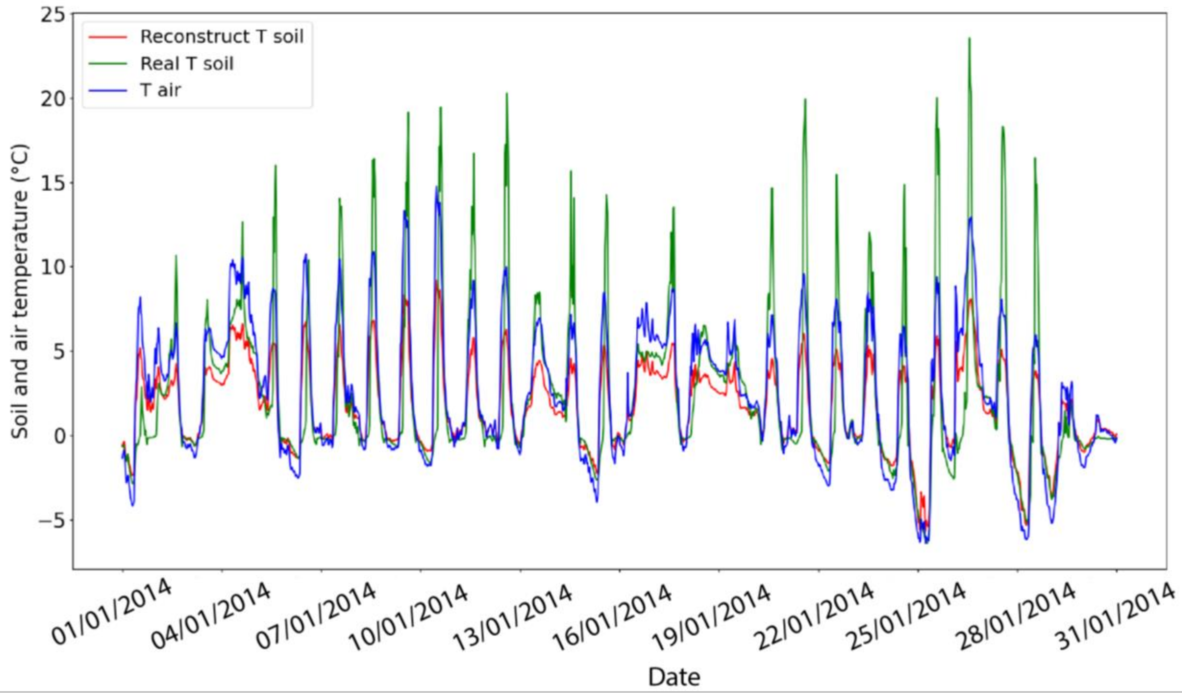
*Line 139: As our interest is on frost weathering, we specifically analysed soil temperatures during the winter season, from October 18th to March 31st. This period was chosen because negative soil temperatures are almost absent outside these dates. During the periods between April 1st and October 17th, we found that the time spent below 0 °C was on average less than 0.4% of the total time in a year and represents less than 4% of the time spent below 0 °C during the winter. The time spent in the frost-cracking window outside of the analysed winter period was null for most of the study years. We chose to start the winter season on October 18th because some yearly series miss temperature data for early October.*

8) Lines 144-152. More information about the reconstruction of the temperature data should be provided (even as supplementary material). Maybe, more info about this process should be included adding a plot with the calibration and reconstruction

To add more information about the reconstruction of the temperature data, we added two figures to the supplementary material and modified the text as follows in order to quantify the number of days that were reconstructed over the analysed period (winter period of each year):




*Supplementary Figure 2: Correlation between soil and air temperature measurements on south-facing slope (SF) downhill station between 01/01/2014 and 31/03/2014. Since we are interested in frost weathering, temperatures were cut off at 5 °C for soil temperature and 10 °C for air temperature. We note a two-tier linear relationship between soil and air temperatures, with a steeper correlation for very low air temperatures (r_bk = 0.81). We therefore set a threshold for air temperature (-4°C) and regress the data above and below that threshold separately (grey / orange dots and blue (r= 0.82) / black lines, respectively). The threshold was determined independently for every year where temperatures needed to be reconstructed. The red ellipse show dots aligned along 0 °C soil temperature and record snow cover . Such snow-cover periods were avoided as much as possible when determining the linear regression parameters.*




*Supplementary Figure 3: Comparison of measured soil temperatures (green line) and reconstructed temperatures (red line) for a complete temperature series (January 2014). Note that the model is only calibrated for soil temperatures <5 °C (Supplementary Figure 1). Air temperature is also shown in blue.*

