Response to R #1

We thank the reviewer for their constructive comments and the time that they invested in reviewing the manuscript. Our responses are included in blue text below and proposed additions are included in red.

The authors aim to explore the definition and detection methods of drought-to-flood transitions (DFTs), and emphasizes the limitations of existing threshold-based approaches in detecting such consecutive extreme hydrological events. Using eight case study catchments, authors compare the detection effectiveness of three threshold methods (fixed, seasonal, and dynamic thresholds) for DFT events and verifies the results by referring to media reported disaster events. The topic is interesting and meaningful, but it also has certain limitations. My comments are outlined below.

Major comments

1. Although the title of this manuscript is mentioning "what is a drought- to-flood transition", in the definition of drought-to-flood transitions (2.3.2), I only find the definition of the transition time. There are no restrictions on the duration of droughts and floods. Long - term droughts, short - term droughts, or interrupted droughts have different effects and need clear definitions. Therefore, I suggest that the authors include additional content to address and clarify this point.

   We thank the reviewer for this comment.

   In the literature to date, “drought to flood transition” has been used as a general term and different definitions and methods have been used to define the events as well as their transition (from drought to flood). We agree that drought type, duration, and severity certainly play key roles in determining the effects of a drought or drought to flood transition period. Thus, exploring how the different definitions of these extreme events (flood and drought) affect the identification of drought-to-flood transitions, is the key motivation for the study.

   We will add the following to the limitations section:

   “In the literature to date, “drought to flood transition” has been used as a general term and different definitions and methods have been used to define the events as well as their transition (from drought to flood). The definitions of drought to flood transitions explored here do not address, in detail, how different drought types e.g. meteorological or soil moisture would fit into this framework. This is not relevant to our study because we focus on streamflow conditions. Neither do we address how different durations, intensities or severity of drought or flood would play into this dynamic relationship. Long-term droughts are likely to have a very different relationship with floods than short, or less severe, events, for example. Understanding how individual event characteristics influence transition occurrence represents a key area for further research.”

   And the following to the methods:

   “It is important to note that droughts of different durations, intensities, drivers, and intermittency can occur and that each of these events can have different effects. The intention of this manuscript is to highlight pitfalls and discuss methodological differences in the context of impactful events. For this reason, we do not distinguish between drought and flood types in the current analysis.”

2. Figure 6 shows that different threshold methods vary significantly in seasonal catchments, but the mechanism is unclear. I suggest adding how dynamic thresholds suppress or amplify seasonal anomalies.

Daily or seasonal varying thresholds are, by definition, adapted to the seasonal norm. This means that if there are pronounced low and high flow seasons, daily or seasonal varying thresholds will be lower and higher in those seasons respectively. We will add the following text:

"Using a daily varying threshold (v) for drought, for example, in a highly seasonal catchment, would result in a relatively high low-flow threshold in the high flow season. Thus, one may select a drought in the high flow season given that the flow is significantly lower than normal (high) flow value, which may not be considered to be a drought by some, as it does not necessarily imply absolute low water levels. Rather, the flow is an anomaly as compared to the normal flow for the time of the year.

Using a fixed seasonal threshold (s: one for summer and one for winter) in catchments with two pronounced low flow seasons, allows drought events to be selected in both summer and winter. In these catchments one may identify a higher number of drought-to-flood transitions given that flood occurs in both seasons. Using a fixed threshold (f: based on a percentile from the whole series) will only select drought events in the dominating low flow season, ensuring that real drought events in absolute low flow terms are identified. However, if the lowest flows occur in winter and the interest lies in summer low flows, a seasonal threshold is preferable."

As for the question raised in particular, the daily varying thresholds are adapted to the seasonal cycle in that they select events that deviate from the daily smoothed threshold all year round, and accordingly these events are commonly referred to as anomalies (rather than drought or flood). Using a seasonal fix threshold will amplify seasonally (given that there are two dominating low flow seasons), whereas a fixed (all-year) threshold will suppress seasonality by selecting drought events from the dominating low flow season.

