We would like to thank both Reviewers for their careful review and constructive feedback of the manuscript. For clarity, we present here the Reviewer's comments in black and our responses in green.

Best regards,

Gabriel Bastien-Beaudet, on behalf of all authors.

Referee #1

Major Comments

1. Role of Reconstructed Streamflow Uncertainty

The study relies entirely on reconstructed streamflow products derived from hydrological modelling and data processing. Because spatial variability is a central conclusion, it is important to clarify to what extent the detected heterogeneity could be influenced by reconstruction uncertainty, model structure, or interpolation effects.

I do not suggest re-analysis, but the Discussion should explicitly address: Whether reconstruction errors may artificially enhance spatial variability. Adding such discussion would significantly improve interpretability and credibility.

The possibility that reconstruction errors may artificially enhance spatial variability cannot be completely excluded, as reconstruction uncertainty arises from multiple factors such has hydrological model structure and spatial interpolation method. However, several aspects of the reconstruction dataset suggest its uncertainty is unlikely to have a strong effect on the observed spatial heterogeneity. First, streamflow estimates were post-processed by assimilating observed data from 279 stream gauges, which greatly reduced reconstruction errors compared to using model outputs alone. In a case study involving 75 hydrometric stations across southern Quebec, optimal interpolation outperformed other streamflow reconstruction methods relying solely on observations or model outputs (Lachance-Cloutier et al., 2017). Second, only stream reaches draining more than 100 km² were retained because uncertainty is known to increase in smaller catchments. Next, the main results of this study were consistent across multiple accumulation periods and drought thresholds, indicating spatial heterogeneity is robust to different drought definitions. As proposed by the reviewer on point #2 below, we will expand on the sensitivity of our findings to the drought definition in the Results section. Finally, interpolation methods combining observations to a background field according to their error covariance structure tends to smooth out random errors rather than creating artificial heterogeneity (Asch et al., 2016). Therefore, while reconstruction uncertainty likely contributes some noise to estimates of drought spatial coverage, it is unlikely to fully explain the spatial heterogeneity and relationships documented in this study.

While this was briefly stated in the Limitations section (4.4) of the manuscript, we will expand the discussion as proposed with the information above.

Asch, M., Bocquet, M., & Nodet, M. (2016). *Data assimilation: methods, algorithms, and applications*. Society for Industrial and Applied Mathematics.

Lachance-Cloutier, S., Turcotte, R., & Cyr, J. F. (2017). Combining streamflow observations and hydrologic simulations for the retrospective estimation of daily streamflow for ungauged rivers in southern Quebec (Canada). *Journal of Hydrology*, *550*, 294-306.

2. Sensitivity to Drought Definition (SSI Threshold)

The manuscript already evaluates multiple SSI thresholds, showing that relationships remain generally consistent across definitions. However, because spatial coherence is inherently sensitive to drought definition, this sensitivity deserves stronger visibility in the main text rather than primarily in appendices.

I recommend briefly summarizing key sensitivity outcomes in the Results section to clarify how robust the central conclusions are across drought thresholds.

We agree with the proposition and will be adding a brief summary of sensitivity to drought definition in the Results section.

In section 3.1, we will add the following text: As the accumulation period lengthens from 1 to 6 months (SSI-1, SSI-3, and SSI-6), drought characteristics varied according to its definition. Longer accumulation periods smooth out short-term fluctuations, leading to fewer distinct drought events. Across all seasons, the median number of droughts per catchment drops by more than half, going from 37 events at SSI-1 to 27 at SSI-3, and 16 at SSI-6. As short-term recoveries were filtered out, individual drought events persisted longer. The median duration increased from 4 months (SSI-1) to 7 months (SSI-3) and 11 months (SSI-6). Accumulated severity increases substantially with longer timescales. The median accumulated severity intensifies from -4.5 (SSI-1) to -6.6 (SSI-3) and -10.3 (SSI-6). Longer accumulation periods leaded to more widespread spatial coverage across the stream network. The median spatial extent increases from 59% of stream length at SSI-1 to 68% at SSI-3 and 80% at SSI-6 (appendix A, table A1). However, the relative variability patterns between catchments remain highly consistent across accumulation periods (Appendix A, table A1).

In section 3.2, we will add the following text: Altering the severity threshold of drought produced the same structural shift in relative frequencies of cumulative spatial extent, regardless of the timescale. With a moderate Threshold (SSI < -1), both SSI-1 and SSI-6 display a dominant peak of widespread events. With a severe Threshold (SSI < -1.5), the high-spatial-extent peak drops, while localized events (near 0.0–0.1 spatial extent) begin to rise. With an extreme threshold (SSI < -2), both timescales transition into a bimodal distribution where widespread events drop, and localized events slowly become the primary occurrence (Appendix A).

