Response to R1

We thank the reviewer for taking the time to comment on our manuscript a second time. Their contributions have improved the manuscript substantially. As a result of this round of recommendations, we have improved the description and application of the GEV method, and clarified additional points as requested. Please find our detailed responses below.

Upon rereading the text, we felt that the revisions from this and the previous round had, in some places, made the text repetitive and cumbersome to read. Therefore, we have additionally proofread the text again and made minor changes to improve clarity and reduce repetitiveness without altering the scientific meaning. Larger edits are marked in red in the main text, and predominantly relate to improving the description of the results in section 3.1.2 which we felt was still not clear after the previous revision.

This is re-review of the manuscript entitled ‘What is a drought-to-flood transition? Pitfalls and recommendations for defining consecutive hydrological extreme events' by Anderson et al. Authors clearly revised the manuscript and replied to my comments. I think that the robustness and depth of this study have been improved. As for this version of manuscript, I have some additional comments and concerns.

1. Lines 114-119: The use of quotation marks around these sentences feels a bit unusual and may be unnecessary. I suggest removing the quotation marks for smoother writing.

This has been adjusted.

2. Figure 1: The selection of case study catchments provides a valuable range of hydrological regimes; however, they are currently concentrated in Europe and North America. The reviewer would encourage the authors to consider including at least one catchment from Africa and one from Asia to enhance the geographic representativeness and broader generalizability of the findings. Many regions in Africa and Asia experience distinct hydrological extremes and transitions. For example, monsoon-driven floods in South and Southeast Asia, or drought-flood cycles in the Sahel and East Africa are not currently represented. Including such examples could provide a more comprehensive picture of drought-to-flood transitions across diverse hydro-climatic contexts and strengthen the applicability of the study's conclusions to global hydrological risk management.

We thank the reviewer for this comment and agree that a wider geographic range would be ideal. We did consider candidate events in Africa, Asia, and additional events in South America including in Peru and Chile. Due to challenges with data availability and reliability none of these were selected for the final analysis.

For example, we considered using data from the 2011 event in the middle-lower Yangtze river in China, which was an extremely large and impactful drought to flood transition event, however, given the restricted nature of hydrological data in the region, inclusion of this case study was not tenable. Similarly, we wanted to analyse the drought-to-flood transition in January 2017 in Peru, and the 2024 floods in Kenya, which followed an unprecedented multi-year drought from 2020 to 2023. However, we were unable to acquire sufficiently long and high-quality streamflow time series corresponding to these events.

Accessibility of daily, or higher, resolution streamflow data globally is a major challenge in hydrological research, and it is an even greater challenge when pairing streamflow data with reported case study events. Although we in no way argue that we have found perfect case studies, we have nevertheless managed to cover a wide range of geographic and hydrological conditions, including a catchment in South America, and a monsoonal regime in Australia. Given the criteria imposed, this was not a straightforward task.

Although valuable to learn from more case studies, we do not believe that adding additional case studies will significantly improve the manuscript. Primarily, as the case studies included are thought to represent a diverse set of regime types (e.g., snow dominated, monsoon type, seasonality), and not necessarily all regions of the world. By demonstrating how commonly used methods for flood and drought detection (and their transition) may succeed (or fail) across these widely different regimes, we emphasise that care is needed in their use and interpretation. Accordingly, we hesitate to include additional case studies at this point in time. Mainly as we do not think it is needed and also (as elaborated above) as it is a highly time consuming task.

We have added the following note to the text:

"Additional case studies in South America, Asia, and Africa were considered, including, but not limited to, the 2011 event in the middle-lower Yangtze river in China, the drought-to-flood transition in January 2017 in Peru, and the 2024 floods in Kenya, which followed an unprecedented multi-year drought from 2020 to 2023. However, we were unable to acquire sufficiently long and high-quality streamflow time series corresponding to these events"

It is worth noting that at the time that this research was carried out, the CAMELS India dataset had not yet been released, although this may have provided some additional options for monsoonal regimes.

3. Lines 175-183: While the authors state that threshold estimation effects are consistent within individual catchments, it should be noted that cross-catchment comparisons may still be influenced if the underlying climatological conditions are not equally represented across sites. How might this affect the robustness of inter-catchment comparisons?

We have adjusted the sentence to the following:

"Although differences in the time series length and period covered could influence inter-catchment comparison, this effect would be consistent across all threshold level approaches and methodological combinations in one catchment, so that the approaches can be fairly compared for each site."

It is important to note that many catchments and events were considered, and these were the most complete time series available. We kept the Chilean case study, for example, despite a slightly limited time series (20 years), after considering more than 20 candidates in the region, because, like the reviewer remarks (see comment 2), we agree that it is important to incorporate a wide range of hydroclimatological regimes and geographies. In the case of the Italian catchment, we considered every sub-basin of the Po River affected by the flood event, and retained the catchments which had the most available data and which were near or within the flooded zone. Alternatively, we could have included the entire Po River basin, which has a longer and more complete time series, but this would have included large areas not affected by the transition event, resulting in a dampened signal. We could have excluded this case study, but both the drought and flood represent some of the worst on record, highlighting the importance of this area of research.