Minor comments

1. Some cases (Chilean and Italy) have short data periods (under 20 years), which may affect the stability of threshold calculations. This needs to be clearly stated in the limitations.

   We will clearly state this in the limitations: "Short streamflow timeseries and data quality issues (such as a bias towards missing data on high flow days in the Chilean case study catchment) could affect the accuracy of the thresholds used for drought and flood definition. However, this bias will be consistent across all methods."

   Additionally, we will include the following in the methods:

   "Limited time series length can affect threshold level estimation in that flow during the available period may differ from long term conditions. However, this effect would be consistent across all threshold level approaches and methodological combinations in one catchment."

2. The "90 days window" for DFT is based on prior studies, yet its applicability across different climate zones (tropical vs. temperate) isn't discussed. It's recommended to add a sensitivity analysis.

Thank you for this comment. We completely agree. The point that 90 days (or 14 for rapid transitions) might not be appropriate is one key aspect that we attempted to reflect on in this paper. Figure 8.b. will be restructured to show which percentile these thresholds correspond to in each catchment. The methodological description will be changed to the following:

"Second, we consider how defining case specific time intervals between drought and flood events may affect the analysis. (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 GEV distribution to the time interval data series using the R package *extRemes* and use the GEV parameters to compute the cumulative distribution function (CDF) for a theoretical period of 0-730 days (2 years). Several candidate distributions were tested, and it was shown that the GEV distribution was a good fit (based on the Kolmogorov-Smirnov and Anderson-Darlings tests, for all case study catchments). This step is necessary because the events are sometimes unevenly distributed or too sparse to reliably estimate time interval probabilities from the empirical distribution directly. (3) From the CDF for each catchment, we extract the probability of 14- and 90-day transition periods."

The presentation in the discussion will be changed as follows:

"This approach also does not consider how time intervals between drought and flood are likely to differ under different hydrological conditions within the catchments. In some locations, the shift between dry and wet conditions may be rapid developing or span of several weeks and in either case be part of a normal seasonal pattern to which local populations and systems are adapted. As an alternative to the top-down selection of time intervals, we suggest that these should be defined based either on impact-specific needs, when possible (e.g. the minimum time frame for changing a reservoir management protocol), or be defined relative to what is normal conditions if this is not feasible.

Here, we tested an alternative approach in which the time intervals between all drought events and the first subsequent flood period were defined probabilistically (Figure 8.b.). The results indicate that, depending on the methodological approach and the flow regime, the probability of a transition can vary widely. For example, in the Norwegian case, the probability of a 14-day time interval between the end of drought and start of flood ranges from 0.02% when fixed thresholds are used (f_f), to 6.5% when variable thresholds are used (v_s). On the other end, the probability of a transition within 14 days in the Swiss case study catchment reaches as high as 14.5% (s_f) and has greater than a 10% chance of occurrence within this time window, regardless of the approach. These results show that the few transitions identified within the selected time windows may differ substantially among the case study catchments, which point to the need for more research on how to best define robust and meaningful transition schemes for hydrological extremes."

We feel that additional sensitivity analysis would overload the manuscript without offering further clarity and that it is outside of the scope of the manuscript.

3. When identifying DFT events, three different drought threshold approaches were used. The paper mentions "calibration" to ensure these methods detect the same number of drought/flood events, but the specific calibration process isn't explained. It's recommended to add technical details or cite relevant literature.

The calibration procedure that was performed involved testing different thresholds until the same number of events per year were detected. We will rephase the description in the text to read: "Defining the threshold based on the number of events (selecting the same number of events) was done so that the different threshold level approaches could be compared more fairly. We repeated the event detection approach using a range of threshold levels until the desired number of events was selected regardless of methodology. Alternatively, the same percentile value could have been used for all methods. Threshold level methods may result in the selection of a proportion of streamflow which is not equivalent to the stated percentile (Brunner and Voigt, 2024), in part because of the imposed minimum duration for drought periods (30 days). Although this may not have given the same number of events, it would have provided the same number of days below the threshold."

4. Fig 2. Figure 2b has no label "b".

We will add the sub-panel label b.