Reviewer 2 comments

Comment 2.1 *I think that this is a well-written and –structured study which highlights the importance of looking at precipitation events beyond the ones in immediate temporal proximity to high flow events. I have one methodological concern though regarding the event definition procedure and its potential impacts on the results/conclusions. In addition, I suggest to reconsider figure choice as the results of the Swiss and European analyses are very similar (one could maybe even say redundant). I think that narrowing down the selection might help to better focus the reader's attention on the main points.*

Answer: Thank you for your positive comments. We hope our responses will answer your concerns.

Comment 2.2 *I suggest to clearly highlight the research gap after the paragraph ending on l.54. What remains to be investigated given the results presented in the authors' previous study?*

Answer: Thank you for this comment; we had not made it sufficiently clear in the manuscript what the novelty of the present study is. We suggest reformulating the last paragraph of the introduction as follows: "*How the impact of TCEP evolves with the timescale of clustering remains however unexplored. Each extreme precipitation event can in principle be associated with a clustering timescale, depending on the lapse of time since the previous extreme event. Tuel and Martius (2021b) only looked at 3-week clusters, analysing together extreme precipitation events at the beginning and end of the clusters. There is also interest in going beyond the borders of Switzerland, to consider a larger number of catchments with more diverse climates and less spatial dependence. Here, we quantify the effects of TCEP on discharge in Switzerland and Europe, specifically on the occurrence and temporal persistence of high discharge. We classify extreme precipitation events according to their clustering timescale, and analyse the sensitivity of results to that timescale, as well as to catchment area and to extreme precipitation magnitude.*"

Comment 2.3 *An overlap of only 10 years between precipitation and discharge events (l.72) seems very little given that the study focuses on extreme events.*

Answer: The thresholds we select are still low enough to capture a sufficient number of extreme precipitation (and high discharge) events (about 3-4 per year for precipitation, and 18 per year for discharge). Admittedly, some of the clustering categories may have very few events, but we take it into account to some extent by showing and discussing the uncertainty in and statistical significance of the results (Figures 6, 7 or 11 for instance).

Comment 2.4 *Please specify how the baseflow filter parameter (l.82) was determined and why it does not vary in dependence of catchment properties.*

Answer: Following a comment by the previous reviewer, we suggest to show the results that include the baseflow since they do not differ much from results with baseflow removed (less so at high elevations).

Comment 2.5 *Precipitation extremes are defined as anomalies (l.97-100) while discharge ex-*

tremes are defined using a fixed threshold (l.115). While I understand the desire to consider precipitation events potentially co-occuring with wet antecedent conditions, I think this is not necessarily achieved by using a seasonally varying precipitation threshold (some events relevant in terms of discharge may still be missed). I think that choosing events based on the discharge rather than the precipitation events would be more consistent with the aim of the study. I think that the effect of choosing one over the other event identification procedure on the results of the analysis should be demonstrated in a small sensitivity analysis (e.g. on a small subset of catchments). Specifically, it would be important to know what effect (a) choosing events based on precipitation has compared to choosing events based on discharge and (b) what an effect choosing a seasonal rather than a fixed precipitation threshold has.

Answer: Some precipitation events relevant for the discharge response are undoubtedly missed by our selection methods, but then this would also be the case with fixed (annual) thresholds, since in some catchments high discharge events do not always co-occur with the heaviest precipitation events. The point of the forward and backward approach is precisely to show that results are consistent whether we start from extreme precipitation events or from high discharge events. We can make this explicit in a revised version.
We chose monthly-varying percentiles to define extreme precipitation events because 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). 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 (Figure R1). Choosing fixed percentiles can also make us miss clustered events that nonetheless bring large precipitation accumulations and can have large discharge responses.

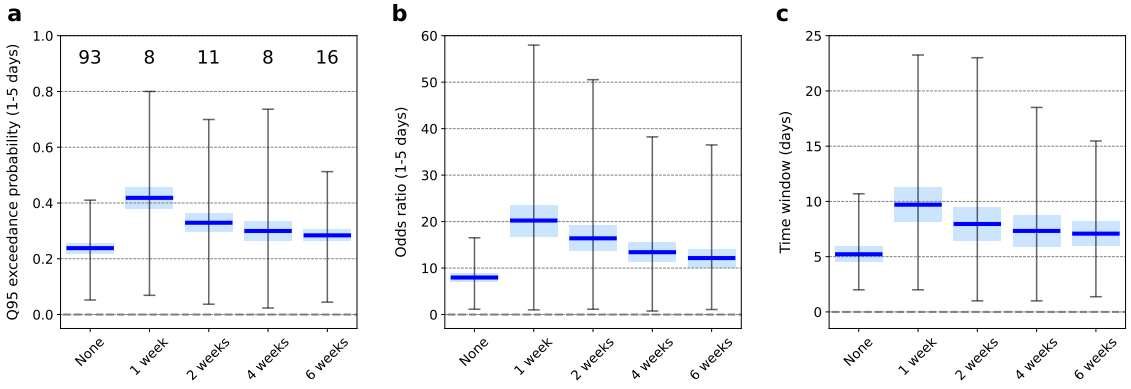

Figure R1: (Fixed (annual) percentiles: compare with Figure 5 of the manuscript) Boxplot of (a) high discharge probability and (b) high discharge odds ratio averaged over day 1-5 following the occurrence of an extreme precipitation event (day 0) for Swiss catchments and various clustering categories. 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.

