Reviewer 1 comments

Comment 1.1 *The manuscript displays an interesting analysis of the impact of precipitation clustering on the duration and frequency of high discharge events in Switzerland and Europe. It uses appropriate data sets and makes a courageous attempt to stratify the analysis along the large number of degrees of freedom that govern the relationship between (clustered) extreme precipitation and high discharge. Overall this analysis and stratification are useful, but the authors seem to address the topic from a pretty methodological point of view, thereby presenting quite a large number of figures and results that are not always directly meaningful to the reader. In particular the large number of spatial maps don't convey a very clear message of spatial structure in the findings, and some reduction in figures and a more concise display of results would be appreciated. On the other hand, some potentially relevant physical processes are discussed somewhat in the discussion section and annex figures, and some of these would have been interesting to elaborate in somewhat more detail. Particularly the notion of (seasonally dependent) soil moisture memory and the impact of catchment size deserve a more explicit discussion and interpretation of the results.*

Answer: Thank you for your positive comments. We reduced the number of figures and simplified some of the spatial maps to narrow the message of the paper. Regarding the potential seasonal dependence of results, we slightly expanded the discussion in the manuscript and enclose a figure in this response. A more detailed seasonal analysis would however take us a bit far and make the paper even more complex.

Comment 1.2 *L30-31: also Khanal et al (2019) address the role of atmospheric clustering and soil moisture memory on Rhine discharge explicitly*

Answer: Thank you for this reference. We suggest modifying the sentence as follows: "*To our knowledge, the impact of sub-seasonal TCEP on discharge has not been explicitly investigated, except briefly in the case of Switzerland by Tuel and Martius (2021b) and of the Rhine river basin by Khanal et al. (2019). Both argued that TCEP increased the likelihood and duration of high discharge events compared to precipitation extremes occurring in isolation.*"

Comment 1.3 *L49-50: Kew et al (2013) looked at the effect of clustering on the probability to have compounding discharge and coastal surge peaks for the Rhine river. This is an interesting application domain of studying temporal clustering/compound events in this context*

Answer: This is an interesting example indeed involving coastal floods. Bevacqua et al. (2019) also analyse this question: https://dx.doi.org/10.1126/sciadv.aaw5531. We will add the reference in the introduction.

Comment 1.4 *L81: given the relatively small effect of removing this baseflow component (L265-269), and the question of whether the resulting runoff data can be interpreted well I would suggest to leave out this baseflow correction and work with total discharge instead*

Answer: This is a good suggestion, and we propose removing the baseflow correction in the revised version altogether.

Comment 1.5 *L101-102: maybe add explicitly that these factors display a seasonal cycle*

Answer: Good point – here is what we suggest: "*This step is motivated by the fact that high discharge is shaped not only by precipitation, but also by seasonally-dependent surface conditions like snow and vegetation cover, soil saturation or evaporative demand.*"

Comment 1.6 *L116-117: the fact that discharge impact is characterized more by absolute than anomalous discharge also applies to the impact of clustered precipitation on discharge. So I find this methodological inconsistency of using absolute or anomalous values not convincing.*

Answer: We define extreme precipitation and discharge thresholds for each catchment independently, as it would be difficult to choose impact-relevant, absolute thresholds for so many different catchments. In addition, the discharge response to extreme precipitation is very dependent of surface conditions, which explains why the annual cycles in extreme precipitation and discharge are not necessarily in phase (Figures A1-A4). Consequently, it is not inconsistent to use monthly-varying precipitation percentiles to identify extreme precipitation events. We could also have used fixed percentiles, and the results averaged across catchments are in fact not very different (see below Figure R1). However, results can differ substantially for individual catchments, especially ones where the seasonality in extreme precipitation and extreme discharge is not in phase, like the Jura. Choosing fixed percentiles can also make us miss clustered events that nonetheless bring large precipitation accumulations and can have large discharge responses.