Modification in the text:

*Line 147: When the missing period is shorter (seasons 2011/2012, 2014/2015, 2015/2016), we searched for a relation between soil temperature and air temperature (recorded at the closest weather station) in order to reconstruct soil temperature during the missing time interval (Suppl. Figures 2 and 3). Over the eleven winter seasons used to calculate our temperature indicators, we reconstructed 56 days, which represents around 3% of the total winter temperature dataset (see Supplementary Table 1 for details).*

9) I think that there is also some confusion about the period you used in the analysis. Some times you indicate 2005-2019 (line 160) but then you indicate 2003-2019 (210). Maybe, as suggestion, it would be nice to include a table (maybe in the Supplementary material) with information about the variables that you have used, where (south-facing or north-facing, high-low slope), and the period that have been used for the analysis. In that sense, it would be really clear.

We used two different time periods for our analyses:

- 2003 to 2019 for the monthly sediment export analysis (4.1, 5.2: Figs. 3 - 5)
- 2005 to 2019 for the temperature indicator analysis (4.2, 5.3 : Figs. 6 - 9)

Because these two results are not directly correlated we used the longer time range available for the sediment-export data to analyse the hysteresis cycle. However, when comparing annual sediment export with the temperature indicators (Fig. 6) we only used the data in the timespan 2005-2019 in order to be consistent with the time range of the temperature datasets.

To reduce the confusion concerning these different time ranges, we added some information at the beginning of Section 3.2.1 part (line155) and mentioned our work on monthly sediment export:

*First, we compared monthly sediment export and monthly rainfall to understand the seasonal dynamics of sediment transport in the Laval and Moulin catchments (Figs. 3 - 5). The analysis was performed for the period 2003 - 2019, during which sediment export was precisely recorded.*

10) In that sense, it is also not clear which information have been obtained and analyzed in north- and south-facing slopes. It is really important, as the literature already have shown that there are significant differences between both expositions. So, if you only calculated some indices for south-facing slopes, some part of the history is missing and it is a pity. Please, think about it because in lines 250-255, you indicate that better correlations are obtained in south-facing slopes, but I think it is not totally true, as one of the variables have not been measured/calculated in the north-facing slope.

We are not sure how we can make this more clear, as we already indicate very clearly in lines 175, 247 and 250-255 what temperature indicators were measured on the north and south-facing slopes, respectively. We acknowledge that it is unfortunate that the deeper temperature probes malfunctioned on the norht-facing slope, disallowing us to calculate FCI on this slope

To take in account your concerns, we slightly modified the text (line 252) and added a recapitulative table on the temperature dataset in the supplementary material.

**Line 250 -255**: *On the north-facing slope, in contrast, the only significant correlation between temperature indicators occurs between the number of freeze-thaw cycles / year and mean negative temperature (R= -0.62). On south-facing slope, the correlation between indicators and sediment-export anomalies show the strongest correlation (R= 0.87) for the frost-cracking intensity indicator but time below 0 °C indicator is also significantly correlated (R= 0.71). On the north-facing slope, the only significant correlation occurs between sediment-export anomaly and time below 0 °C (R= 0.74), but note that frost-cracking intensity was not calculated on the north-facing slope (Supplementary Table 1).*

*Supplementary Table 1: Information about temperature dataset available between 2005 and 2019*

| Depth of probes | Full winter period used in the analysis | Aspects | Hill location | Reconstructed periods |
|---|---|---|---|---|
| -1 | 2006/2007,<br>2007/2008,<br>2008/2009,<br>2009/2010,<br>2011/2012,<br>2012/2013,<br>2013/2014, | South-facing<br>North-facing | Uphill<br>Downhill | **2011/2012 SF uphill** : 24/12 6h30 to 31/12 23h50<br>**2011/2012 all**: 12/03 2h20 to 26/03 12h40<br>**2014-2015 SF**: 7/12 11h10 to 11/12 18h30 |

