We would like to express our gratitude to Dr Bailey Anderson for the detailed review and constructive feedback. We provide below detailed answers (in black) to the remarks made by the reviewer (in blue). Line numbers refer to those from the submitted paper.

The manuscript develops an analytical framework for drought to flood transitions, which focusses on the conditions which make these events important for decisionmakers. They address two questions: 2. how efficient is the framework at detecting hazards, and 2. How can the framework be used to inform decision-makers? They also assess different event severities lead to transitions, and assess different baseflow separation methods.

The paper is excellent, in my opinion. The method that they propose is, to my knowledge, novel, and their conclusions are interesting and valuable for hydrological science in general and for this emerging area of research in particular. It is well-written and easy to follow, and the figures are clear and useful. My primary suggestions relate to moving the comparison of baseflow separation methods to the supplement, and to tidying up the conclusions. I suggest publication with minor revisions.

Thank you for your very positive evaluation and meticulous review by sections. We acknowledge that the paper structure could benefit from moving the baseflow separation methods to the supplementary material. We will also review the conclusions to clarify the main take away messages and improve conciseness.

Introduction:

I congratulate the authors, because both the abstract and the introduction are very well written, easy to follow, and clearly highlight the value and novelty of this paper.

I would acknowledge the comparison of baseflow methods briefly in the abstract, as this is a substantial component of the analysis.

Thank you again for the positive feedback. We will explicitly add in the abstract that we conducted a comparison of baseflow methods in our methodology. We will add it after L13, where we mention the use of baseflow estimation to identify the spells.

Methods:

The methods are very clear for the most part. My main comment is that this paper would benefit greatly from a code release which demonstrates the application of the method presented.

We have released the datasets and we plan to release the code as open source as well. This should be done once we have finished exploring the code in the analysis of data at EU scale within the MedEWSa Horizon Europe project (ongoing work with expected completion October 2026).

Thank you for your suggestion. We will consider moving the comparison of baseflow methods to the supplementary materials in the revised version of the paper.

This is a pertinent remark. The use of daily streamflow smooths sub-daily variability, which can impact the detection of extreme flood peaks and rapid transitions. We acknowledge the recent publication (Anderson et al., 2025), where it is shown that on the Emme catchment in Switzerland, daily mean streamflow failed to capture an extreme flood peak, that, at hourly resolution, corresponded to the 99.99th percentile of daily maximum time series. In our study, the use of daily mean data would mainly impact the detection of short duration HFS in small and flashy catchments, particularly in the Mediterranean region, where convective rainfall can produce intense and flashy flood peaks that may be attenuated below the threshold when averaged into daily scale. This has implications for both the number of detected HFS and the characterization of rapid transitions. At hourly resolution, we would potentially detect more short duration HFS and rapid transitions from LFS to HFS. More importantly, we found that rapid transitions from HFS to LFS are predominantly preceded by antecedent low flow conditions (L500 to L504), therefore using hourly resolution, additional cases of these rapid back-and-forth sequences (LFS to HFS to LFS) could be detected, as flashy flood peaks that are smoothed out in daily aggregation would be captured, potentially increasing the frequency of these transition patterns.

On the other hand, the detection of LFS, which has timescale of weeks to several months (we consider only those with at least 7 days duration), and the characterization of medium and long duration HFS would be minimally impacted by hourly resolution. We therefore expect that our main findings, especially concerning the spatial patterns of transitions at the SPC level, would remain unchanged, while the frequency of short duration HFS and rapid transitions could be underestimated in flashy catchments. Our use of two threshold levels partly mitigates this, as the less severe threshold could capture spells (to some extent) that might be missed by the more severe threshold. We will discuss these implications in the revised paper.

The choice of these time windows is based on the disaster risk management cycle. When consecutive disaster occur before recovery from an initial event is complete, the cascading impacts are amplified (De Ruiter et al., 2020), and a society's ability to cope with subsequent disasters deteriorates (Ward et al., 2026). According to the UNDRR, disaster response and rehabilitation happen in the days and weeks following an event, while recovery extends over months to years (Risks and Disasters | UN-SPIDER Knowledge Portal, 2026). Furthermore, Emrich et al. (2022) classifies short-term recovery as lasting days to weeks, and intermediate recovery as weeks to months. Therefore, our within-a-month category (≤ 30 days) captures transitions where

a second hazard arrives during a highly vulnerable active response and short-term recovery phase. The seasonal category (31 to 90 days) corresponds to the early to intermediate recovery phase, where society's coping capacities are still compromised.

