Review of the manuscript „Replication of ecologically relevant hydrological indicators following a modified covariance approach to hydrological model parameterisation“ by Visser et al.

Dear editor and authors,

I thank the authors for rigorously answering and addressing many of my comments. I highly appreciate their efforts. I only have a small number of comments left – some of them were already part of the last review round, but still need some clarification. Besides that, I would like to emphasise that I personally think that the manuscript is now of good quality and that the study proposes and tests an interesting approach to model parameterisation. It is definitively a valuable contribution for the (eco-) hydrological modelling community.

Comments

1) P11 L 24-25: The introduction now includes a part that defines the “traditional” approach. However, the definition is still relatively vague, mostly because there is a lack of examples. Is the Nash-Sutcliffe efficiency a traditional objective function? What about the Kling-Gupta efficiency? What about the flow duration curve? These questions can be easily clarified by adding the information (e.g. “…traditional objective functions, such as the NSE and KGE…”). Also, it would be interesting to know which objective function was used in the studies that are referenced right after the definition (Shresta et al., 2014; Vis et al, 2015). This would certainly help to define what a traditional objective function is.
2) P12 L10: You mention the selection of ER HIs to guide model parameterisation as one limitation you want to address in this study. The limits of acceptability concept of the covariance approach also needs a selection of ER HIs. Can you shortly comment on where the difference is?
3) P12 L11 and P28 L27-32: Is the recalibration of a model, such as GRJ, really a limitation in terms of computational power? I assume the bigger challenge is rather to find a set of parameters that manages to replicate multiple ER HIs. Having such a “common” set of parameters is hydrologically more meaningful than having a separate set of parameters for different ER HIs.
4) Table 1: Catchment steepness is missing for the Tarland Burn catchment. I assume that it could be calculated relatively easily using the catchment outline and a digital elevation model.
5) This comment is on step 2 and 3 of the covariance approach. In Fig. 1 and the corresponding text you make a clear distinction between step 2 and step 3. I am not convinced that it is meaningful to separate the two steps. To my understanding you do the following: In step 2, you make a plot of the observed and *n* simulated covariances and each ER HI. If the observed moments lie inside the “cloud” of simulated moments, then the model structure is valid. In step 3, you take a subset of the “cloud” that is within the limits of acceptability. This gives you the valid parameter sets. You therefore use a stricter test to validate parameter sets than to validate the model structure. My question is now: does that make sense? An extreme example: imagine you have 10,000 parameter sets, a validated model structure, but only 1 parameter set is within the limits of acceptability. Would you say you have a valid model structure if only 1 parameter set actually manages to give you simulations of covariance and ER HI that are realistic enough to work with? Personally, I would doubt that my model structure is valid. Based on your text and comments I understand that step 2 is exactly as proposed by Vogel and Sankrasubramanian and that step 3 is the extension. However, I think what you want to do is modifying their approach to 1) use it for multiple indicators, and 2) to weight indicators according to their importance. To do so, you can skip step 2 and directly apply step 3.

6) P28 L23: As you mention, equifinality is reaching the same outcome by different means. This is mainly due to uncertainties in data and model structure/parameters and the limited type of data used to evaluate the model. Given the same model and data, I don't fully understand why the covariance approach reduces equifinality compared to a traditional approach.
7) P30 L8: You mention that an improvement in consistency is reached by the covariance approach. Is this due to the covariance approach or due to the fact that all available ER HIs are used to select parameter sets, i.e. the ER HIs are part of the calibration?
8) Finally, I fully agree that your approach is different from the GLUE approach proposed by Beven and Binley (1992). However, they have a lot of similarity and this is why I think it is fair to shortly comment on that somewhere in the manuscript (two to three sentences are enough). I would like to add some thoughts to your previous answer to this topic: a) both, the GLUE and the covariance approach need a performance metric (you use the covariance and ER HIs; any objective function (also called likelihood function) can be used in GLUE), b) statistical importance could be considered in GLUE using the approach of fuzzy limits of acceptability (Beven, 2006), and c) a range of indicators can be considered in GLUE (e.g.Blazkova and Beven, 2009).

References

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of hydrology*, *320*(1-2), 18-36.

Blazkova, S., & Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, *45*(12).