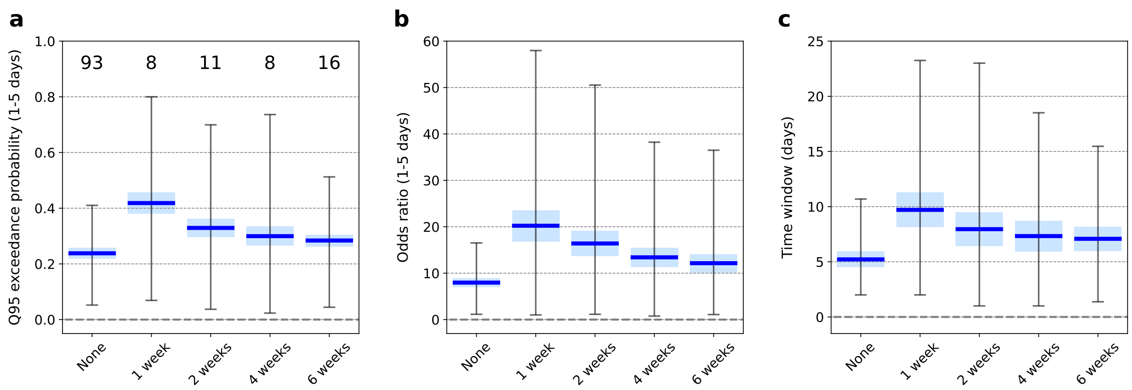

Figure R1: (Fixed (annual) percentiles: compare with Figure 5 of the manuscript) Boxplot of (a) high discharge probability and (b) high discharge odds ratio averaged over day 1-5 following the occurrence of an extreme precipitation event (day 0) for Swiss catchments and various clustering categories. Numbers at the top in (a) indicate the average number of extreme events in the respective categories. (c) Boxplot of response timescale for Swiss catchments and various clustering categories.

Comment 1.7 *L117: "smaller" à "lower"*

Answer: Noted, thanks.

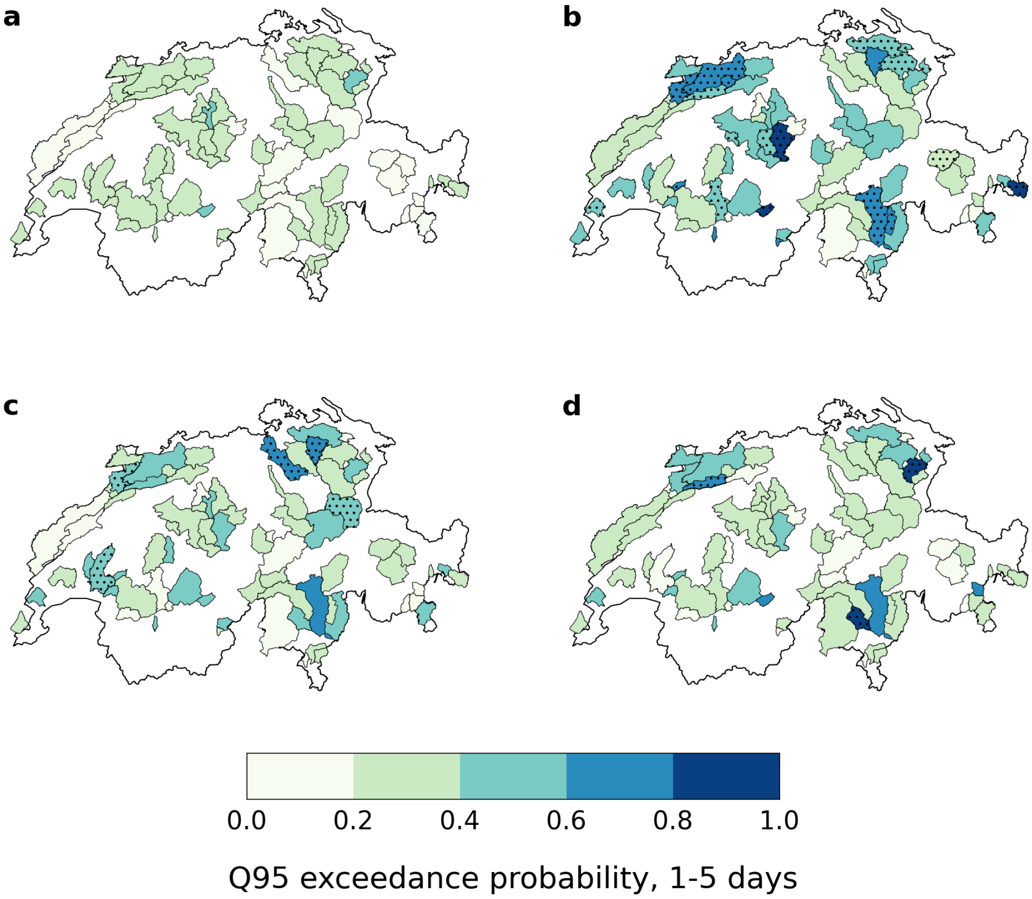

Figure R2: (Fixed (annual) percentiles: compare with Figure 6 of the manuscript) Average high discharge probability in day 1-5 following an extreme precipitation event, for (a) non-clustered, (b) 1-week clustered, (c) 2-week clustered and (d) 4-week clustered events, in the Swiss data. Hatching in (b-d) indicates catchments where values are significantly different from those in (a) at a 10% level.