These timescales also align with operational forecasting horizons where forecast information can be acted upon. For example, the European Flood Awareness System (EFAS) provides sub-seasonal outlooks extending to 6 weeks, and White et al. (2017) link these sub-seasonal-to-seasonal timescales to disaster preparedness through the "Ready-Set-Go!" framework adopted by the Red Cross Climate Centre, where sub-seasonal forecasts trigger early warnings and resource pre-positioning, and seasonal forecasts support contingency planning. Our 30-day and 90-day thresholds thus correspond to the windows where forecast information can be translated into actionable resource allocation. Transitions exceeding 90 days are more likely to represent a return to normal conditions where initial response and intermediate recovery may be completed. While a subsequent spell remains relevant for risk management, the transition itself carries less operational significance for short-to-medium term emergency preparedness.

L207-215 I know that the Hermans-Rasson test addresses this to some extent, but have you considered how possible bimodality in event seasonality might influence this?

We thank the reviewer for this relevant question. Bimodality in event seasonality would make the calculation of a single mean transition date meaningless, as the mean would fall between the two modes. Our methodology addresses this through the combined use of two circular uniformity tests with different properties. The Rayleigh test is designed to detect unimodal departures from uniformity, more specifically, it tests whether the mean resultant vector length is significantly greater than zero (Ley and Verdebout, 2017). Therefore, it fails to reject the null hypothesis when the distribution is bimodal with opposing modes, because the resultant vector length would be close to zero. The Hermans-Rasson test, on the other hand, is sensitive to any departure from uniformity, including multimodal distributions (García-Portugués and Verdebout, 2018; Landler et al., 2019).

In our analysis, we only retain stations where both tests reject the null hypothesis of circular uniformity (p value < 0.05; L213). This dual requirement effectively filters out bimodal stations: a station with bimodal seasonality would pass the Hermans-Rasson test (which detects any non-uniformity) but fail the Rayleigh test (which requires a dominant single direction). As a result, only stations with unimodal seasonality are included in Figure 13 and the associated analysis. Stations that were tested but did not meet these criteria are marked as grey crosses in Figure 13. We will clarify this filtering logic more explicitly in the methods section of the revised manuscript, stating that only stations with unimodal distributions are analyzed.

Results:

L 313 "as expected" or just, "by definition"

Thanks for the suggestion. We will update the sentence to "As expected, the majority of spells..."

We thank the reviewer for the comment on this particular figure. Figure 9 is key to show the proportions of the different transitions. It does create many lines that intersect/overlap and this can create difficulties when reading the figure. In the revised paper, we will adapt the figure to not show the proportions of transition of the same spell type and duration category. Our first tries for this new visualization showed that this can decrease the complexity of the figure and declutter the chart.

We thank the reviewer for the suggestions. From our experience, the visualization of compound event analysis is a challenge in itself! For this specific figure, the suggestion of removing the LFS to LFS and HFS to HFS transition panels can indeed improve readability, and we will consider it in the revised version of the paper.

Thank you for the suggestions, which help us improving our figures' readability.

We will review the text font and the color choices in all figures to make sure the figures remain readable.

For Figure 13 in specific, we agree that there is not much variability in the concentration index for the stations that had a significant transition date (those indicated in the maps). In the revised paper we will simplify the figure as suggested, by removing the size information related to the concentration index. We believe this will indeed improve readability.

Discussion

In a preliminary analysis (not reported in the paper), we investigated the use of other approaches for the definition of the thresholds, such as using seasonal and daily thresholds. For the daily thresholds, we compared two approaches, which we show on the Figure 1 here below. One approach uses a 31-day centered moving average (green line), and the other, uses a rolling percentile for each day of the year by pooling ± 15 days around it (pink line). We observe that using a smoothing average creates lower thresholds than the monthly threshold that we finally adopted in the paper. The use of the daily threshold pooling ± 15 days results in a more similar pattern to the monthly variable threshold, with the middle of each month having the same value of the