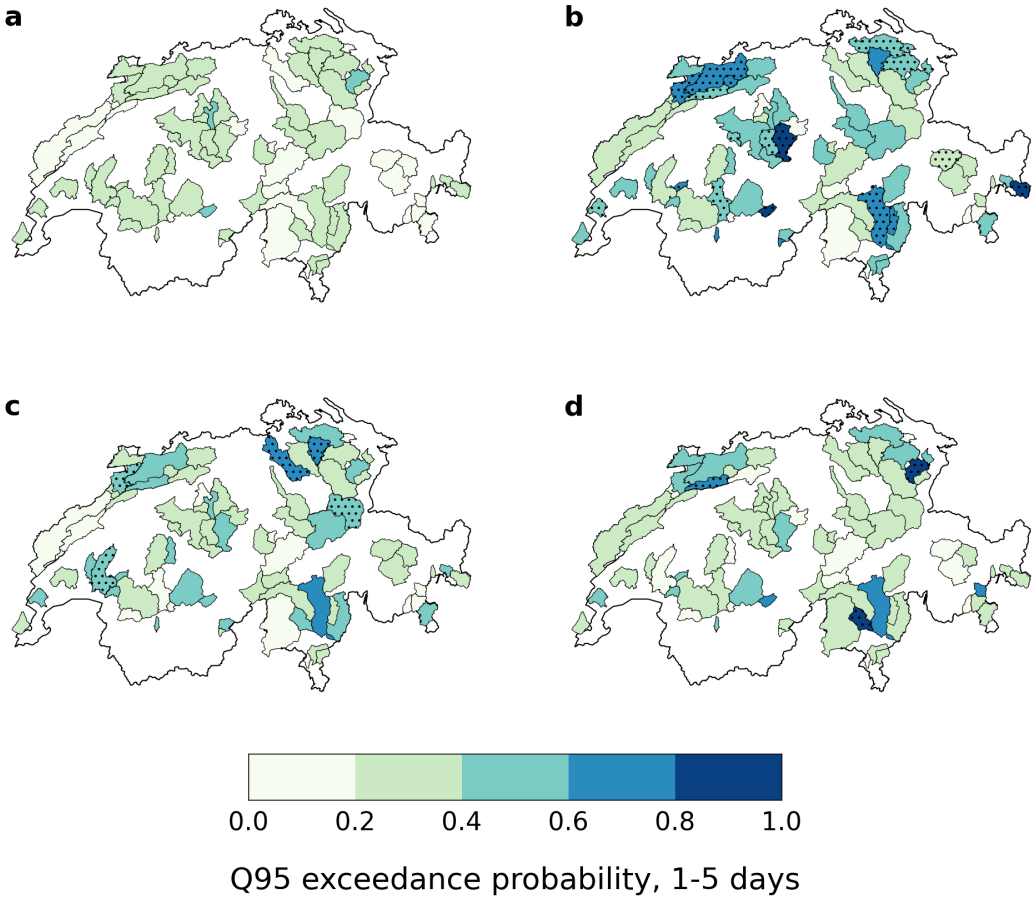

Figure R2: (Fixed (annual) percentiles: compare with Figure 6 of the manuscript) Average high discharge probability in day 1-5 following an extreme precipitation event, for (a) non-clustered, (b) 1-week clustered, (c) 2-week clustered and (d) 4-week clustered events, in the Swiss data. Hatching in (b-d)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

Comment 2.6 *Were the percentiles computed empirically (l.155)?*

Answer: Yes, they were.

Comment 2.7 *What about uncertainty in percentile computation because of small event sample sizes (l.159)? How does the analysis deal with cases where there are only very few events of the same length and what event sample sizes are we generally talking about?*

Answer: This 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^{th} *percentile. Observed percentiles are then said to be significant if they exceed this value.*".

Comment 2.8 *I think that the European analysis is nice to put the Swiss analysis into broader perspective. However, I also think that it leads to the presentation of slightly too much material. I therefore suggest to just show a subset of the European results. I would try to focus the reader's attention to the most important information.*

Answer: We agree that the figures were too numerous and sometimes a bit too complex. We suggest replacing Figures 11 and 12 by more simple boxplots (since we do not discuss the spatial variability much in the end) and to move the revised Figure 12 to the supplementary material, since it is similar to the Swiss results.

Comment 2.9 *l.3: 'this question' lacks a reference in the previous sentence.*

Answer: We can replace by "this topic", which refers to the "potential effects on discharge" (cf. "Its potential effects on discharge have received little attention. Here, we address this topic by analysing...").

Comment 2.10 *l.10: the influence of temporal clustering on what? I understand that the statement refers to high flows and think that this could be made clearer.*

Answer: You are correct; we suggest reformulating as follows: "*The influence of temporal clustering on discharge decreases as the clustering window increases; beyond 6-8 weeks the difference in discharge response with non-clustered events is negligible.*"

Comment 2.11 *l. 310: specify what 'this bias' refers to.*

Answer: We suggest "*Likewise, for some catchments, TCEP events occur in the season with the largest precipitation extremes, like in Southern Switzerland, which can bias the result since clustered events will also tend to be the heaviest.*"

Comment 2.12 *Figure 1: suggest to reconsider color scale choice as a continuous variable is presented using a non-continuous color scheme.*

Answer: This is a standard color map for elevation, so we would prefer to keep it as it is.

Comment 2.13 *Figure 2: I like figure 2. However, the horizontal lines in panel b should be*

explained in the legend.

Answer: You are right: "*The high discharge probability threshold of 0.1 is shown by the horizontal red line, and the baseline high discharge probability (0.05) by the horizontal black dashed line.*"

Comment 2.14 *Figure 11: legend entry for grey color needed.*

Answer: The grey shaded catchments were the ones where the result was not statistically significant; in the revised Figure 11 this would be made explicit.