Comment 1.8 *L126: the 30 and 60 days are not clear to me, I compared them to the max 8 weeks of clustering and couldn't find the logical match. Please explain somewhat better*

Answer: These two time windows correspond to the maximum time horizon we analyse after extreme precipitation events. The discharge response after that cannot be statistically separated from background noise. But having a different value for the Swiss and European data is unnecessarily confusing. In practice a 60-day window works for both. Thus we suggest reformulating this paragraph to make the point clearer:

" *We quantify the effect of temporal clustering of precipitation extremes on discharge by considering several simple metrics. For each catchment and each clustering category, we calculate for each day following extreme precipitation events:*

1. *daily discharge percentiles averaged across all events in the corresponding clustering category;*

2. *daily high discharge probabilities;*
3. *daily high discharge odds ratios*

In practice, we limit the analysis to 60 days after extreme precipitation events, beyond which we do not find a significant discharge response."

Comment 1.9 *L147: what motivated this combination of N and L? It came somewhat unannounced.*

Answer: The idea here is to look at sub-seasonal timescales (hence L between 10 and 40 days), and to set the threshold for "persistent" high discharge at 50% of high discharge days within a given period. To make this all clearer we suggest reformulating the beginning of the paragraph as follows: "*Following Tuel and Martius (2021b), we identify periods of persistent high discharge at sub-seasonal timescales as periods of 10 to 40 days when discharge exceeds its 95*$^{\text{th}}$ *percentile at least half of the time. In practice, we look for L-day periods with at least N high discharge days, with* $(L, N) \in \{(10, 5), (20, 10), (40, 20)\}$*. We also consider an additional category,* $(L, N) = (10, 1)$*, to characterise non-persistent high discharge events.*"

Comment 1.10 *L166: "less variability": less than what?*

Answer: We meant less uncertainty compared to the clustered categories, but the sentence can be removed altogether.

Comment 1.11 *L208: please refer forward to discussion section when introducing "karst effects" (and in the discussion section: please explain in some detail what this effect is about)*

Answer: We suggest adding the following detail to the discussion: "*The Jura is a region that shows strong karst effects (where soluble limestone rocks dominate, leading to high permeability and complex subsurface flows) [...].*"

Comment 1.12 *L222-224: this should be discussed in the section on Swiss results, not here*

Answer: Good point, we will move it there.

Comment 1.13 *L225-228: also this feels that it belongs to the methods section, not to the results section*

Answer: You are right; we suggest moving several sentences that explain how we tested for the influence of elevation and of precipitation magnitude to the methods section:
"*Finally, because the phase of the precipitation and its magnitude impact the discharge response to precipitation extremes, we also analyse the Swiss results as function of catchment elevation – a rough proxy for the influence of snow – and of extreme precipitation magnitude. We separate Swiss catchments into two groups (one with mean elevation below 1500m, the other above 1500m) and average the results for each group. Discharge in high-elevation catchments is typically snow- or glacier-dominated, and we expect the discharge response to precipitation extremes to differ with elevation. We do not investigate the influence of elevation in the European data; first, because it covers a much narrower range of elevations (only 10 catchments have a mean elevation*

higher than 1500m); second, because mean elevations are less representative of the elevation distribution in larger catchments; and third, because unlike in Switzerland, the presence of snow is dictated by other catchment characteristics (chiefly latitude). We also explore the sensitive to the event magnitudes, for this we separate extreme precipitation events in each catchment into two groups based on their absolute magnitude (bottom and top half), and average the results across catchments, for each group separately."