monthly threshold. Our conclusions were that the monthly threshold already captures the inter-annual variability that we were aiming for when we chose this threshold. Having said, and as discussed on lines L559-561, we believe that using a daily threshold could eventually enhance the representation of the spell termination dates, notably avoiding the staircase patterns we have when using the monthly thresholds. As mentioned, this point is included in the Discussion section (L557 to L561).
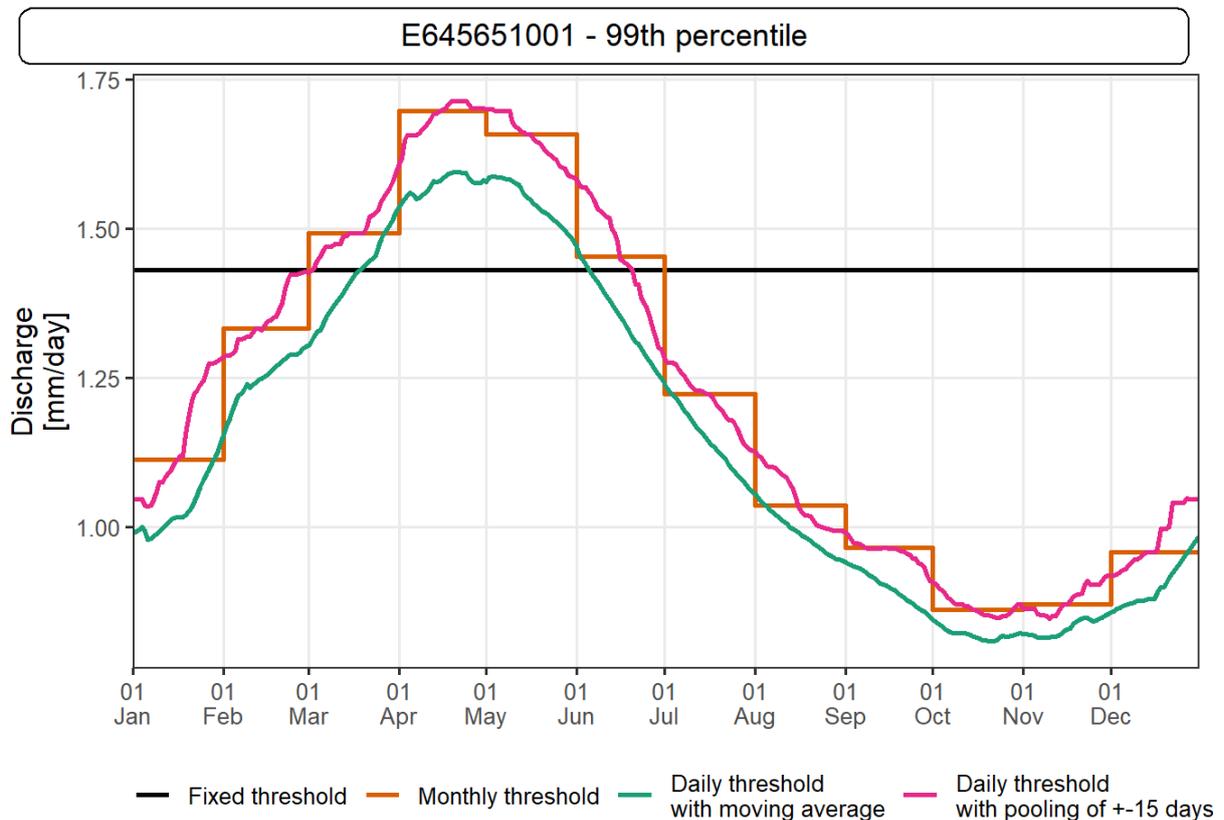

Figure 1 - Comparison of different threshold methods for the more severe high flow spell threshold (HFS 99th percentile) for station E645651001.

L 569 "our methodology allows"

Thanks for spotting the typo. We will remove the comma in L569.

Conclusions

I like that you have chosen to summarize your conclusions with bullet points, but I think that these should be refocussed to directly address the research aims outlined in the introduction.

Additionally, some of the conclusions, such as discussion of methodologies, should be moved to the discussion. In my opinion, this includes everything after line 672.

Overall, I would suggest reworking this section to be more succinct and focussed on the key contributions made by this paper.

