Reply to RC2

We thank the referee for his/her review and constructive comments. Original review comments are shown in **black** while our replies are provided in blue.

The manuscript "Flood drivers and trends: a case study of the Geul River Catchment over the past half century" by Tsiokanos et al., analyses the long-term temporal variability of flood drivers for the Geul river catchment. The study adopts an interesting multi-temporal approach to analyze temporal trends of floods and their drivers and finds that 1-day extreme precipitation alone does not explain flood changes, rather heavy prolonged rain, and wet initial conditions. The manuscript is well written, and the analyses and results are presented in a convincing way. Please find my comments below:

"Major comment"

1. **The aim of the study should be clarified. Is it to develop a methodology (L61-64) or to understand flood trends and their drivers in the catchment (L69-72)? These lines appear quite disconnected in the introduction. Furthermore, is the multi-temporal trend approach new (L7, L61-64), or it was proposed by Hannaford et al. (2021) and Murphy et al. (2020) as stated in line 52-53? Please clarify.**

Answer: The main objective of this study is to identify the primary drivers of high-flow/flood events in the Geul river catchment and examine their long-term trends. To achieve this, we employ an event-based approach (examining the relative contributions of extreme precipitation, prolonged heavy rainfall, extreme initial conditions, and compound extremes in generating high flows) and we use a multi-temporal trend analysis. Multi-temporal trend analysis is not novel, as it has been employed to detect temporal variabilities in the past. However, in this work, we build on the multi-temporal approach and propose a new methodology to assess the consistency or stability of trends in this analysis.

We aim to contribute valuable insights to the Geul area without presenting the article solely as a case study. Our combined approaches (integrating an event-based approach with multi-temporal analyses) and the proposed trend consistency methodology can be applied to diverse studies. Furthermore, in future work, we seek to extrapolate our findings to yield useful outcomes for similar regions worldwide (Section 4.2). We will further clarify the focus of the study in the revised version accordingly.

"Specific comments"

1. **"critical precipitation" terminology. In several parts of the manuscript (abstract, introduction and discussion) the authors draw conclusions on the "critical precipitation (precipitation that leads to floods)". It is not fully clear to what of the analyzed precipitation indices they refer to. Please clarify.**

Answer: With the term "critical precipitation" in the aforementioned sections we refer to the consistently increasing trends in winter. In this respect, we find that the rise in severe precipitation is mostly caused by more rain on wet days and that prolonged events have impacted the area and caused floods in winters. At the same time, we show that heavy storms in combination with wet antecedent conditions should be considered. As a result, with the term "critical precipitation" in winter, we refer to the increases in prolonged heavy events P_{kD} for k > 3 days and/or P_{MD} but also to the increases in daily extremes in P_{1D} and P_{95} as together with wet antecedent conditions they can lead to compound events and thus flooding. We will further clarify this in the revised version.

2. **L145-148: It is not clear if these lines describe an extra criterion used. How do you practically ensure that P_{MD} is higher than P_{99} ? What do you do when this is not the case (L148)?**

Answer: The P_{MD} events are defined using the 95th percentile of all 4-day accumulated (rolling) precipitation sums and the P_{99} events using the 99th percentile of wet days (days with more than 1 mm precipitation). For each of the five precipitation stations considered the P_{MD} 95th percentile was calculated and it was found to be higher than the 99th percentile used for the definition of P_{99}. Line 148 refers to very extreme events when the 24-hour precipitation events can cause at the same time both P_{99} and P_{MD} which is unavoidable (especially in the way we defined P_{99}, using two days -see definition line 180-), and not in cases when the P_{MD} 95th percentile is lower than the P_{99} (which is not feasible). These lines will be reformulated to avoid confusion. Generally, this effect is taken into account in the results (see lines 256-260) and is further discussed in subsection 3.1.1, especially in lines 304 – 308.

3. **L158: how is FE defined?**

Answer: Past flood events (FE) come from the definition/description given earlier in subsection 2.2 (Data sets), lines 93-98. We will provide additional clarification in the text, or we may opt not to use the acronym "FE" due to the limited number of occurrences in the manuscript.

4. **L207: "Trends in P_{kD} are based on the. annual maximum values". What does it mean? Do you refer to annual maximum discharges and the fact that P_{kD} is calculated using k days hat preceding flood events? Please clarify.**

Answer: Trends in P_{kD} are not connected to the occurrence of Q_{max}. They are calculated based on the highest k-day total precipitation per year. So actually, a summation moving window with different lengths (1, 3, 5, 7, 10, 15, 30, and 40 days) is applied over the whole time series from the 1950s to 2021 and the annual maxima (in the season half-years) are extracted. This will be also clarified in the revised version of the manuscript to avoid confusion, especially with Table 4 (correlation coefficients between the winter half-year discharge maxima and their antecedent k-day precipitation depths).

5. **L218-220: Why are different assumptions used for the MK test for precipitation and discharge trends? Why do you account for autocorrelation in annual maximum discharge series? Annual maximum values are typically considered uncorrelated by construction as they belong to different blocks/years.**

Answer: We used the original MK test for precipitation, assuming no (auto)correlation, because precipitation time series are considered less prone to autocorrelation due to the inherent variability in weather patterns (strong random variation in the daily time series). For discharge, we applied a modified MK test that accounts for autocorrelation to ensure the statistical robustness of trend detection (autocorrelation can impact trend detection accuracy). While it is true that annual maximum discharge values are typically considered uncorrelated by construction, low (first-order) autocorrelation might be present. In any case, the difference between the original and the modified MK for the discharge time series appears to be minor. We will reevaluate the necessity of employing a modified MK test for discharge.

6. **L221: What do you consider in the analyses?**

Answer: We are not entirely sure about the specific clarification the reviewer is seeking with this comment. Trends are considered statistically significant at $\alpha=0.2$. The criteria used to categorize trend's consistency are described in lines 228-233. These criteria are based on the statistical significance (percentage of time *t* for which trends are statistically significant) and their directions (number of the detected statistically significant trends that are in the same direction, i.e. increasing or decreasing) mentioned in line 221. For example, to have a consistent trend we need 25-45% of all the calculated trends in the multi-temporal analyses to be statistically significant while at the same time the majority (more than 60%) of the detected significant trends should be in the same direction (increasing or decreasing).

7. **Table 2: Last column. Shouldn't it be "Reverse *relative* frequency?**

Answer: The last column in Table 2 shows the relative frequencies of the reserve situation: given that a driver has occurred, what is the relative frequency of it being followed by a Q_{max} event. It is again a relative frequency calculation (column (3) divided by column (2)). We will consider reformulating this to avoid confusion with Fig. 4 (relative frequencies but for a different scenario).

8. **L304-308: these lines were not fully clear to me.**

Answer: We define P_{MD} using the 4-day accumulated precipitation amount exceeding the 95th percentile thresholds (the percentile is calculated from all 4-day accumulated rolling sums). As was explained, for very extreme 24-hour events this amount can sometimes be exceeded and can cause at the same time both P_{MD} and P_{99}. In that case, the P_{MD} is not caused by the 4-day accumulated amount but by the one-day event. However, when we are using a longer accumulation period for P_{MD} (i.e. 5, 6, 7, 8, 9, and 10 days) the corresponding 95th percentile increases, as the moving/accumulated period is extended, and becomes much larger than the 99th percentile used for the definition of P_{99}. Thus, the separation between P_{MD} and P_{99} becomes clearer. Irrespective of the applied duration our results remain relatively stable.