We would like to express our gratitude to the Reviewer for their meticulous review and constructive feedback. We provide below detailed answers (in black) to the remarks made by the reviewer (in blue). Line, section and figure numbers refer to those from the submitted manuscript.

This manuscript uses an analytical framework to detect and characterize the spatial variability and timescales of transitions between high-flow and low-flow spells. The topic is timely and relevant for the analysis of extreme hydrometeorological events. The methods are original, the work is quantitative, and the overall quality of the manuscript is high. I believe the manuscript is a good fit for *HESS*, and I recommend it for publication after revisions.

We thank the reviewer for the positive evaluation and the detailed review provided.

I have the following comments for the authors to consider.

Major comments

1. The manuscript refers to extreme-event impact databases, but I do not fully understand the role in this study. In hydrology, flood and drought severity is commonly assessed using streamflow time series and deviations from normal conditions. While extreme events can certainly lead to economic losses and other societal impacts, these aspects do not seem to be directly addressed by this paper. I suggest clarifying the relevance of these databases to the present analysis. Otherwise, use less words about the impacts database.

   We acknowledge that our study relies only on streamflow time series and threshold-based approaches for spell detection and characterization, not using impact databases. The main reason we included this in the introduction is that compound events are, by definition, associated with impacts on the society and/or ecosystems. However, in practice, these databases are rarely available or difficult to put in place. Our methodology allows us to proceed with the compound analysis even if impact databases are not available or usable. We consider it is important to provide the overall picture and explain how our study differs by proposing an approach that uses only streamflow thresholds rather than impact data. We also note that, in our results and discussion, we cross-checked several detected spells against well-documented historical events to verify that our methodology captures events known to have caused significant impacts. This qualitative cross-examination, while not a systematic use of impact databases, illustrates the potential of our methodology to detect relevant hydrological spells and be complementary to other approaches that rely more on impacts. We also included in the conclusions (L664 to L669) that future validation against impact databases would be a valuable next step. Therefore, we believe that the literature review provided about impact databases (L59 to L65) is key to understanding the positioning of our study in the landscape of compound events analyses. We will clarify this relevance in the revised version of the manuscript, while maintaining this part of the introduction short and to the point.

2. The authors compare three different baseflow separation methods. However, based on Section 3.1 and Figure 6, the LH and UKIH methods appear to produce quite different results for the description of high-flow and low-flow spells. This raises the question of how robust the results are across methods. Please discuss more explicitly how much confidence can be placed in the findings given these methodological differences?

   We thank the reviewer for this important comment. In our spell detection procedure (L142 to L154), the spell start date is determined only by the threshold exceedance (non-exceedance) for HFS (LFS), and is therefore identical regardless of the baseflow method used. The baseflow only influences the determination of the spell end date, and even then, indirectly: once the baseflow end date ($t_{e,b}$ see Figure 1) is identified, we search backwards to find the last date when the streamflow remained above (below) the threshold ($t_{e,f}$) for HFS (LFS), which defines the final spell end date. This means that even when two baseflow methods have different baseflow end dates, they can still produce the same spell end date ($t_{e,f}$) if the backward search converges to the same threshold date crossing. In practice, the only situation where two methods produce different results is when one method's baseflow returns to normal conditions earlier, causing it to split what the other method detects as a single continuous spell into two or more separate spells.

   This mechanism explains the very high correlations observed in Figure 4 between the LH and UKIH methods (r = 0.99 for HFS, r = 0.89 to 0.92 to LFS). Since the total number of spells per catchment is a direct consequence of both start and end dates, matching spell counts between methods implies that the vast majority of individual spells have the same start and end dates. The cases shown in Figure 6, where the methods have different results, were chosen deliberately to highlight how the baseflow estimation method can impact the results, specifically in groundwater dominated catchment with long duration spells. This illustration is not representative of the typical behavior across all the catchments in our study, but it has its importance as it may be the case in other areas, if our methodology is applied to a different dataset. Therefore, we believe our results are robust in terms of the description of high-flow and low-flow spells in our dataset when using the LH method.

