Responses to the Comments of Reviewer #3 on ⟨hess-2022-282⟩

Peishi Jiang Pin Shuai Alexander Sun Maruti K. Mudunuru
Xingyuan Chen

December 28, 2022

> General comments:
> This paper proposes a knowledge-informed deep learning method that can reduce the computational demand required by the calibration of the computationally expensive environmental model. I like the proposed MI sensitivity analysis best because it is able to disclose the sensitivity of parameters varying along with time which traditional sensitivity analysis is not capable of. Please see the comments in the attached PDF file for suggestions and questions.

Thank you for reviewing our manuscript. We made modifications per the comments below.

> Line 13: ‘‘all observations’’ --> observations covering all time steps

Revised.

> Line 137: ‘‘continuous discharge (Q)’’ --> observed daily

Revised.

> Line 214: When this inverse mapping is trained, how to select the significant time steps from all time steps for each parameter? The union of the time steps that are significant in using Q only and using ET only?

The significant modeled responses (either Q or ET) are identified prior to the development of a knowledge-guided inverse mapping that uses these responses as the inputs. A response at a time step is considered as significant to a parameter if its corresponding mutual information is non-zero based on a statistical significance test, which is described below:

"(L190-L193)In this study, we follow a similar strategy of [2] to estimate p using 10 evenly divided bins along each dimension and perform SST tests to filter out any non-significant MI value with a significance level of 95% based on 100 bootstrap samples. In other words, the computed MI is set to zero if the statistical significance test fails."

In the case of using both Q and ET to estimate a parameter, we took those Q and ET that have non-zero mutual information and concatenated them into an array to the inputs of the knowledge-guided inverse mapping (as elaborated in Figure 3(b)).

> Lines 225-226: It's not clear what the epsilon is because there is no definition of or equation defining the noise and observation error.
>
> I assume what you mean by the observation error is the standard deviation of the observation, and epsilon is 1/3 of observation standard deviation. I am not sure whether my understanding is right.

To clarify, we modified the associated sentences as follows:

" (L226-L230) To this end, we generate 100 realizations of noisy observations, denoted as $\mathbf{o}_n$, such that $\mathbf{o}_n = \mathbf{o} + \epsilon \times \mathbf{o} \times \mathbf{r}$, where $\mathbf{o}$ is the vector of the original observations, $\mathbf{r}$ is the random vector with the same size as $\mathbf{o}$ and is drawn from a standard normal distribution, and ϵ is the standard deviation of the random vector $\mathbf{r}$ and is usually taken as 1/3 of a given observation error. Following [1], ϵ is set to 0.0166 for a 5% observation error in this study. "

> Line 273: ''redish'' --> greenish

Revised.

> Without reading Cromwell et al (2021) and Mudunuru et al (2021) about the original inverse mapping that estimates all parameters from all responses(illustrated in Figure 3(a)), I have no idea why the NSEs using the same response (for example Q) for different parameters are different. To my understanding, the test data (46 out of 396 realizations including all time steps instead of selected time steps) are the same in each case (q, et or qet) for different parameters in the original inverse mapping since all the parameters are estimated together. Hence, the NSE using the same response (for example Q) for different parameters should be the same. How are the NSEs using data from the original inverse calculated for each of the parameters in the three cases?
>
> To make the complicated figure easy to digest, you might consider to remove 7(a) in the upper pannel.

We calculated NSE and mKGE for the estimation of each parameter, instead of all the parameters. The purpose is to compare the parameter estimation by the original and the knowledge-guided mappings. We clarify this point in the following sentence:

" (L263-L265) The performances of these mappings are further evaluated on the two magnitude-independent metrics, NSE and mKGE. To have consistent comparisons between mappings with and without being knowledge guided, both metrics are computed for the estimation of each parameter based on the test dataset. "

> Could you add the default ATS run which uses the default parameter values instead of the estimated parameter values in Figure 8 and Figure 9?