Thank you for the feedback. We will directly address the research aims outlined in introduction L89 to L92 in the Conclusions section in the revised version of the paper, refocusing the bullet points presented. We will move the parts related to methodological discussions (lines 673 to L690) to the Discussion section. We will review the conclusions to clarify the main take away of the study, which is twofold: (i) that baseflow can be used as an effective indicator of catchment recovery to jointly detect high and low flow spells from streamflow data; and (ii) that the resulting characterization of spell transitions at SPC level in France reveals distinct regional patterns, such as rapid transitions from LFS to HFS in Mediterranean catchments, that are directly relevant in operational context.

Once again, we appreciate the comments made by the reviewer and we are grateful for it, as it allows us to improve our paper.

References:

Anderson, B. J., Muñoz-Castro, E., Tallaksen, L. M., Matano, A., Götte, J., Armitage, R., Magee, E., and Brunner, M. I.: What is a drought-to-flood transition? Pitfalls and recommendations for defining consecutive hydrological extreme events, Hydrology and Earth System Sciences, 29, 6069–6092, https://doi.org/10.5194/hess-29-6069-2025, 2025.

De Ruiter, M. C., Couasnon, A., Van Den Homberg, M. J. C., Daniell, J. E., Gill, J. C., and Ward, P. J.: Why We Can No Longer Ignore Consecutive Disasters, Earth's Future, 8, e2019EF001425, https://doi.org/10.1029/2019EF001425, 2020.

Emrich, C. T., Aksha, S. K., and Zhou, Y.: Assessing distributive inequities in FEMA's Disaster recovery assistance fund allocation, International Journal of Disaster Risk Reduction, 74, 102855, https://doi.org/10.1016/j.ijdrr.2022.102855, 2022.

García-Portugués, E. and Verdebout, T.: An overview of uniformity tests on the hypersphere, https://doi.org/10.48550/arXiv.1804.00286, 3 April 2018.

Landler, L., Ruxton, G. D., and Malkemper, E. P.: The Hermans–Rasson test as a powerful alternative to the Rayleigh test for circular statistics in biology, BMC Ecol, 19, 30, https://doi.org/10.1186/s12898-019-0246-8, 2019.

Ley, C. and Verdebout, T.: Modern directional statistics, 1st ed., Chapman & Hall CRC, New York, 190 pp., https://doi.org/10.1201/9781315119472, 2017.

Risks and Disasters | UN-SPIDER Knowledge Portal: https://www.un-spider.org/risks-and-disasters (last access: 19 March 2026).

Ward, P. J., Buijs, S. L., Ciurean, R., Claassen, J. N., Daniell, J., De Polt, K., Duncan, M., Gottardo, S., Hochrainer-Stigler, S., Šakić Trogrlić, R., Schlumberger, J., Tiggeloven, T., Torresan, S., van Maanen, N., Warren, A., Álvarez-Albelo, C. D., Banks, V., Blanz, B., Casartelli, V., Correa, J., Crummy, J., Daloz, A. S., de Ruiter, M. C., Díaz-Hernández, J. J., Díaz-Pacheco, J., Dorta Antequera, P., Ferrario, D., Geurts, D., García-González, S., Gill, J. C., Hernández-Martín, R., Jäger, W. S., López-Díez, A., Ma, L., Mysiak, J., Ngoc Nguyen, D., Padrón Fumero, N., Petrescu, E.-C., Reiter, K., Sillmann, J., Smale, L., and Stolte, T.: Reducing risk together: moving towards a more holistic approach to multi-hazard and multi-risk assessment and management, Natural Hazards and Earth System Sciences, 26, 1325–1345, https://doi.org/10.5194/nhess-26-1325-2026, 2026.

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., Vitart, F., Perez, E. C. de, Ray, A. J., Murray, V., Bharwani, S., MacLeod, D., James, R., Fleming, L., Morse, A. P., Eggen, B., Graham, R., Kjellström, E., Becker, E., Pegion, K. V., Holbrook, N. J., McEvoy, D., Depledge, M., Perkins-Kirkpatrick, S., Brown, T. J., Street, R., Jones, L., Remenyi, T. A., Hodgson-Johnston, I., Buontempo, C., Lamb, R., Meinke, H., Arheimer, B., and Zebiak, S. E.: Potential applications of subseasonal-to-seasonal (S2S) predictions, Meteorological Applications, 24, 315–325, https://doi.org/10.1002/met.1654, 2017.