Review of "Ensembling Differentiable Process-based and Data-driven Models with Meteorological Forcing Datasets to Advance Streamflow Simulation"

The authors have comprehensively answered the points raised during the previous review round, and incorporated changes that improved the manuscript. I consider that it can be published in its current state. Nevertheless, the following minor comments may contribute to further improve the final paper. The line numbers correspond to those in the file with highlighted changes.

L182: Rephrase. Suggested alternative: "Three additional modifications are included to address high-flow simulation challenges:…"

Section 2.5: I just wondered if the performance metrics from the different ensembles vary much depending on the season of the year. However, I realize that this might be out of the scope of the current study.

Fig. 5: If I understand it correctly, the plot compares single seeds (dashed translucent lines) versus the case when using 10 seeds (solid lines), while no lines show the results when averaging 3 seeds. However, in the text (L416-420) it is discussed as comparing the average 3 seeds versus the average of 10 seeds. It would be helpful to refer to in the text which are the NSE values obtained by averaging 3 seeds that are to be compared with the values shown in Fig. 5 obtained by averaging 10 seeds. Alternatively, consider adding lines to Fig. 5 corresponding to the average of 3 seeds.

L546: It seems it should be cited as "Bellmore et al., 2017": Bellmore, J.R., Duda, J.J., Craig, L.S., Greene, S.L., Torgersen, C.E., Collins, M.J., et al. (2017) Status and trends of dam removal research in the United States. *Wiley Interdisciplinary Reviews: Water*, **4**(2), e1164. Available from: https://doi.org/10.1002/wat2.1164

Fig. B1: Clarify in the caption what is the difference between the top and bottom panels.