

Hydrol. Earth Syst. Sci. Discuss., author comment AC1
<https://doi.org/10.5194/hess-2022-276-AC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Erin Towler et al.

Author comment on "Benchmarking high-resolution hydrologic model performance of long-term retrospective streamflow simulations in the contiguous United States" by Erin Towler et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-276-AC1>, 2022

We appreciate the comments from Referee #1, and wanted to post a **short, general response** to their main concerns before providing a comprehensive response that addresses every comment point-by-point (which we will do once the Open Discussion is closed and all comments have been received).

The reviewer has raised several constructive, general comments that we can address to increase the impact of our research paper. While this study was spurred by programmatic needs and priorities specific to our model development teams, this feedback helps us to leverage our research efforts to increase the international appeal and general interest. To this end, one of the main suggestions is to provide performance context for both models. The approach suggested in Knoben et al. (2020) is a good option, whereby the KGE values for each model are compared with KGE values calculated based on a climatological benchmark (e.g., mean annual flow or interannual daily means and/or medians). Adding this type of analysis and more of the existing benchmarking literature would address their main concern #3. Second, with this additional focus on the KGE metric and its context, we could remove some of the other metrics investigated (like NSE and logNSE), which would reduce the number of bias-based metrics (main concern #1). However, we focus on magnitude because of its relevance to our application of water availability, and bias metrics are interpretable and fit-for-purpose in this use case. In our revision, we will make these points more explicit and refer to the suggested literature on other aspects of the hydrograph that could be used for evaluation of other applications. Another issue raised (main concern #2) is the difficulty of relating a model's performance to model processes, or to concrete suggestions for improvement; however, the reviewer acknowledges that this is widely recognized as not a trivial thing to do. To this point, using the reviewer's previous suggestion, we can look more closely at the sites that perform worse than the climatological KGE benchmark. We have performed several preliminary analyses to this end, finding that most sites that perform worse than the climatological benchmark are influenced by human activities (i.e., "Non-Reference" sites). This can help shed some light on "how much" improvement we would need at these sites, for example from adding a management module, potentially providing a more concrete goal for model development. This would be of widespread interest, as the hydrologic modeling community grapples with how to account for the anthropogenic influence on watersheds, and most studies to date focus on minimally disturbed sites. This is our current thinking on how we can reshape the paper, and will provide more details/analysis in our comprehensive response after the Open Discussion closes. Finally, in thinking about the reviewer's concern #4, we could

post our codes (we have draft Jupyter notebooks and R codes), but we will continue to ponder this, for although we certainly could post them, one reason we didn't initially was because most of the metrics we calculated were straightforward and already in existing R libraries (e.g., hydroGOF), and were detailed in Table 1.

We thank the reviewer for bringing up these points and will continue to think through how to address their main concerns. We look forward to providing a more comprehensive response and an improved revision once the Open Discussion closes.