REVIEW of the paper "Improving Runoff Simulation in the Western United States with Noah-MP and VIC" by Lu Su et al.

This manuscript studies streamflow forecasts improvement in the Western U.S. using VIC and the Noah-MP model, also evident in the title. The authors describe a systematic calibration of parameters for VIC and Noah-MP resulting in model accuracy improvement. The calibrated parameters were extended to ungauged basins and the entire region using the donor-basin regionalization method. Both models showed improvement in the high and low flow simulation capabilities after calibration and regionalization. The structure and organization of the paper is coherent. The study uses suitable models and perform highly actionable simulation. Developing parameter sets regionalization across all HUC-10 basins in the WUS seems relatively novel. The topic of this work is of interest to the regional water management practitioners. The study is valuable for the regional streamflow simulation and prediction in the Western United States. It fits the scope of the journal, very relevant for HESS. The manuscript has a clear potential for publication, though there are a few aspects that need to be clarified. Based on my comments below, I recommend moderate revision before the manuscript can be published.

Major arguments

Two models used: The authors need to provide a better explanation of why they decided to use two models separately to improve runoff simulation. In fact, there are two hydrological models applied in previous studies. Two models are selected as

representatives of different levels of model complexity to see how model complexity differences impact findings, and are used to provide a reliable empirical assessment in the experiment (Shen et al., 2022). Overall, this manuscript may not explicitly provide a direct comparison of the forecast results for the VIC and Noah-MP models. Though possible explanations on VIC outperformed Noah-MP both pre- and post-calibration are given, and there is a quick mention of the regionalization enhancement greater for the Noah-MP model compared to the VIC model. The limited comparison is more like explaining the results by corresponding to the previous text. Instead of just stating that both the VIC model and the Noah-MP model are used for streamflow simulation improvement, there should be a satisfying reason to use two models. It is not clear why you study both models when they each could improve the simulation accuracy. I believe the authors should distinguish at least slightly between the two models used. Perhaps there could be some discussion between the two models, which model works better in which situation. Or, you suggest what to do with the two results to forecast.

Parameter sets: The paper concludes that gridded parameter sets were developed for both the VIC and Noah-MP models to all 4816 HUC-10 basins across the WUS after calibration and regionalization. However, the process of obtaining the parameter set seems a bit vague, with few direct mentions in the manuscript. The authors calibrated 6 parameters for VIC and 5 for Noah-MP. The next regionalization process requires basin-specific features taken into account, introducing more information from the ungauged basins. Will this result in the necessity of more parameters? Will free parameters be brought in? Perhaps the size of the parameter space could be clarified. The particular applications of the two models are also assumed to be different. Probably additional modifying parameters should be involved in the process to make the model transferable across space (Feigl et al., 2022). The gridded parameter sets could be further explained to indicate a centralized view. I think how the parameter sets are developed should be further discussed.

Best regionalization features: Selection of relevant catchment features is imperative for the success of regionalization (Bastola et al., 2008). It is not clear to me exactly how the best regionalization features are derived. The authors describe that the addition of further features doesn't improve KGE. It is not evident how you defined further features in the best regionalization features. Based on what is stated in line 360, it seems that each feature is added in a particular order. But the sequence is not specified. I think the authors should give more explanations on the applied iterative approach. In fact, relationships could be found between features. Therefore, these features could be fixed on the basis of the correlations, for example. Then the iterative process was employed by varying other features (Narbondo et al., 2020). I suggest the authors to be clearer on this point. Perhaps there could be a list indicating the importance of the features to give the rank. I would like to see more discussion here.

Minor arguments

The study only considers the KGE metric for model evaluation, which may not capture all aspects of streamflow simulation performance. The results could be supplemented by other evaluation metrics.

Check your references. Some of the references are not shown in the references part even they are put in the main text. Please complete this section to provide sufficient details so that readers can locate the source of each citation.

Section 3.1: I suppose the obtained VIC model parameters seem to be too region-specific. Perhaps indicate if the simulation can be replicated in a different area.

p9, Table 1: The first column could probably have a better layout.

p10, line 213-216: This sentence might be split to express.

p12: Perhaps move this paragraph forward, not to put up "3.2 Noah-MP parameterization" alone.

p14, line 169: It might be good to have a reference for such a statement.

p16, Figure 4: The figure name can be shown in full, adding (3), (6).

p20, line 341: Please consider regionalization performance doesn't show significant increase when using more than 4 catchment descriptors to compute the Similarity Index (Poissant et al., 2017).

p22, line 373: The references do not have the evidence that geographical similarities are most significant (Burn and Boorman, 1993). Perhaps remove "This suggests that geographical similarities are the most important factor in parameter information transfer from gauged to ungauged basins."

Section 6: The limitations of the study should be emphasized in the discussion section.

References

Bastola, S., Ishidaira, H., & Takeuchi, K. (2008). Regionalisation of hydrological model parameters under parameter uncertainty: a case study involving topmodel and basins across the globe. Journal of Hydrology, 357(3-4), 188–206. https://doi.org/10.1016/j.jhydrol.2008.05.007

Feigl, M., Thober, S., Schweppe, R., Herrnegger, M., Samaniego, L., & Schulz, K. (2022). Automatic regionalization of model parameters for hydrological models. Water Resources Research, 58(12). https://doi.org/10.1029/2022WR031966

Narbondo S, Gorgoglione A, Crisci M, Chreties C. Enhancing Physical Similarity Approach to Predict Runoff in Ungauged Watersheds in Sub-Tropical Regions. Water. 2020; 12(2):528. https://doi.org/10.3390/w12020528

Shen, H., Tolson, B. A., & Mai, J. (2022). Time to update the split-sample approach in hydrological model calibration. Water Resources Research, 58(3). https://doi.org/10.1029/2021WR031523