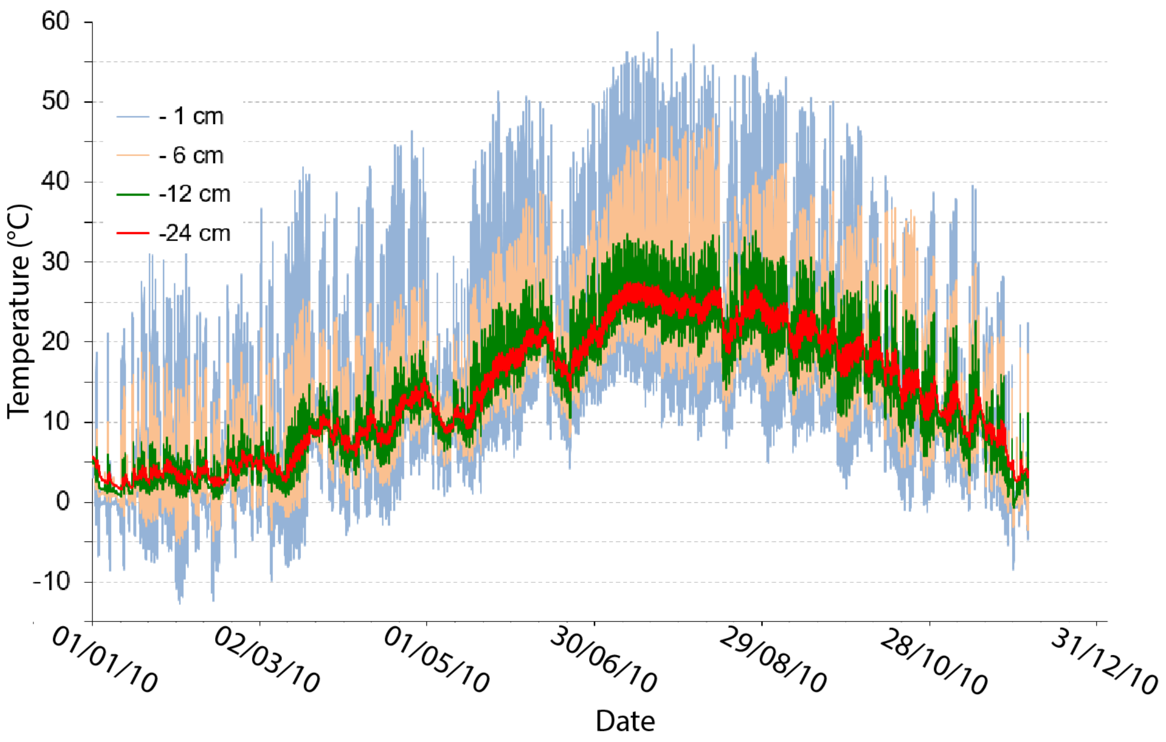
| | | | | |
|---|---|---|---|---|
| | 2015/2016, 2016/2017, 2019/2020 | | | **2015/2016 all**:4/11 18h00 to 7/11 8h40, 21/11 00h00 to 14/12 15h20, 27/12 4h00 to 29/12 18h20, 01/019h10 to 5/01 23h50 |
| -24 | 2006/2007, 2007/2008, 2008/2009, 2009/2010, 2013/2014, 2014/2015, 2016/2017, 2019/2020 | South-facing only | Uphill (except for 2016/2017)<br><br>Downhill | No reconstruction possible |

**11)** I suggest to include the complete temperature database (or at least the period you consider) in the manuscript and not only 2-3 months (Figure 2). I think it is really important to show the temperature pattern in this area, at different depths, once you have already measured this variable. And also the information about the different temperature variables that you have measured (maybe boxplot or similar plots). I'm also really surprise about the high temperatures you recorded in the winter months (January temperature higher than 30º). In Figure 2, you should indicate if these data are average daily temperatures.

We chose to show only a short period for the temperature database in order to more clearly identify the different patterns of each probe. We thought that the annual cyclic variations of temperatures were less relevant for this study. However, in response to this comment, we have added a graph showing a complete annual temperature series to the Supplementary material, allowing the reader to visualise the lower-frequency trend of temperatures during an entire year.

Moreover, we added to the legend of Figure 2 that data are recorded every 10 min. The high peak temperatures of the -1 cm probe (even during winter) are due to radiative heating of the heat-absorbing black marls during sunny periods.

*Figure 1: Example of raw temperature series (1 measurement every 10 min) (A) Typical soil-temperature series recorded with four probes at different depths (from south-facing uphill location). (B) Example of soil-temperature series (from south-facing downhill location) biased because of climatic conditions (snow cover), buried or loosened probes. A full year of temperature measurements is shown in Supplementary Figure 1. High temperature values are observed at -1cm even in winter when black marl heats up in sunny periods.*




*Supplementary Figure 1: Annual time series of raw temperature data from four different depth probes.*

12) In the analysis presented in Figure 6, and also explained and discussed in the text, it is not clear what is the meaning of rainfall above 50 mm. Please could you define it? Rainfall events, flood events with rainfall > 50 mm. In that sense, Figure 6 X axis is the number of rainfalls above 50 mm/h. Please specify it in the figure caption and clarify in the manuscript.