   The point raised by the reviewer is an important one and we will clarify it in the revised version of the manuscript. We will add it in the discussions, by also bringing to this section the part of the text about using other baseflow methods that is currently in the conclusions L673 to L677, as already suggested by Reviewer 1.

3. The characteristics of the different spells and their spatial patterns are interesting, and the authors present them thoroughly. However, I wonder whether climate change has affected these spells over the past 30 years. For example, have the duration or transition characteristics of HFS/LFS increased or decreased over time, or are there other temporal trends? Even if a full trend analysis is beyond the scope of this paper, some discussion would be valuable.

   We thank the reviewer for this comment. We agree that investigating temporal trends in spell characteristics and transitions is an interesting question. In our study, it can be, however, challenging, due to limitations in the number of years of observed streamflow in the studied catchments. In our dataset, the data period of complete hydrological years varies from 25 years to 52 years, according to the catchment. This can be short for a robust statistical analysis of trends. In addition, the relatively low frequency of transitions at the catchment scale (especially for the more severe thresholds) would also present a limit to the application of trend analysis.

   We will add a comment on this aspect in the perspectives for further studies in the revised version of the manuscript.

4. Please explain more explicitly how the two research questions are answered in the manuscript. At present, this connection is not clear?

   Our two main research questions are (L89 to L92):

   1. When we rely only on streamflow time series, how efficient is the framework in detecting hazards using a threshold-based approach combined with baseflow as an indicator for catchment recovery to identify spells of high and low flows?
   2. When applied to characterize hazards and their transitions, how can the framework be used to inform decision-makers in risk management and reduction?

   These questions are answered throughout the manuscript, as follows: question 1 is addressed in Section 3.1 (impact of baseflow separation on spell detection) and Section 3.2 (detection and characterization of spells at catchment scale), and it is further discussed in Section 4.1 (integrating baseflow estimates into spell detection); question 2 is addressed in Section 3.3 (assessment of temporally compound events), and discussed in Sections 4.2 (characterization of hydrological spells) and 4.3 (added value of the approach to understand spatial patterns of hydrological transitions).

   Based on our results, the answers to these questions can be stated as follows:

   1. The framework efficiently detects both high and low flow spells across 643 French catchments using only daily streamflow data. The mixed threshold approach captures seasonal variability while filtering out non-extreme conditions, and the integration of baseflow as a catchment recovery indicator effectively pools successive flood peaks or low flow episodes into single spells. The comparison of three baseflow methods shows that the LH and UKIH methods produce highly consistent results ($r = 0.99$ for HFS; $r = 0.89$ to 0.92 for LFS), with the LH method better suited for capturing long-duration spells in groundwater dominated catchments.

2. The framework provides operationally relevant information for decision-makers at the scale of French flood forecasting centers (SPCs). It reveals that transitions are predominantly between consecutive spells of the same type, with consecutive HFS being most common. It also identifies distinct regional patterns, such as within-a-month transitions from LFS to HFS in the Mediterranean region with median duration as fast as 8 days, while in Loire-Bretagne and Seine-Normandie region they take 17 to 22 days. The standardized frequency analysis (Figure 11) identifies which SPCs experience significantly more or fewer transitions than the national average, enabling targeted preparedness. The seasonality analysis shows when within-a-month transitions predominantly occur, providing forecasting centers with information on when to be prepared for such transitions. Together, these results offer a characterization of transition dynamics that can support the adaptation of flood and drought management strategies to local context.

In the revised manuscript, we will ensure the conclusions address both research questions more explicitly (addressing also a comment from Reviewer 1).

5. Also I also have the impression that the manuscript is somewhat long, and the discussion of baseflow separation occupies substantial space. Since the main analysis appears to rely on the Lyne-Hollick (LH) method, the authors may wish to consider moving some of the detailed comparison of baseflow separation methods to the Supplementary Material.

This suggestion was also raised by Reviewer 1, and we agree that it can be beneficial to the paper to move the detailed comparison of baseflow separation methods to the Supplementary Material. We will do it in the revised version of the manuscript.