In section 3.3, we will add the following text: Despite the sensitivity in spatial coverage to the drought threshold (section 3.2), the statistical conclusions remained robust across definitions. Linear mixed models confirmed a consistent, statistically significant positive relationship between spatial extent and drought severity across all thresholds ($p < 0.001$). As thresholds grew more negative, model intercepts increased (rising from 0.522 for SSI < -1 to 1.825 for SSI < -2) and regression slopes decreased (from 1.229 to 0.523), demonstrating that while strict definitions naturally elevate baseline severity, the positive coupling between spatial extent and severity holds regardless of the threshold used. The relationship also held with different accumulation period, with expected changes in model's coefficient. Increasing the accumulation period increased the intercepts (from 0.919 for SSI-1 to 1.697 for SSI-6) while the slopes remained similar, reflecting the aggregation hydrological deficits over a larger time window (appendix C).

3. Strengthening Mechanistic Interpretation

The Discussion proposes plausible drivers of spatial drought heterogeneity, including climate variability and local catchment processes. However, these explanations remain largely conceptual.

Without requiring additional modelling, the manuscript could be strengthened by: more explicitly linking observed spatial patterns to known hydrological processes in cold, humid environments (e.g., snowmelt dynamics, storage buffering, groundwater contributions), or clarifying which mechanisms are supported by evidence versus proposed as hypotheses.

This would enhance the broader scientific contribution beyond empirical characterization.

We appreciate this suggestion and agree with the reviewer that identifying the processes controlling within-catchment drought variability is an important research focus. However, the goal of this study was mainly to describe and quantify within-catchment hydrological drought variability across a spatially extensive stream network, a subject poorly documented in the current literature, rather than attribute this variability to specific hydrological processes happening at a sub-catchment scale. The study design did not include observations of hydrological processes (e.g., groundwater dynamics, snowpack, or distributed storage measurements) at the same sub-catchment scale as streamflow, hence why we avoided drawing mechanistic conclusions to our findings.

Consequently, we propose to modify the discussion accordingly, by explicitly stating that processes such as snowmelt variability, groundwater contributions, and catchment storage are proposed as plausible hypotheses that should be tested in future studies, rather than mechanisms demonstrated by the present analysis. Overall, we consider our study to be an important first step. By establishing that substantial within-catchment variability exists, it provides the foundation for future studies aimed at identifying the processes responsible for these contrasting drought responses.

Minor Comments

We thank the reviewer for the minor comments. We think they are reasonable suggestions and concerns and will be addressed in the revised manuscript.

- Please clarify the rationale for selecting SSI over alternative hydrological drought indicators.

We selected the SSI over alternative hydrological drought indicators for its ability to enable direct spatial comparability across hydrologically distinct regions without scale bias. Because the index values are standardized, an SSI of -1.5 represents exactly the same probabilistic rarity in all catchments, regardless of their hydrological context. This feature made the SSI well-suited for this study, which aimed to compare drought variability across catchments over a broad spatial domain.

- Provide a short justification for using drought severity (rather than duration) in the mixed-effects modelling framework.

Section 2.5, lines 174-176 mentions "Given the strong correlation between duration and severity (R^2 = 0.87), statistical modelling focused exclusively on severity as it combines duration and intensity together into a single indicator.". We will however clarify the distinction between the different drought characteristics as proposed by Reviewer #2 with the specific comment #3 below, which should refine the justification.

- Discuss more explicitly the transferability of conclusions beyond cold humid regions.

We will expand the Discussion to emphasize that the applicability of our findings beyond cold, humid regions remains uncertain, as the hydrological processes driving within-catchment spatial variability are likely to differ among hydroclimatic settings. We will highlight the need for future studies to assess spatial variability of hydrological drought across contrasting climatic contexts to determine the generality of the patterns observed in our study, thereby avoiding interpretations beyond what our results can directly support• Statements regarding implications for monitoring network design may be slightly overstated and should be framed with appropriate caution and context.

We agree that the implications for monitoring network design could be presented more cautiously. We will revise the wording throughout the manuscript to avoid overstatement and ensure that these implications are framed appropriately within the scope of our results.

Referee #2

General comments

1. Most importantly, spatial coherence is defined in terms of simultaneous drought occurrence, whereas the principal spatial metric represents the cumulative proportion of the stream network affected at any point during a grouped event.

And

The manuscript would benefit from a clearer distinction between cumulative spatial extent and simultaneous spatial coherence, together with appropriate revision of the analyses and interpretation.

While spatial coherence is often defined as the simultaneous occurrence of drought, we chose to use cumulative spatial extent to measures how much of the stream network is affected during a catchment-scale drought instead. However, we agree with the reviewer that those terms deserve a clearer distinction. We will revise the wording throughout the manuscript and refer to *cumulative spatial extent* rather than spatial coherence and focus the discussion and interpretations on cumulative spatial coverage.