We acknowledge that the case studies selected are not necessarily perfect, but given the scarcity of hydrologic data in many parts of the world, we maintain the view that our efforts to select representative case studies with sufficient data have facilitated a novel and valuable scientific contribution.

Besides, given the importance of data integrity for extreme event detection, it would be helpful if the authors could clarify whether any additional quality control procedures were applied to the hydrological data prior to analysis.

We also removed all NULL and negative values and visually inspected the time series as a sanity check. We have added the following to clarify:

“after removing any negative or null values, which would be indicative of data error”

4. Line 213: The term “POT” (Peaks Over Threshold) appears for the first time in the main text without being defined.

This has been changed to “peaks over threshold (POT).”

5. Lines 249-252: The use of the GEV distribution to model time intervals between drought and flood events is interesting, but the justification could be strengthened. It would help to explain why the GEV is theoretically appropriate for modelling such intervals, e.g., is the time interval considered an "extreme" variable in this context?

We thank the reviewer for this point.

Choosing an appropriate distribution for this exercise is challenging and important, given the influence that parametric assumptions might have on the outcomes. In our case, the variable of interest is the time interval between drought and flood events, where extremely short intervals can be viewed as extreme minima. Empirically, the GEV performed well: the Kolmogorov–Smirnov (KS) test failed to reject the null hypothesis in 91% of cases, and although the Anderson–Darling test (which gives greater weight to the tails) was less conclusive, failing to reject in 56% of cases, visual inspection of QQ plots with parametric 95% simulation envelopes focused on the lower 30% of the distribution showed that the GEV provided a reasonable fit to the very short transition times in all case studies except California.

Therefore, although we do not claim the GEV is the ‘true’ underlying law of transition time intervals, we view it as a reasonable approximation for capturing the tail behavior relevant to our study.

We have made several edits to the text to clarify what we have done and to justify the use of the GEV distribution.

Additionally, we realized an error in our original formulation of this approach: the typical application of the GEV distribution focusses on the right tail of the distribution, while here we want to focus on the left tail, because we are interested in extremely *short* transition intervals, rather than extremely long ones. We have reformulated our code to fit the GEV distribution to the negative of the observed time interval values to address this. The description of the method is now as follows:

“Second, existing literature typically defines transitions by pre-selecting the time interval between drought and flood event, which is not explicitly linked to transition probability or physical processes. Therefore, next, we consider how the probability of these preselect windows differs between catchments. (1) We begin by calculating the time interval between each drought event and the first subsequent flood event, resulting in a series of time interval values for each catchment. (2) Next, we fit a Generalized Extreme Value (GEV) model to the negative time interval series using L-moments for parameter estimation (chosen for stability with small samples). We then map results back to the original scale to obtain the cumulative distribution function (cdf). Model adequacy was assessed using QQ plots with a parametric 95% simulation envelope focused on the lower 30\% of the distribution, given our interest is in the probability of short duration transitions. Based on visual inspection, the GEV provided a reasonable fit in all case studies except California, where it likely underestimates the probability of very short intervals. (3) From the cdf for each catchment, we extract the probability of 14- and 90-day transition periods. To quantify uncertainty, we use a parametric bootstrap to simulate datasets from the fitted GEV distribution, refit using L-moments, recompute the cdf and take the 2.5\% and 97.5\% pointwise percentile intervals from 500 replicates.”

We have separated the description of the results of this analysis into an additional section to improve clarity. Small adjustments are marked in the text and the following statements have been added:

“From the previous sections, it is clear that the time window between drought and flood events impacts the detection of transitions. This is especially true since different methods can result in different drought end dates, thus potentially excluding events from being defined as transitions if the end date is shifted by a few days. These time intervals, which are typical based on previous literature, have also been selected without explicit consideration of typical conditions in a catchment, physical processes driving response, or impacts. While quantifying physical process changes and impacts is outside of the scope of this analysis, we do assess how the preselected time intervals between events relate to what might typically be expected in a catchment. We defined here the probability of time intervals between all drought events and the first subsequent flood period and related this to the preselected time windows (Figure 8.b). “

“...For instance, while 14 days may be quick from some perspectives and in some locations, a transition with a 20% chance of occurring is hardly extreme. The approach tested here offers one potential avenue for exploring extremeness in transition duration. However, it is important to acknowledge that the confidence intervals for these estimates generally overlap between methods, that the theoretical distribution applied here to estimate probabilities may not represent a good fit in all locations. Further, different parametric assumptions may bias probability estimates, especially when sample size is small.”

Correcting the method to use the negated GEV distribution results in some changes in the results presented in Figure 8.b. which have now been updated in the text directly.

6. Figure 8b: Where are the error bars in the panel?

Thank you for catching this, we have added in the missing CIs as well as a description of how these are calculated, as above.