Minor comments

1. Lines 23–28: Was the 2022 drought more severe than the 2017–2018 drought? I am not sure, but I am curious. Otherwise, it may be better to mention both drought events (2022 and 2017–2018).

In the literature, the 2022 drought was reported as more severe in terms of soil moisture anomaly, spatial extent, and total economic losses in France (Bevacqua et al., 2024; CCR, 2024). The 2018-2020 multi-year drought was, however, considered an unprecedent event in over 250 years (Rakovec et al., 2022). Bevacqua et al. (2024) acknowledge that its persistent soil moisture deficits might have contributed to the 2022 drought event.

2. The following paper on compound floods fits well with the content of the second paragraph and may be worth citing: DOI: 10.1126/sciadv.adl4005.

We thank the reviewer for this suggestion. We will include it in the revised manuscript.

3. Line 49: I do not fully agree with the statement that fewer studies investigate this topic using discharge time series than using climate data alone. Please consider rephrasing this statement.

   In our statement, we wanted to highlight the fact that studies based on climatological indices (SPI, SPEI) have emerged earlier than studies based on indexes derived from streamflow time series. We acknowledge the studies that have addressed the Standardized Streamflow Index (SSI) (see, for instance Anderson and Schilling, 2024; Ren et al., 2024; Lema et al., 2025). However, while studies using climatological indexes can be traced back to early 1990s (McKee et al., 1993), studies on SSI date back to early 2000s (Modarres, 2007). Also, from our experience, hydrologic studies have often considered floods or droughts separately in their analysis. This probably explains the fact that studies that consider hydrological transitions between these two extremes are mostly emerging only now. Hammond et al. (2025, p.2), for instance, remark that "One aspect of repeated or sequential extreme events that is only starting to be explored is the topic of drought-to-flood transitions", and similarly mentioned by Anderson et al. (2025, p.6070) with "although rapid changes between wet and dry conditions have been widely studied in the meteorological sciences, attention to this phenomenon is relatively new in hydrology [...] only a few large sample analyses of drought-to-flood transition events have been published to date".

   We will rephrase the statement in the introduction to better reflect our main point.

4. Lines 48–52: Explainable AI also offers useful approaches for analyzing this type of problem. Please consider adding a sentence from this perspective.

   We fully agree with the reviewer. Research on compound events in hydrology could greatly benefit from the use of explainable AI (XAI) as well. For example, Jiang et al. (2024) applied XAI methods to isolate and evaluate the interacting drivers of floods on a global scale, identifying that the influence of these compound effects tend to grow as flood magnitude increases. On a broader perspective, Slater et al. (2025) reviewed the opportunities of explainable AI in large-sample hydrology, highlighting its potential to discover relationships between catchment characteristics and hydrological responses that traditional models may not be able to capture. In addition, Jiang et al. (2022) further demonstrated how interpretable machine learning can be used to identify dominant flood generating processes from large sample datasets. We will include these references in the introduction in the revised version of the manuscript.

5. Lines 59–69: It is not clear why such a detailed explanation of observed impacts and impact databases is needed here. Traditionally, the severity of floods and droughts in hydrology has been assessed based on streamflow. Please consider shortening or rephrasing this paragraph.

   We agree with the reviewer that, traditionally, the severity of floods and droughts in hydrology has been assessed based on streamflow. Impact-based analysis, and impact-based forecasts as well (see, for example Potter et al., 2025, who reviewed current gaps and challenges in developing impact-based forecast and warning systems), have recently emerged as an alternative approach to better connect hazard detection with societal consequences. In the framework of compound events and multi-risk analysis, the use of impact datasets is recommended in order to capture how successive or interacting hazards amplify damages beyond what single-hazard assessments would predict. For example, Lumbroso et al. (2025) highlight the limitations of existing disaster databases, such as EM-DAT, to capture compound event dynamics, noting that current databases often have incomplete records and fail to adequately represent the complex interactions between multiple hazards. As we mentioned earlier in our replies to the reviewer, we think it is important to set this context in the introduction of the paper, which allow us also to highlight how our analysis can be complementary to impact-based compound event analyses.