In order not to complicate the two figures, we now add a figure in the Appendix (Figure A7) that compares the default and the calibrated ATS runs, showing the improvement of the model performance using the knowledge-informed inverse mapping.

> Lines 345-347: I have no idea what this sentence is for.

This sentence is used to indicate the importance of the discharge fluctuations during the low flow period of the dry year in model calibration. We revised the sentence as follows:

“ (L) Our finding on the significance of dry year discharge in model calibration indirectly supports some recent studies. [3] found that high flow provides limited information to calibrate models in snow-dominated catchments. This is mainly because there are fewer discharge fluctuations during snow melting or high flow period than rainfall-fed catchments [4]. The decreased role of high flow, in turn, enhances the importance of the low flow period in calibration, particularly in dry years. Indeed, in this watershed, we do observe stronger diurnal discharge fluctuations during the low flow period of the dry year (i.e., WY2018) than the other two wetter years (see Figure A3 in the appendix), which facilitates the better calibration result using observations from the dry year. ”

> `Figure 12: Could you please plot the observation in a different color? Maybe red?`

Figure 12 (now Figure 13) is updated now.

> `Line 388: It’s not clear to me what the number of input means. Does it mean the total time steps of the selected time steps? Please indicate the time step of observed Q and modeled Q from ATS in section 2.1 and 2.2.`

The input refers to the input neurons of a neural network. For knowledge-informed inverse mapping, the input is an array concatenating the responses (i.e., with non-zero mutual information) to be assimilated within a given calibration period. So, the number of inputs is the number of these selected responses used for parameter estimation. We revised the associated texts as below:

“ (L248-L253) Each mapping was developed using a multilayer perceptron (MLP) model as follows. The input of an MLP is an array concatenating the responses to be assimilated within a given calibration period. The output is the model parameter(s). Let’s denote the number of input neurons, output neurons, and hidden layers as N_i, N_o, and N_l, respectively. N_i depends on the type of inverse mapping (with or without being knowledge guided), the selections of the response variable(s), and the number of calibration years, varying from $\sim$100 using one year of Q to 1,785 using all three years of Q and ET. ”

> `Line 394: So the Adam optimization is used to tune the values of the number of hidden layers Nl and the learning rate within the sets listed below?`

The Adam algorithm is a stochastic gradient descent approach to optimize the parameters of a deep learning model and is used for each MLP development. The hyperparameter tuning was done through a grid search to find the optimal hyperparameters, where each trial/training employs the Adam algorithm to optimize the loss. The related sentences are revised as follows:

“ (L259-263) We trained each MLP using mean square error (MSE) as the loss function over 1,000 epochs with a batch size of 32. The Adam optimization algorithm, a stochastic gradient descent approach, was used to train the neural network. We performed hyperparameter tuning on each MLP using grid search to find the optimal result by varying the number of hidden layers $N_l = [1, 3, 5, 7, 9, 10]$ and the learning rate $l_r = [1e-5, 1e-4, 1e-3]$. ”

References

[1] E. Cromwell, P. Shuai, P. Jiang, E. T. Coon, S. L. Painter, J. D. Moulton, Y. Lin, and X. Chen. Estimating watershed subsurface permeability from stream discharge data using deep neural networks. *Frontiers in Earth Science*, 9, 2021.

[2] P. Jiang, K. Son, M. K. Mudunuru, and X. Chen. Using mutual information for global sensitivity analysis on watershed modeling. *Water Resources Research*, 58(10), 2022.

[3] S. Pool, D. Viviroli, and J. Seibert. Value of a limited number of discharge observations for improving regionalization: A large-sample study across the united states. *Water Resources Research*, 55(1):363–377, 2019.

[4] D. Viviroli and J. Seibert. Can a regionalized model parameterisation be improved with a limited number of runoff measurements? *Journal of Hydrology*, 529:49–61, 2015.