I appreciate the authors' explanations and their effort to add a comparison to FF networks. Apart from that, not too much has changed in the manuscript, so I don't have a lot of comments. You will not be surprised that I continue to be unhappy about the LSTM setup. To me, the framing of investigating how well you do if you don't follow the best practice is not entirely convincing. Shouldn't we rather educate the "non-expert user" about the correct modeling setup? Wouldn't that user be interested in how much better their model could be, especially in an investigation of "the lower limits of data availability"?

The small data size argument is indeed frequently heard, but it is largely a fallacy. E.g., for Australia, there exists plenty of data (Camels-AUS, Caravan, or the authors' data set) which anyone can use in conjunction with their individual catchment.

Overall, I feel that in its current form, the paper becomes one of the many many papers that compare single-basin LSTM vs. conceptual model X on gauge set Y. Nothing about it is *wrong* (in fact, this is one of the much better ones because |Y| >> 1), but it also doesn't show the full picture, and I fear that adding to this type of papers will only further distort the picture of LSTM-based modeling in the literature.