Comment 1.14 *L236-237: "Cumulative precipitation ... periods.": I don't know what you want to convey with this statement*

Answer: We meant to say that even though both persistent and non-persistent high discharge periods were preceded by high cumulative precipitation, the precipitation percentiles were even larger before persistent periods than non-persistent ones. We suggest reformulating as follows: "*Most high discharge periods, whether persistent or not, are preceded by intense precipitation (90$^{\text{th}}$ percentile or higher) in the three preceding days (Fig. 11-a,b,c). Still, accumulated precipitation tends to be even larger before persistent periods than before non-persistent ones, except at high elevations. The difference is largest for the most persistent periods (compare panels a and c), especially in the Jura and Southern Switzerland where values larger than the 98$^{\text{th}}$ percentile are found.*"

Comment 1.15 *Fig 11: the set of panels don't convey a very clear message of a spatially meaningful structure. Also it takes me a long time to make up my mind of what is actually shown in the different rows and columns. Also: what does "non-significant" (grey shading) mean in these panels? I don't really understand the grey shading in all basins of panel 11-d. I feel this display of information is too extensive.*

Answer: Figures 11 and 12 were indeed too complex, especially since we hardly discussed the spatial variability of the results. We propose to replace Figure 11 with more simple boxplots that still convey most of the information (see Figure R3), and to move Figure 12 to supplementary (also as a boxplot).
Additionally, your comment made us realise that we had forgotten to explain how we assessed the significance of the results shown on Figure 11. For each catchment and (L,N) value, we obtain a number (say, m) of persistent high-discharge periods. We calculate our various metrics (cumulative precipitation percentile in days 0-2 preceding those periods, etc.) and then assess their statistical significance by comparing them to metrics calculated from 1000 randomly generated samples of m periods of same length L at about the same time of year ($\pm$20 calendar days) as the actual persistent high discharge periods. We will add the following paragraph to the methods section:
"*The cumulative precipitation percentiles are calculated with respect to all periods of the same length within $\pm$20 calendar days of observed persistent high discharge periods. Their statistical significance is assessed with a Monte-Carlo approach. For each catchment and (L, N) category, assuming we observe m periods of persistent high discharge, we generate 1000 random samples of m periods occurring within $\pm$20 calendar days of observed high discharge periods, calculate cumulative precipitation percentiles for these random periods and obtain their 90$^{\text{th}}$ percentile. Observed percentiles are then said to be significant if they exceed this value.*".

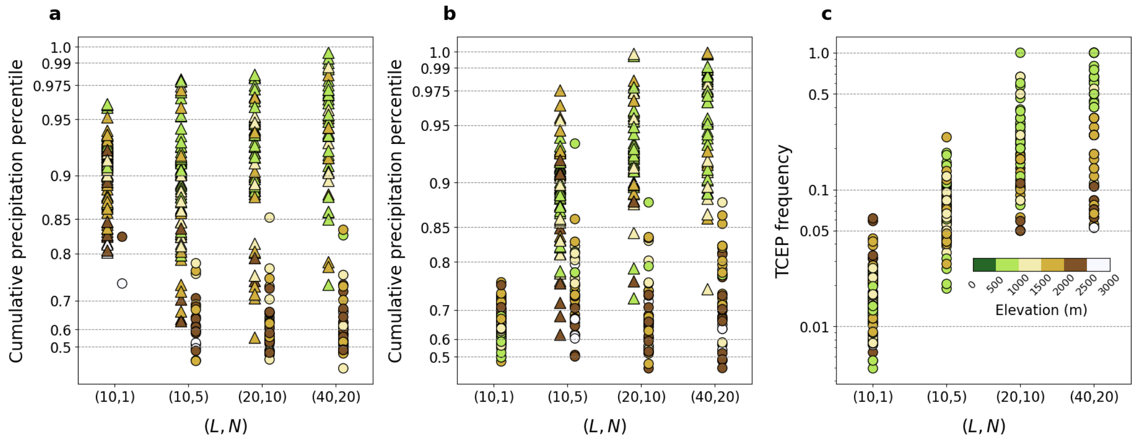

Figure R3: (a) Average percentile of cumulative precipitation during day 0-2 before and (b) during persistent high discharge periods, and (c) fraction of high discharge periods with two or more extreme precipitation events (TCEP) between day 2 before to the end of the period, averaged by catchment for the Switzerland data. Values are coloured according to catchment-mean elevation, to highlight the difference between high- and low-elevation catchments. In (a-b), triangles (resp. circles) indicate values that are (resp. are not) statistically significant at a 10% confidence level (see methods).