2. The event-grouping procedure and statistical model formulation also require clarification, particularly because they may influence the reported relationships among drought extent, duration, and severity.

As in all drought studies, the identification of events requires an operational definition. The lack of an established operational definition for catchment-scale drought assessment highlights the novelty of our study, but it also required us to propose a definition appropriate to our study objectives. Accordingly, catchment-scale events were defined by grouping reach-scale droughts that overlapped by at least one month. We chose this definition to characterize droughts that were affecting multiple parts of the stream network as a result of a same hydrometeorological event or chain of such events. Alternative grouping could have certainly be chosen, but we believe this was the best definition for our study purpose. Additionally, using a consistent rule across all catchments allowed robust comparison of spatial drought characteristics over the study area.

While the event-grouping procedure determines which reach-scale droughts are considered part of the same catchment-scale event, drought severity and spatial extent quantify different properties of those events. Severity represents the median cumulative SSI deficit across affected reaches, whereas spatial extent reflects the cumulative proportion of the hydrometric network experiencing drought during the event. The observed relationship therefore is not imposed by the event definition itself but reflects the tendency for more spatially extensive droughts to also exhibit greater cumulative deficits.

We will expand the Methods section to clarify the rationale and procedure used to group events, to ensure the grouping procedure and statistical model are clear.

3. The Discussion would benefit from more process-based interpretation of the results. In particular, please explain the hydrological mechanisms that may cause droughts to remain localized or become widespread within a catchment, and how factors such as climate variability, tributary differences, and flow regulation may influence the observed patterns.

Please refer to response to major comment #3 of referee #1. We agree this would be a helpful contribution to the manuscript, but we believe our study design do not allow for such interpretation. We therefore propose a modification to the discussion to mention some plausible hypotheses that should be tested in future studies.

Specific comments

1. Abstract, lines 15-20: The reported values of 37% widespread and 14% localized events apply specifically to SSI < −1.5 and change considerably with the selected threshold. Please state this threshold explicitly in the abstract and highlights. Also, is there any information about how these missed drought events vary as a function of catchment size that could be included in the abstract? The abstract also states that approximately 30% of drought events would be missed by a downstream gauge, whereas the Results report 37% when all events are considered; the 30% value applies only after single-reach events are excluded. Please clarify this distinction.

   This will be clarified.

2. Abstract should also be more explicit about the spatial domain of the study and should clearly state that the analysis is based on a modeled streamflow dataset rather than observations at all reaches.

   We agree and will modify accordingly.

3. Abstract, lines 20-21: The finding that widespread droughts are also more severe is somewhat surprising, particularly because severity may be interpreted by some readers as intensity. Please clarify the distinction between drought severity and intensity and ensure that this result is adequately discussed in relation to the existing drought literature.

   While the definitions of drought severity and intensity presented in the manuscript follow standard practices associated with the Standardized Streamflow Index (SSI), we acknowledge that the presentation could be clearer in the abstract and for a broader readership. In the revised manuscript, we will refine these definitions and provide additional context from the existing literature to ensure clarity, in the abstract but also in the main text.

4. Introduction, lines 59-60, and Sections 2.4, lines 163-171: The Introduction defines spatial coherence as the simultaneous occurrence of drought across locations. However, the calculated spatial-extent metric includes all reaches affected at any point during a grouped event. It therefore represents cumulative spatial coverage or event footprint rather than simultaneous spatial coherence. Please distinguish these concepts throughout the manuscript. Either use "cumulative spatial extent" consistently or provide a time-resolved metric, such as the mean or maximum proportion of the network simultaneously experiencing drought during each month.

   Please refer to response to general comment #1.

5. The Introduction would also benefit from more discussion of existing work on the spatial extent of other types of drought and its relationship with intensity and/or severity. This would help establish the study's contribution and provide context for interpreting the reported relationship between widespread drought and greater severity.

   While we agree with the reviewer, existing work the spatial extent of hydrological droughts is currently limited, and we included as much relevant literature as possible. We will expand the Introduction to also include relevant work on the spatial extent of meteorological droughts, providing a broader context for our study.

6. Section 2.3, lines 145-155: The best-fitting distribution is selected independently for each reach. Please justify this choice, report how often each distribution was selected, and assess whether it creates spatial inconsistencies in SSI classification. Maybe a map of which SSI distribution was selected per reach, to check whether distribution selection is spatially structured or not.

