Reviewer 3 comments

Comment 3.1 *l.110 You identified clusters of precipitation events with a pairwise approach, categorizing each event depending on the distance in time from the previous closest event. The whole precipitation cluster is not identified, and consequently some events belonging to different categories may in reality be dependent, being part of the same precipitation cluster. I imagine that some events in a specific category may be preceded by other precipitation events and the whole story may influence the final discharge characteristics, rather than only the last two events of the cluster. Does this occur in your datasets? How much do you think this may affect the results?*

Answer: You are correct that extreme precipitation events are categorised based on their distance in time to previous events only. For instance, an event occurring on day 100, 10 days after another event on day 90, would be put in the "2-week" category, but the event on day 90 may belong to a different category, depending on the time of the previous event. This choice is consistent with the physical perspective: the first event in a cluster conditions the discharge response to later events, but not the other way around. Thus, it doesn't make sense to automatically put the first event of a cluster in the same category as the later events.
If we did that, however, notwithstanding the physical inconsistency, we expect the results to be generally less significant and the discharge response after clustered events to be closer to the one after non-clustered events (since many events previously categorised as non-clustered would then be classified as clustered).

Comment 3.2 *l.99 and l.114 Have you tried also lower precipitation quantiles or different pairs of precipitation and discharge quantiles? Are the results sensitive to this choice? Also (l.117) I do not think to be meaningful to chose discharge quantile starting from the chosen quantile of single precipitation events. Mainly considering that you show that extreme discharge is more often caused by precipitation events close in time than isolated.*

Answer: The results are qualitatively unchanged for lower precipitation percentiles (e.g. 95th percentile, compare Figures R1 and R2 with Figures 3 and 5 in the manuscript). The choice of a discharge percentile lower than that for precipitation is common in hydrology. It is meant to capture as many of the discharge extremes following extreme precipitation events as possible. Extreme discharge events are indeed not exclusively triggered by extreme precipitation events, and surface conditions in particular strongly modulate the discharge response to extreme precipitation. An extreme (>99th percentile) precipitation event therefore does not necessarily lead to a discharge response above the 99th percentile.

Comment 3.3 *l.77 Why did you not operate the same selection of catchments used for the Switzerland dataset (human influence, lakes, stationarity of the series...)?*

Answer: We forgot to mention that we did test for the stationarity of the series in the GRDC dataset as well (by testing for significant trends in annual discharge maxima with a Mann-Kendall test; for the Swiss data, the stationarity analysis is performed by the FOEN). We would specify it in the revised version. As to the other criteria, however, the metadata of the GRDC