Comment 1.16 *L237: "increase": of what relative to what?*

Answer: Please see our proposed reformulation of this sentence in your previous comment.

Comment 1.17 *L245: events are classified by percentile, by persistence, by significance. It's not easy to disentangle all these attributes. What's the key point you want to convey here?*

Answer: The key point is that persistent high discharge periods are characterised by intense precipitation accumulations, unlike non-persistent periods. The statistical significance of the results is here secondary (more of a safety check, really). We can reformulate slightly to make it easier to read: "*Precipitation accumulations during high discharge periods are by contrast very different between persistent and non-persistent periods (Fig. 11-b). Small precipitation accumulations characterise non-persistent periods, whereas persistent periods are associated with high event precipitation totals except at high elevations.*"

Comment 1.18 *L246: "often result from TCEP.": how can I deduce this statement from these panels? They look quite similar to panels 11d-f. And how does "cluster frequency" shown in Fig 11 translate to TCEP?*

Answer: The revised Figure 11 (Figure R3 here) shows the distribution of TCEP frequency during high discharge periods instead of spatial maps and should hopefully be easier to understand. Across catchments, we see for instance that 5-100% of persistent periods with (L,N)=(40,20) are associated with two or more precipitation extremes (i.e. TCEP), while these figures are only 1-7% for (L,N)=(10,1).

Comment 1.19 *L249: "Overall, the connection to TCEP is weaker for less persistent high discharge periods.": how can I see this from the figure? What should I compare to what to understand this statement? Same for statement in L254*

Answer: Again, we hope that the revised Figure 11 (Figure R3) should be clearer.

Comment 1.20 *L272-273: I didn't read the Bevacqua reference, but doesn't this combination of arguments imply that it is the change of the wet-day frequency and R95p that is responsible for this, without a significant change in clustering?*

Answer: Bevacqua et al. (2020) do find a significant change in cyclone clustering, namely that cyclone clusters will get shorter, which partly compensates for the increase in average precipitation associated with individual cyclones. Changes in clustering (of both signs) also matter for climate projections, even if in most regions the projected trends in extreme precipitation frequency dominate clustering trends (Tuel and Martius 2021a).

Comment 1.21 *L274-275: this soil moisture memory plays a smaller role in winter/spring time, where the highest discharge occurs, I would reckon. A seasonal instead of a spatial analysis would have been more interesting (see also statement in L306)*

Answer: Soils indeed tend to be more saturated in winter so that soil moisture memory may be less important then. In summer, dry soils can nevertheless be less permeable and conducive to a large surface discharge response. The seasonality is worth exploring, which is challenging because it involves baseline discharge (Figures A1, A3), surface conditions and precipitation characteristics (Figure A2, A4). A detailed seasonal analysis by catchment is also difficult due to the potentially small number of clustered events (some clustering categories already have few events, so divided into four seasons they would be too few to say something remotely robust).
One way to tackle this problem is to simplify the clustering categories by grouping together all extreme events clustered at 1-4 weeks, and all others into a "non-clustered" category. This yields enough events for each catchment and category to get somewhat stable estimates of the various metrics. Results are shown on Figure R4 (on which we separated catchments based on elevation, since discharge in high-elevation catchments typically has a sharp peak in summer). As you suggested, the effect of clustering seems weaker in winter. Still, this remains a very rough analysis: as we said above, to reach robust conclusions, it would be necessary to simultaneously take into account spatial variability, seasonality in extreme precipitation and discharge, and variability in clustering timescales. But first, before even talking about clustering, one should detail the seasonality in the discharge response to extreme precipitation. This consequently goes beyond the scope of our study, which we see more as a first step to understanding the role of temporal clustering.
We suggest expanding the discussion section to bring up those points as follows: "*The role of the pre-conditioning through soil moisture is likely to vary across the year. In winter, soils are more likely to be saturated, so that the discharge response to small clustering windows may not be significantly higher. Yet, to explore the seasonality in TCEP impact on discharge, one would have to take into account seasonality in discharge and extreme precipitation magnitude (Figures S1-S4), in TCEP frequency (Tuel and Martius 2021a,b) and in surface conditions. All these factors make for a complex analysis which goes beyond the aim of the present study and would*

likely require hydrological modeling, since at seasonal timescales clustered events might be too few to obtain robust statistical results.".