The SSI is strongly influenced by the fitted probability distribution function (Stagge et al., 2015). Vicente-Serrano et al. (2012) concluded that using any single distribution was inappropriate in regions with high variability in the magnitude of streamflow and in seasonal regimes. Choosing a best-fitting distribution for each reach and calendar month ensure the SSI values represent the local and temporal conditions accurately which allows them to be compared effectively across regions, seasons and regimes. It also prevents unrealistically extreme values of SSI that are produced when an inadequate distribution is fitted to the monthly series of streamflow. We will add a table reporting how often each distribution was selected and comment it accordingly in the Results section. While we appreciate the suggestion, we believe it is not necessary to assess whether the choice of distribution creates spatial inconsistencies as this approach has proved to be an accurate way to compute the SSI (Vicente-Serrano et al., 2012). Additionally, our findings remained robust to other methodological choices such as the aggregation period for the computation of SSI and the choice of drought threshold. In the context of drought assessment, we believe the consequences of a badly fitted distribution (e.g. distortion of distribution's tails, loss of normality and probabilistic meaning of indices, seasonal bias) are worth the risk of small spatial inconsistencies from a multi-distribution approach.

Stagge, J. H., Kohn, I., Tallaksen, L. M., & Stahl, K. (2015). Modeling drought impact occurrence based on meteorological drought indices in Europe. Journal of Hydrology, 530, 37-50.

Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., & Morán-Tejeda, E. (2012). Accurate computation of a streamflow drought index. *Journal of Hydrologic Engineering*, *17*(2), 318-332.

7. Section 2.4, lines 163-171: Reach-level drought events are grouped whenever they overlap by at least one month. This procedure may create transitive event chains: event A may overlap event B, and event B may overlap event C, although events A and C never occur simultaneously. Such chaining may increase the estimated catchment-scale duration, severity, and cumulative extent. Please describe the grouping algorithm more clearly, quantify how frequently event chaining occurs, and assess sensitivity to a more restrictive event definition.

   The chaining of events described by the reviewer is accurate and is a result of the way we defined a catchment-scale event. This chaining stays in line with our definition of spatial extent, which is the measure how much of the stream network is affected during a catchment-scale drought, a definition we will clarify in the manuscript. A hydrological drought caused by a specific meteorological event for example, could cause the chaining example provided by the reviewer, that is, a local precipitation deficit induces a drought in reach A that propagates in reach B and resolves itself before the deficit continues its propagation in reach C. Thus, we believe that such chaining is not artificially inflating drought characteristics but adequately describing them. Using a more restrictive grouping algorithm would certainly decrease duration, severity and spatial extent of events but would correspond to a different definition of drought. While we appreciate the suggestion, we believe it would not enhance the manuscript to add such analysis.

8. Section 2.5: Although the section is titled "Statistical modelling of spatial coherence," the model uses drought severity as the response variable and spatial extent as a predictor. It therefore evaluates whether spatially extensive droughts are more severe rather than directly testing whether severe droughts are more spatially coherent. Please clarify and justify the direction of inference.

   This will be addressed, as mentioned in the response to general comment #1.

9. I recommend adding at least one spatial figure: examples of catchment maps illustrating a widespread and a localized drought event, to make the central finding visually concrete and to allow assessment of whether localized events cluster within sub-basins or appear spatially incoherent.

   We agree such a figure would improve the manuscript. It will be added in the revised manuscript.

Technical corrections

We thank the reviewer for the careful review of the manuscript. We will address the technical corrections accordingly.

1. Abstract Lines 9-10: This is a big claim, is it supported in the introduction? It's hard to say "studies don't do x" without a very clear overview of those studies. Consider alternative framing.
2. Lines 12-13: Change "one of the first large-scale assessment" to "one of the first large-scale assessments."
3. Line 13: Change "streamflow timeseries" to "streamflow time series."
4. Line 25: Delete the stray number "32" after "drought."
5. Lines 141-147: Use consistent terminology. Lines 141-142 refer to the monthly streamflow time series of each "catchment," whereas lines 146-147 refer to each "stream reach." Replace "catchment" with "stream reach" if this reflects the analysis.
6. Line 155: Reconcile the publication year of Gudmundsson and Stagge. The text cites 2014, whereas the reference list gives 2016.
7. Line 161: Change "McKee et al., (1993)" to "McKee et al. (1993)."
8. Line 197: Change "the number of stream gauges ... require to detect" to "the number of stream gauges ... required to detect."
9. Line 216: Change "experienced droughts conditions" to "experienced drought conditions."
10. Line 226: Delete the duplicated period after "Table 2."
11. Lines 241-242: Change "reaches not experienced drought conditions" to "reaches not experiencing drought conditions."
12. Line 288: Replace the reference to "Fig. 3" with "Fig. 5."
13. Line 368: Correct "corelate" to "correlate."
14. Line 421: Change "contains incorporates uncertainty" to either "contains uncertainty" or "incorporates uncertainty."
15. Table A1 caption: Change "fall = September/October/December" to "fall =September/October/November."
16. Use "time series" rather than "timeseries" throughout the manuscript.
17. Use consistent capitalization for "Figure," "Table," and section references throughout.