6. Line 80: This is a good point and makes the paper particularly interesting.

   Thank you for the positive feedback.

7. Figure 3: Could the two subplots be shown using the same regional division? For example, one of the major regional boundaries or the 17 SPC regions.

   We appreciate the reviewer's suggestion. The two panels intentionally use different levels of spatial aggregation because they serve complementary purposes. Panel (a) shows the spatial distribution of the 643 catchments outlets used in our manuscript, colored by the six major hydro-administrative regions, which provides the reader with an overview of the station coverage and density. Panel (b) introduces the 17 SPCs (flood forecasting centers), which are the primary spatial units for the regional analysis presented in Sections 3.3 and 4 (Figures 10, 11 and 12). We believe that combining both levels of information into a single panel would result in an overly crowded figure or in omitting information that the reader needs to interpret the results. We therefore prefer to keep the current layout.

8. Figures 7 and 8: The current colors do not show the spatial patterns very clearly. Please consider using a more divergent color scale.

   We thank the reviewer for the suggestion. Since the mapped values are strictly positive (frequencies in Figure 7 and cumulative volumes in Figure 8), the implementation of a divergent color scale, without having a meaningful midpoint, would be confusing to they eye, producing an artificial discontinuity. To improve the readability of the maps in the revised version of the manuscript, we will search to adjust some of the color class breaks to better match the data distribution across panels, while maintaining a consistent legend across panels to allow cross-panel comparison.

9. Lines 530–542: These findings are not particularly surprising given the climatology of the study area, with high flows in winter and low flows in summer. A more interesting question would be whether the timing of these transitions has changed over the years, for example becoming earlier or later as a result of climate change. However, I understand that this may be beyond the scope of the present manuscript, especially given its current length.

   The result on the seasonality of consecutive HFS (winter) and consecutive LFS (summer) transitions is indeed expected given the climatology of France. This is rather ensuring as it validates the methodology developed for the detection of spells and their transitions. It thus serves as a consistency check for our methodology. More importantly, we note that the seasonality of transitions between different spell types is less intuitive, and our results reveal some informative spatial patterns: the October to December timing of LFS to HFS transitions agrees with the broader European patterns (Brunner et al., 2025), while differing from findings in the US (Götte and Brunner, 2024), as discussed in Section 4.3.

   Regarding the question of whether transition timing has shifted over the years, as we mentioned in our response to the reviewer's major comment no. 3, although trend analysis would be an interesting evaluation to be carried out, it remains out of scope of this study and topic for future studies, as we will mention in the conclusions section in the revised version of the manuscript.

10. Line 594: This statement is valid based on the historical data, but how might it change under a changing climate? Please consider adding a brief discussion.

   We agree with the reviewer that it might change under climate change. A study of spell detection and analysis on a dataset of hydro-climatic projections under climate change would be interesting, but remains out of scope of this manuscript, as mentioned earlier in our replies.

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.

Anderson, E. S. and Schilling, K. E.: Expanding the applications of the standardized streamflow index through regionalization, JAWRA Journal of the American Water Resources Association, 60, 837–850, https://doi.org/10.1111/1752-1688.13205, 2024.

Bevacqua, E., Rakovec, O., Schumacher, D. L., Kumar, R., Thober, S., Samaniego, L., Seneviratne, S. I., and Zscheischler, J.: Direct and lagged climate change effects intensified the 2022 European drought, Nat. Geosci., 1–8, https://doi.org/10.1038/s41561-024-01559-2, 2024.

Brunner, M. I., Anderson, B., and Muñoz-Castro, E.: Meteorological and hydrological dry-to-wet transition events are only weakly related over European catchments, Environ. Res. Lett., 20, 084013, https://doi.org/10.1088/1748-9326/ade72c, 2025.

CCR: Les Catastrophes Naturelles en France - Bilan 1982-2023, Caisse Centrale de Réassurance (CCR), Paris, France, https://www.ccr.fr/wp-content/uploads/2026/03/20240605_BILAN_CAT_NAT_DIGITAL_05.06.2024_compressed-1_AvertL.pdf (last access: 18 March 2026), 2024.