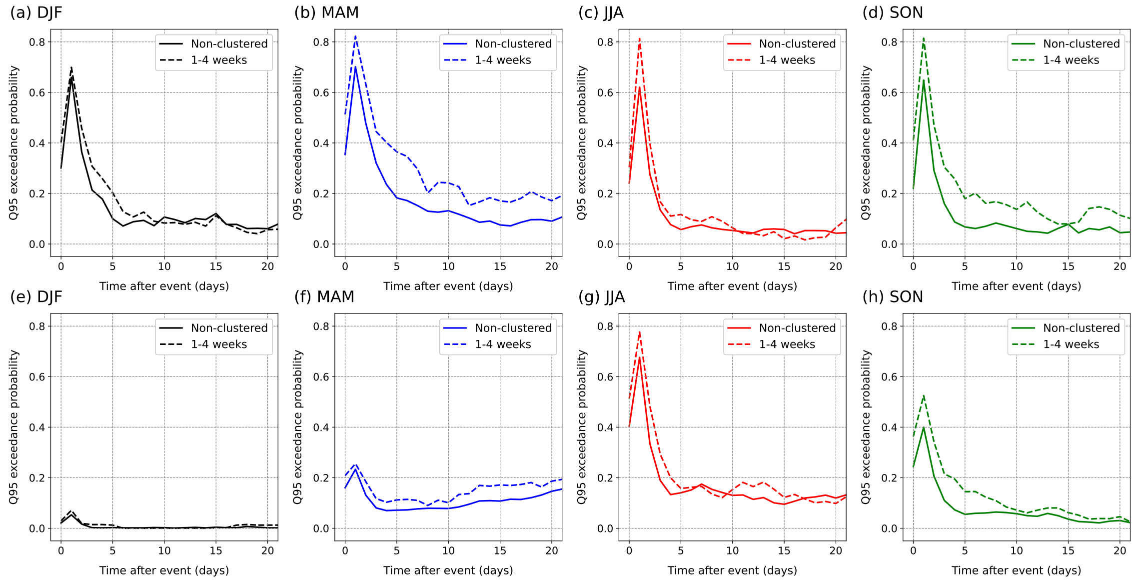

Figure R4: (a-d) Daily average probability of high discharge (defined as the exceedance of the respective 95th daily discharge percentile) averaged across FOEN catchments with a mean elevation of 1500m or less, for the non-clustered and 1-4 week clustered categories of extreme precipitation, in (a) DJF, (b) MAM, (c) JJA and (d) SON. (e-h) Same as (a-d), but for FOEN catchments with mean elevation larger than 1500m.

Comment 1.22 *L282: figures 11 and 12 don't show soil moisture impacts*

Answer: We suggest removing this sentence.

Comment 1.23 *L307: "largest events": discharge or precipitation events?*

Answer: We meant extreme precipitation – we should reformulate by saying "*The heaviest extreme precipitation indeed generally occurs in summer and fall (Figure S1)* [...]."

Comment 1.24 *L317: double use of "high"*

Answer: Thanks for noticing – we can replace by "*large high discharge probabilities persist for much longer.*"

Comment 1.25 *L324: 1000 km2 is pretty small. I would assume that clustered precipitation extremes can give high discharge in much larger basins*

Answer: There was a typo here, the figure should be 10000 km^2 (see Figure 14).

Comment 1.26 *L330: "vary" à "varies"*

Answer: Noted, thanks.

Comment 1.27 *L336: this seasonal signature is of great interest and could be promoted to the main text*

Answer: These maps are not a result of our study and would divert the reader away from the main point, since we do not explicitly analyse seasonal variations.

Comment 1.28 *L351-352: both "antecedent soil moisture" and "timing of precipitation" refer to (temporal and spatial) clustering, so are attributes that are within scope of the current analysis*

Answer: You're correct and the wording was inadequate. We suggest "*Still, whether high discharge translates into a flood, particularly a disastrous one, depends on other factors related to the exposure and vulnerability of human systems, like the presence of infrastructure and its management, or the performance of early warning systems (Merz et al. 2021). The most disastrous floods tend to result from compounding effects between hazards, exposure and vulnerability.*"

Comment 1.29 *L355: do "cross-catchment analyses" imply that you would show more results like fig 14? That would be great!*

Answer: The idea would indeed be to analyse several catchments with similar characteristics together. Currently, our cross-catchment analyses are rather coarse, since we average all available catchments together or, at best, divided into two groups based on elevation, which further research should refine.