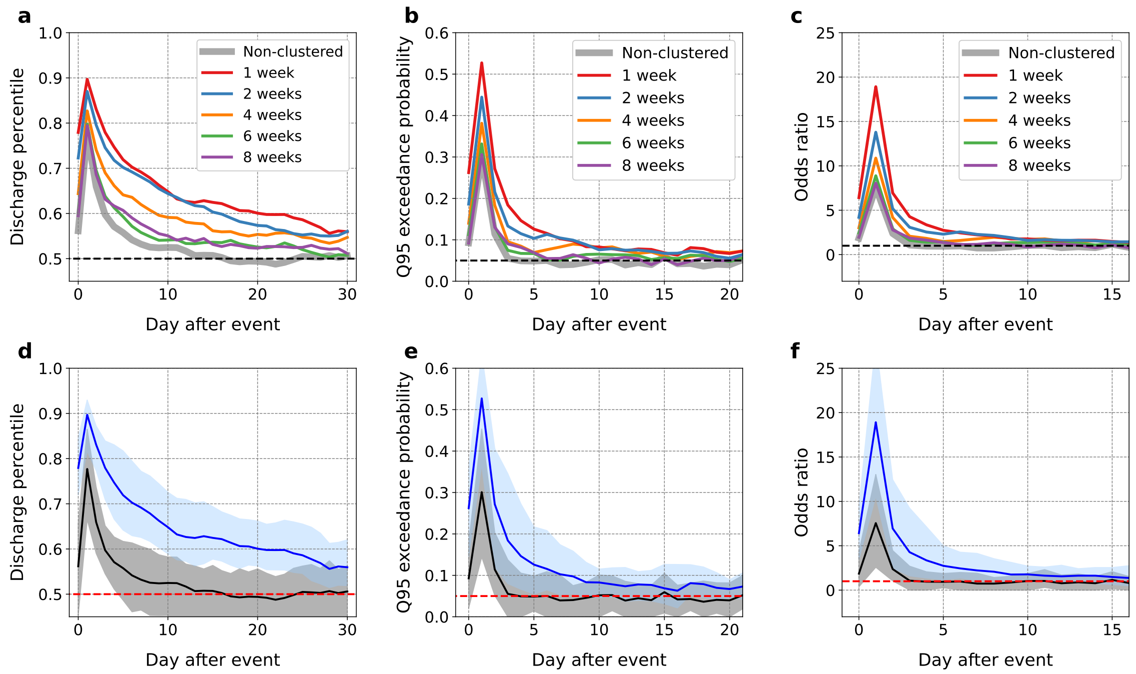

Figure R1: Daily average (a) discharge percentile, (b) probability of high discharge (defined as the exceedance of the respective 95^{th} daily discharge percentile) and (c) odds ratio of high discharge, averaged across FOEN catchments with a mean elevation of 1500m or less, for the different clustering categories of extreme precipitation (exceeding its 95th percentile). Black dashed lines indicate baseline values of 0.5 for discharge percentiles in (a), 0.05 for high discharge probability in (b) and 1 for odds ratios in (c). (d-f) Same as (a-c), but for the non-clustered (black) and 1-week clustered (blue) categories only, with the 95% range of values across catchments shown in light grey and blue shadings, respectively. Baseline values are shown by horizontal red dashed lines as in (a-c).

dataset is unfortunately insufficient to determine which catchments strongly anthropogenically influenced or contain major lakes.

Comment 3.4 *l.145-147 From this section it seems that all combinations of periods here reported are considered as persistent high discharge periods, included the pair (10, 1). However, also reading l.235, this is a non persistent period, right?*

Answer: You are correct and the initial formulation was misleading. We suggest reformulating 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^{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 3.5 *Fig.11 I did not understand what is reported with cluster frequency.*

Answer: We should have said "TCEP frequency", meaning the fraction of high discharge periods with two or more extreme precipitation events.

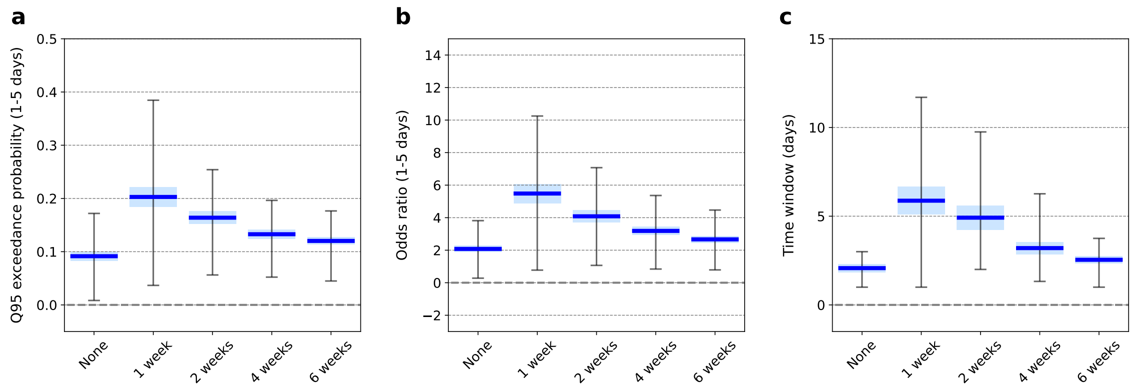

Figure R2: Boxplot of (a) high discharge probability and (b) high discharge odds ratio averaged over day 1-5 following the occurrence of an extreme precipitation event (day 0, 95th percentile) for all FOEN catchments and various clustering categories. (c) Boxplot of response timescale for all FOEN catchments and various clustering categories.

Comment 3.6 *l.250 Why did you not use a measure, normalized for L, for example, that is comparable between the different pairs of (N, L)?*

Answer: For persistent high-discharge periods, we already choose N=L/2 for all three categories (L=10, 20 or 40 days). We suggest making it explicit in the methods by adding "*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.*

Comment 3.7 *l.293 Have you considered also the possibility of using rainfall rather than total precipitation?*

Answer: The available gridded precipitation data does not discriminate between solid and liquid precipitation. Admittedly, we could use air temperature as a proxy, but only one value per day is available (and values in mountain areas where snow is most frequent are less reliable since stations are less numerous there). We prefer here to keep the analysis simple, even if it adds limitations, and explore the role of snow at first order only (Figures 3-4).

Comment 3.8 *Fig. 13 Why did you choose an absolute precipitation magnitude rather than a magnitude relative to the quantile in each specific catchment (i.e. the excess over the threshold)?*

Answer: The selection of the two classes is done on each catchment separately, so we do use a relative precipitation threshold. To make it clear we suggest reformulating the corresponding sentence (l.301) as follows: "*A simple way to tackle this is to separate, for each catchment separately, extreme precipitation events into two groups based on their absolute magnitude.*"

Comment 3.9 *l.69 these catchment → these catchments*

Answer: Corrected, thanks.

