Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/hess-2022-282-RC2, 2022

Comment on hess-2022-282

Anonymous Referee #2

Referee comment on "Knowledge-informed deep learning for hydrological model calibration: an application to Coal Creek Watershed in Colorado" by Peishi Jiang et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-282-RC2, 2022

This study aims at basin scale parameter calibration for a physical hydrologic model (ATS) using DL-based inverse method. The authors leveraged the mutual information (MI) for the global sensitivity analysis to identify the relation between parameters and model simulations, which was later applied to the input selection of a MLP parameter inverse model. They executed different groups of simulations and analyses to comprehensively evaluate the proposed framework. The MS is well-written with overall structure easy to follow. I provide my suggestions below regarding better clarifying several points and hopefully they can be useful to further improving the quality of this study.

As my understanding on this study, the title “knowledge-informed DL” is mainly represented by the MI sensitivity analysis used in the input selection for the following inverse modeling. Knowledge informed learning, generally in my mind, is applying physical laws or constraints to the data driven model based on our domain knowledge. To bridge the proposed MI and physical processes together and better strengthen the headline of this study, I suggest the authors try to link the MI results with physical processes of the study area and give some physical explanations of the results from sensitivity analysis. This can further highlight the physical representations of this study.

I am still confused at the details about how the inverse framework is set up and trained. My understanding is that you first run some simulations with ATS (how are the parameters first initialized here?) and use the simulations and parameters to train an inverse mapping with inputs selected by MI, and then replace ATS simulations with real observations to estimate parameters. Does the “responses” mentioned throughout the paper mean the simulated ATS discharge and ET? What are the training targets and how do you develop the structure, tune the hyperparameters and train the DL framework? What are the training and testing dataset separation?... Maybe I didn’t understand some parts very well, but indeed expect the authors can better clarify their methodology and results to make readers more easily understand this work.

I didn’t understand the result of Figure 7 well and hope the authors can give more explanations. Which variables are the NSE and mKGE calculated on, estimated parameters or model simulations? If they are simulation metric, are theses simulations from the model forwarding with parameters estimated from real observations (Q & MODIS ET inverse)? For each individual parameter evaluation, how do you set up the values of other parameters when doing ATS forwarding. The caption notifies the performance is reported on testing data, but I didn’t see how the authors divide testing and training data.

I am thinking this multiple-years training VS one-year training discussed in section 3.3. As for multiple years, you choose to increase the input neuron number, or keep the one-year structure not changed and just use multiple years data as more training samples? I think the latter one could be more beneficial because inputting three-year time series once to the model would require large amounts of parameters in the input layer which can be inefficient and overfitted to small training data.

Another point I would be interested in is whether the authors have tried adding meteorological forcings to the inputs of inverse modeling. I feel the forcing-hydrologic response pair is very important to inform the characteristics of basin processes reflected in model parameters. I am expecting the paired input may bring more benefits to this study.

Specific comments

Line 76 Do you intend to discuss the overfitting problem here? Large number of weights and limited realizations as training data may cause overfitting with a complicated model.

Line 177 Please also give explanations for H(Y|X) to help readers’ understanding.

Line 258 and 259 How did the authors safely draw the conclusion of “improves the MI estimations” and “the parameters are falsely considered” based on the differences of preliminary and full analysis? Additionally, is it possible that in the preliminary analysis some parameters are not identified but actually behave sensitive if you include them in the full MI analysis?

Figure 8 The inputs to the inverse model here are real observations or simulated responses?