Götte, J. and Brunner, M. I.: Hydrological Drought-To-Flood Transitions Across Different Hydroclimates in the United States, Water Resources Research, 60, e2023WR036504, https://doi.org/10.1029/2023WR036504, 2024.

Hammond, J., Anderson, B., Simeone, C., Brunner, M., Muñoz-Castro, E., Archfield, S., Magee, E., and Armitage, R.: Hydrological Whiplash: Highlighting the Need for Better Understanding and Quantification of Sub-Seasonal Hydrological Extreme Transitions, Hydrological Processes, 39, e70113, https://doi.org/10.1002/hyp.70113, 2025.

Jiang, S., Bevacqua, E., and Zscheischler, J.: River flooding mechanisms and their changes in Europe revealed by explainable machine learning, Hydrology and Earth System Sciences, 26, 6339–6359, https://doi.org/10.5194/hess-26-6339-2022, 2022.

Jiang, S., Tarasova, L., Yu, G., and Zscheischler, J.: Compounding effects in flood drivers challenge estimates of extreme river floods, Science Advances, 10, eadl4005, https://doi.org/10.1126/sciadv.adl4005, 2024.

Lema, F., Mendoza, P. A., Vásquez, N. A., Mizukami, N., Zambrano-Bigiarini, M., and Vargas, X.: Technical note: What does the Standardized Streamflow Index actually reflect? Insights and implications for hydrological drought analysis, Hydrology and Earth System Sciences, 29, 1981–2002, https://doi.org/10.5194/hess-29-1981-2025, 2025.

Lumbroso, D., White, C. J., Brown, E., and Kolusu, S. R.: Rethinking Impact-based Forecasts and Warnings (IbFW) for multi-risks, npj Nat. Hazards, 2, 105, https://doi.org/10.1038/s44304-025-00157-5, 2025.

McKee, T. B., Doesken, N. J., Kleist, J., and others: The relationship of drought frequency and duration to time scales, in: Proceedings of the 8th Conference on Applied Climatology, 179–183, 1993.

Modarres, R.: Streamflow drought time series forecasting, Stoch Environ Res Ris Assess, 21, 223–233, https://doi.org/10.1007/s00477-006-0058-1, 2007.

Potter, S. H., Kox, T., Mills, B., Taylor, A., Robbins, J., Cerrudo, C., Wyatt, F., Harrison, S., Golding, B., Lang, W., Harris, A. J. L., Kaltenberger, R., Kienberger, S., Brooks, H., and Tupper, A.: Research gaps and challenges for impact-based forecasts and warnings: Results of international workshops for High Impact Weather in 2022, International Journal of Disaster Risk Reduction, 118, 105234, https://doi.org/10.1016/j.ijdrr.2025.105234, 2025.

Rakovec, O., Samaniego, L., Hari, V., Markonis, Y., Moravec, V., Thober, S., Hanel, M., and Kumar, R.: The 2018–2020 Multi-Year Drought Sets a New Benchmark in Europe, Earth's Future, 10, e2021EF002394, https://doi.org/10.1029/2021EF002394, 2022.

Ren, M., Jiang, S., Ren, L., Weng, B., Wang, M., Cui, H., and Xu, C.-Y.: A new non-stationary standardised streamflow index using the climate indices and the optimal anthropogenic indices as covariates in the Wei River Basin, China, Journal of Hydrology: Regional Studies, 51, 101649, https://doi.org/10.1016/j.ejrh.2023.101649, 2024.

Slater, L., Blougouras, G., Deng, L., Deng, Q., Ford, E., Hoek van Dijke, A., Huang, F., Jiang, S., Liu, Y., Moulds, S., Schepen, A., Yin, J., and Zhang, B.: Challenges and opportunities of ML and explainable AI in large-sample hydrology, Philos Trans A Math Phys Eng Sci, 383, 20240287, https://doi.org/10.1098/rsta.2024.0287, 2025.