Reply to Anonymous Referee's comments (RC2)

We thank the anonymous referee (RC2) for their constructive comments and careful evaluation of our manuscript. Their feedback has been very helpful in improving the clarity and robustness of the paper. Below, we provide a detailed, point-by-point response to all comments. The referee's remarks are shown in black, and our responses and corresponding revisions are provided in blue.

This paper investigates and compare several instances of multi-model approaches over a large set of catchments:

- **mosaic methods** that assign a single model to each catchment,
- **combination methods** that merge multiple models using static or dynamic weighting schemes.

By all means an excellent work: clear presentation, smooth writing, a pleasure to read and (why not) an example to cite.

We sincerely thank the referee for this very positive evaluation of our work. We greatly appreciate the recognition of the clarity of presentation and the overall contribution of the study.

[introduction]
If you know who first introduced the expression "multi-model mosaic", write it.

The term "multi-model mosaic" has emerged in recent hydrological literature to describe approaches that assign a single model to each catchment (or sub-catchments) based on local performance. One of the early explicit uses of this term appears in the context of the NextGen project (Ogden et al., 2021; Johnson et al., 2023; Ogden et al., 2026), and it has been adopted subsequently in other recent studies (e.g., Knoben et al., 2025; Thébault et al. 2025). We have now added these references and a brief historical note in the introduction to clarify the usage of the term.

Johnson, J. M., Fang, S., Sankarasubramanian, A., Rad, A. M., Kindl Da Cunha, L., Jennings, K. S., Clarke, K. C., Mazrooei, A., and Yeghiazarian, L. (2023). Comprehensive Analysis of the NOAA National Water Model: A Call for Heterogeneous Formulations and Diagnostic Model Selection, J. Geophys. Res.-Atmos., 128, e2023JD038534, https://doi.org/10.1029/2023JD038534

Knoben, W. J. M., Raman, A., Gründemann, G. J., Kumar, M., Pietroniro, A., Shen, C., Song, Y., Thébault, C., van Werkhoven, K., Wood, A. W., & Clark, M. P. (2025). Technical note : How many models do we need to simulate hydrologic processes across large geographical domains? Hydrology and Earth System Sciences, 29(11), 2361-2375. https://doi.org/10.5194/hess-29-2361-2025

Ogden, F., Avant, B., Bartel, R., Blodgett, D., Clark, E., Coon, E., Cosgrove, B., Cui, S., Kindl da Cunha, L., Farthing, M., Flowers, T., Frame, J., Frazier, N., Graziano, T., Gutenson, J., Johnson, D., McDaniel, R., Moulton, J., Loney, D., Peckham, S., Mattern, D., Jennings, K., Williamson, M., Savant, G., Tubbs, C., Garrett, J., Wood, A., and Johnson, J. (2021). The Next Generation Water Resources Modeling Framework: Open Source, Standards Based, Community Accessible, Model Interoperability for Large Scale Water Prediction, in: AGU Fall Meeting Abstracts, vol. 2021, AGU Fall Meeting 2021, held in New Orleans, LA, 13–17 December 2021, H43D–01, Bibcode: 2021AGUFM.H43D..01O, https://ui.adsabs.harvard.edu/abs/2021AGUFM.H43D..01O/abstract

Ogden, F. L., Jennings, K., Clark, E. P., Coon, E., Cosgrove, B., da Cunha, L. K., Farthing, M. W., Flowers, T., Frame, J. M., Frazier, N. J., Garrett, J. L., Graziano, T. M., Hughes, J. D., Johnson, J. M., McDaniel, R., Moulton, J. D., Peckham, S. D., Salas, F. R., Savant, G., … Wood, A. (2026). The NextGen Water Resources Modeling Framework : Community Innovation at the Intersection of Hydrologic, Data and Computer Sciences. JAWRA Journal of the American Water Resources Association, 62(1), e70089. https://doi.org/10.1111/1752-1688.70089

Thébault, C., Perrin, C., Legrand, S., Andréassian, V., Thirel, G., & Delaigue, O. (2025). What can be expected from a semi-distributed multi-model approach for streamflow forecasting? Tailoring the structure and size of a super-ensemble on the Rhône basin. Journal of Hydrology, 661, 133589. https://doi.org/10.1016/j.jhydrol.2025.133589

I may be old-fashioned, but I like to pay tribute to the "eminent forebears" at least in an introduction. I would suggest to cite here to initial ambition of Linsley (1982) who defended the idea of a single model arguing that it should be no longer "necessary for each hydrologist to develop his or her own model for each catchment". One of the arguments of Linsley was that "a new model for every application eliminates the opportunity for learning that comes with repeated applications of the same model." Note that your "Spatially and temporally static combination" represent to some extent a "multi-model" extension of this ambition.

We thank the referee for this historically grounded suggestion. We have now incorporated a reference to Linsley (1982) in the introduction to acknowledge the early discussion on the ambition of applying a single general model rather than developing a different model for each catchment.

Eventually, you could cite the work of van Esse et al. (2013) as an (unsuccessful) attempt in mosaic-type approaches.

We agree and have now added a reference to van Esse et al. (2013) in the introduction to acknowledge earlier applications of mosaic-type approaches in hydrological modelling.

[Materials and methods]
147 : "couple" -> "coupled"

Corrected.

[Results]
Why did you wait the results section to mention the 15 catchements that you excluded? I would have said it from the beginning.

We agree and have now moved this information to Section 2.1 (Catchments and hydrometeorological data).

[Conclusion]
Based on the surprising result you obtained with your benchmark, one of the first things I would personally try would be to test the Oudin et al's (2006) multi-calibration approach with this one-size-fits-all structure!

We thank the referee for this suggestion. In their original work, Oudin et al. (2006) demonstrated the benefit of using a seasonal index to weight hydrographs derived from two distinct calibrations (one favoring high flows and the other low flows), depending on the filling level of the production store in GR4J. Implementing such a framework in our context would require identifying an appropriate internal state variable within the selected FUSE configuration to serve as a proxy for hydrological conditions, as well as recalibrating the model separately for high- and low-flow regimes. While this represents a promising direction, it is beyond the scope of the present study.

Also, it would be extremely interesting to check whether, even if a multi-model is not much superior to a single one-size-fits-all structure, it does not prove more robust in a climate-change perspective. A possibility would be to run some kind of climate robustness test (for example the RAT of Nicolle et al., 2021).

We thank the referee for this suggestion. Assessing the relative robustness of single- and multi-model approaches under climate change conditions is indeed an important and highly relevant extension of the present work, as already highlighted in the concluding section of the manuscript. Implementing such an analysis, for example using a climate robustness framework such as the Robustness

Assessment Test (RAT) proposed by Nicolle et al. (2021), would require dedicated climate perturbation experiments and a specific evaluation protocol designed to assess transferability under non-stationary conditions. This would substantially expand the scope of the present study, which focuses on comparative performance under observed historical conditions. Nevertheless, we fully agree that robustness under changing conditions is a critical aspect of model evaluation, and we now explicitly acknowledge this perspective in the conclusion as a priority direction for future research.