To clarify Figure 6, we will modify the x-axis title, the caption and text as follow:

*x-axis title: Cumulative rainfall (mm) above a threshold*

*Caption Figure 6: Linear correlation between annual total sediment export from the Laval catchment and the cumulative rainfall above an instantaneous intensity threshold of 50 mm/h for the years 2005 to 2019 (blue dots). Regression line is in black; grey shaded area shows 95% confidence interval. Most outliers occur for low cumulative rainfall above the threshold (< 40 mm). (B) Coefficient of determination ($R^2$) between annual sediment export and cumulative annual rainfall above threshold for different intensity thresholds. Optimum correlations are found for intensity-threshold values between 50 and 55 mm/h.*

We also made this variable more explicit in the text line 228:

*Based on annual data records since 2005 and previous work (see Methods section), we established a correlation between the cumulative rainfall above an instantaneous intensity threshold and sediment export (Fig. 6).*

13) Discussion sections 5.2 and 5.3. Once I have read the discussion 5.2, I really miss some information about geomorphological dynamics in humid badland areas and explanations about weathering processes, including some references with similar studies in humid badland areas worldwide. It is true, that then once I started reading section 5.3 I partially found this information. I suggest to link both sections and organize the ideas presented in both previous sections.

- In that sense, some significant references that I missed in the manuscript and that could be checked are:
    - Clarke and Rendell, 2006. Process-form relationships in Southern Italian badlands: erosion rates and implications for landform evolution. ESPL 31.
    - Clarke and Rendell, 2010. Climate-driven decrease in erosion in extant Mediterranean badlands. ESPL 35.
    - Nadal-Romero and Regüés, 2010. Geomorphological dynamics of subhumid mountain Badland areas- weathering, hydrological and suspended sediment transport processes: A case study in the Araguás catchment (Central Pyrenees) and implications for altered hydroclimatic regimes. Progress in Physical Geography 34.
    - Gallart et al., 2013. Thirty years of studies on badlands, from physical to vegetational approaches. A succinct review. Catena 106.
    - Gallart et al., 2013. Short- and long-term studies of sediment dynamics in a small humid mountain Mediterranean basin with badlands. Geomorphology 196.
    - Bollati et al., 2019. Alpine gullies system evolution: erosion drivers and control factors. Two examples from the western Italian Alps. Geomorphology 327.
    - Llena et al., 2020. Geomorphic process signatures reshaping sub-humid Mediterranean Badlands: 2. Application to 5-year dataset. ESPL 45.
    - Llena et al., 2021. Do Badlands (always) control sediment yield? Evidence from a small intermittent catchment. Catena 198.

We feel that paragraphs 5.2 and 5.3 treat quite separate subjects and, as such, we prefer to keep them separate. However, upon rereading paragraph 5.2, we agree it can be sharpened by more appropriate comparisons with other similar study sites. Thus, we have studied the references provided by the reviewer and incorporated these in a revised version of paragraph 5.2.

Line 301: *Numerous studieshave reported annual hysteresis cycles between rainfall or discharge on one hand, and sediment export on the other; these studies, have commonly focused on large catchments, e.g. in the Andes or Himalaya (e.g., Andermann et al., 2012; Armijos et al., 2013; Tolorza et al., 2014; Li et al., 2021). For such large catchments, the annual hysteresis cycle is explained by the role of subsurface water storage (Andermann et al., 2012), dilution effects (Armijos at al., 2013) or variations in the contributive erosive area (Li et al., 2021). Hysteresis cycles for smaller catchments have generally been analyzed at the event scale, and have been interpreted in terms of the proximity of sediment sources and the spatio-temporal heterogeneity of rainfall (Klein, 1984; Buendia et al., 2016). More directly comparable to our results, several studies have been carried out in small (< 15 km²) Mediterranean badland catchments where climate can vary between arid to humid conditions. Llena et al. (2021) reported a seasonal sediment dynamic with lags between sediment production and sediment yield and highlighted the role of the channel network in the sediment transfer. Several catchments in Northern Spain that are very similar in size and lithology to our study site have* *been well studied and suspended sediment transport processes have been reconstructed at the event (Soler et al., 2008; Nadal-Romero et al., 2008) and seasonal scale (Nadal-Romero and Regüés, 2010). Counter-clockwise and clockwise hysteresis loops in these catchments are associated to dry and wet seasons, respectively, and are inferred to be driven by infiltration and saturation processes on hillslopes.*

*The annual hysteresis cycle between rainfall and total sediment export observed in the Draix-Bléone CZO (Fig. 4A) presents two loops with successively anti-clockwise and clockwise patterns, reflecting the rapid seasonal changes in erosion regime in these badlands*