Comment 3.10 *l.70 The data is → The data are*

Answer: Corrected, thanks.

Comment 3.11 *l. 71 We selected catchments among all available ones based on several criteria → I would rewrite it like this: "Among all available catchments, we selected a subset of them based on several criteria:"*

Answer: Good suggestion, thanks.

Comment 3.12 *l.76 Daily discharge data for Europe comes from the Global Runoff Data Center dataset (GRDC). → The second dataset consists of daily discharge data for Europe and it comes from the Global Runoff Data Center dataset (GRDC).*

Answer: Good suggestion, thanks.

Comment 3.13 *l.92 yielding two precipitation datasets → I would remove this. You had two precipitation datasets also before.*

Answer: Good suggestion, thanks.

Comment 3.14 *l.109 events which occurred between n-1 and n weeks after another event are put into the "n-week" category, where $n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$. → I would substitute another event with the previous extreme event, this to make it clear that you are selecting the smallest window, or otherwise rewrite it more similarly to the explanation in the caption of Fig. 2.*

Answer: Good point; we suggest to reformulate as follows: "*For each catchment, we then classify precipitation extremes into different categories based on their degree of sub-seasonal temporal clustering (Figure 2-a). For each extreme event, we look for the previous event closest in time, by exploring progressively longer time windows of n weeks ($n \in \{1, 2, 3, 4, 5, 6, 7, 8\}$). We choose the first (i.e., shortest) window that contains the closest previous event.*"

Comment 3.15 *l.122 I think this part not to be reported clearly. In particular, I find misleading that you write that the probability is averaged across all extreme precipitation events belonging to a clustering category. Also, I would explicitly specify that the days (60 and 30 days) are the ones after each precipitation event. This is what I would probably write, but feel free to write it differently:*
For each catchment, clustering category, and for each of the 30 (Switzerland data) or 60 (GRDC data) days following extreme precipitation events, we calculate
1. daily discharge percentiles, averaged across all extreme precipitation events in each clustering category;
2. daily high discharge probabilities;
3. daily high discharge odds ratios.

Answer: Thank you for the comment. We suggest reformulating this paragraph as follows:
"*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 3.16 *Sec. 3.1.1 I believe that it is not always clear if the comments reported refer to all the catchments or only low/high elevation ones. In ll 172-187, for example, you are referring only to Fig. 3, that collects results of low elevation catchments but at the same time you are comparing it with Fig. 5, that, if I understood correctly, reports results for both subsets.*

Answer: The original Figure 5 indeed grouped together both high- and low-elevation catchments. We suggest to replace it with one that separates the two groups of catchments:

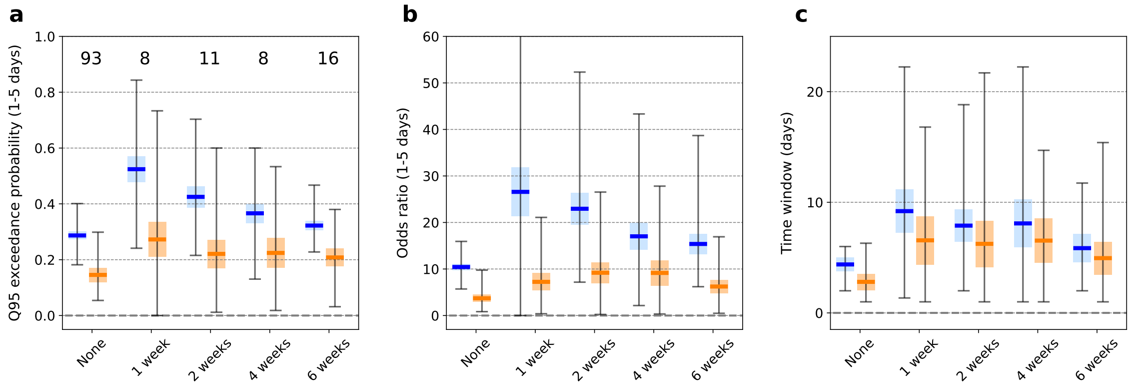

Figure R3: Boxplot of (a) high discharge probability and (b) high discharge odds ratio averaged over day 1-5 following the occurrence of an extreme precipitation event (day 0) for FOEN catchments with mean elevation lower (blue) and higher (orange) than 1500 m, and various clustering categories. Numbers at the top in (a) indicate the average number of extreme events in the respective categories. (c) Boxplot of response timescale for FOEN catchments and various clustering categories.