**Reply to Editors' and Reviewers' Comments**

**Legend**

Editors' and reviewers' comments

Authors' responses

Direct quotes from the revised manuscript

**The Editor:**

Thank you for your revisions of the manuscript. Both reviewers are satisfied that the manuscript is suitable for publication, but one of the reviewers has some suggested some further minor changes. I am therefore happy to accept this article for publishing subject to minor revisions in response to the reviewer comments.

Response: We sincerely thank you for your positive assessment of our revised manuscript and for considering it suitable for publication subject to minor revisions. We are grateful that both reviewers are satisfied with the manuscript. In this revised version, we have carefully addressed the remaining minor comments raised by the reviewer. Specifically, we added a brief discussion of the hydroclimatic differences between the calibration and validation periods and clarified the rationale for using NSE together with other performance measures in the context of hydrological drought analysis.

We believe these revisions have further improved the clarity and methodological justification of the manuscript. We appreciate your and the reviewers' constructive comments throughout the review process.

Kind regards

Keirnan Fowler

**Reviewer #1:**

General comments

This is the second round of reviews following major revision. I have reviewed in detail author responses to my original queries, and responses to those made by the editor and second reviewer. In general, I think the authors have done a good job in revising the manuscript to improve the clarity and methodological robustness of the study. The discussion section is much improved. However I have some small considerations for the authors to address below. These corrections can be made without another round of review.

Response: We sincerely thank the reviewer for the positive and encouraging comments. We are pleased that the revisions have improved the clarity, methodological robustness, and Discussion section of the manuscript. We have carefully addressed the remaining minor comments in the revised manuscript and provide detailed point-by-point responses below.

Original query:

Line 233: What are the differences in the hydroclimate regime in the validation period compared to the calibration period? Normally the differential split sample method you are describing here is supposed to contrast calibration and validation performance between two different climatic regimes to demonstrate out-of-sample performance for the calibration parameter set.

I see you have changed the text here, but I do think from a hydrological perspective it would be good to briefly discuss any contrasts in prevailing hydroclimate conditions between the calibration and validation time periods. This could be as simple as reporting divergence from long-term mean rainfall/flow, or just qualitatively describing conditions as “wet,” “dry,” or “average.” This does have implications for the ability of the model to generalise to other hydroclimate conditions.

Response: We thank the reviewer for this helpful suggestion. We agree that the hydroclimatic contrast between the calibration and validation periods is important for interpreting the out-of-sample evaluation. In the revised manuscript, we added a brief clarification that the mean hydroclimatic conditions during the validation period were broadly comparable to those during the calibration period, with no pronounced wet or dry anomaly. We therefore clarified that the chronological split primarily evaluates temporal out-of-sample performance under similar historical hydroclimatic conditions, rather than transferability across strongly contrasting hydroclimatic regimes.

Lines 263-266.

Mean hydroclimatic conditions during the validation period were broadly comparable to those during the calibration period, with no pronounced wet or dry anomaly (He et al., 2023). Thus, the chronological split primarily evaluates temporal out-of-sample performance under similar historical conditions rather than transferability across contrasting regimes.

Original query:

Line 267: A note here that while NSE is commonly adopted in hydrological studies, its' squared error formulation means it will place far more emphasis on high flows compared to low flow performance. This is fairly well established in hydrology literature with various other objective functions or transformations used to overcome the issue when low flow performance is important. Can you offer some commentary on why high flows are more important for your study, considering your focus on hydrological drought?

Your response to this is very relevant that much of your analysis on power generation focusses on metrics across the flow regime. I would add some of this to the methods section so the choices are justified within the paper (copy and paste is fine).

Response: We thank the reviewer for this helpful suggestion. We agree that the implication of using NSE as the calibration objective should be clarified in the Methods section, especially given its tendency to emphasize errors during high-flow periods. In the revised manuscript, we added a statement in Section 3.1.3 explaining that NSE was used to calibrate the overall inflow and release dynamics across the full flow regime, which supports the subsequent assessment of hydropower generation, and the power generation guarantee rate. We also clarified that hydrological drought conditions were evaluated separately using SSI and run-theory-based duration and severity metrics, and that the remaining sensitivity to NSE-based calibration is discussed in Section 4.5.

Lines 309-314

Although NSE tends to emphasize errors during high-flow periods because of its squared-error formulation, it was used here to calibrate overall inflow and release dynamics across the full flow regime. This choice supports the subsequent assessment of hydropower generation, and the power generation guarantee rate (Section 3.3). Hydrological drought conditions were evaluated separately using SSI and run-theory-based duration and severity metrics, and the remaining sensitivity to NSE-based calibration is discussed in Section 4.5.