Author Response

hess-2020-487

Editor (Dr Fabrizio Fenicia):

[Comment] Reviewers are generally satisfied with the outcome of the review process, and have some minor recommendation to further improve the manuscript, which should not require much additional work. I foresee acceptance after these minor issues have been resolved.

[Response] We are pleased with your decision. We would like to express our gratitude for your time and effort dedicated to providing valuable feedback to help in improving this journal paper. It is a great honor to have you as the editor to evaluate our research work. We have addressed all the minor recommendations of the reviewers through this document and the revised manuscript.

Reviewer 3:

[Comment] This study designed a semi-distributed model named MIKE-SHA, and it shows better performance than a lumped model ML-RR-MI. This paper shows that the experiment is reasonable and the results are credible. There are some minor issues and suggestions listed below.

[Response] We are pleased with your decision. We would like to express our gratitude for your time and effort dedicated to providing valuable feedback to help in improving this journal paper.

R3.1

[Comment] Line 146. "...without depending on domain knowledge." It could be more accurate to use "...with limited dependent on domain knowledge." Many studies have shown that if you use more domain knowledge or physics-guided models, it would be much better. This means domain knowledge is still needed and may be important in deep learning as well.

[Response] Yes, we agree with your comment. Therefore, in the revised manuscript, the sentence has been corrected as follows.

"Data-driven techniques have made it possible to develop implementable models with high prediction accuracy using the available data with limited dependence on domain knowledge." – Page 5, Line 145 - 146

R3.2

[Comment] From Line 153. You mentioned data-driven models on "ungauged basins" multiple times in this paragraph. I thought you may address this with your models, but I found that you did not mention the predictions in ungauged basins in the rest of your paper. This part could be simplified in your introduction.

[Response] Following your recommendation, the part on ungauged basins has been simplified (shortened) in the revised manuscript. – Page 5, Line 152 - 154

R3.3

[Comment] Line 249. "There is no catchment size limitation for applying the ML-RR-MI toolkit." It would be more careful to say "There is no catchment size limitation in the design of the ML-RR-MI toolkit." In addition, I am interested in that what is the largest catchment you have tested using ML-RR-MI toolkit?

[Response] Our objective of extending lumped modelling capacities of the ML-RR-MI toolkit towards semi-distributed modelling through the MIKA-SHA framework is to incorporate spatial heterogeneities of catchment properties and climate variables into the model building phase. In general, it is expected that large catchments are more spatially heterogeneous than small catchments. Hence the lumped representations in ML-RR-MI for small catchments are much meaningful. On the other hand, semi-distributed representations in MIKA-SHA are more meaningful for large catchments where spatial heterogeneities of catchment properties and climate variables are expected to be significant. However, for data-limited large catchments, inducing a semi-distributed model through MIKA-SHA may not be sensible due to the comparatively large number of model parameters. In such situations, applying ML-RR-MI to induce a lumped model may be more accurate and meaningful. This was the rationale behind adding such a sentence in the manuscript.

However, following your comment, we have rephrased the sentence in the revised manuscript (Page 8 – 9, Line 246 - 247). As we have limited the utilization of ML-RR-MI to relatively smaller catchments, the presented catchment in the manuscript (Rappahannock River basin near Fredericksburg, Virginia, United States – 4134 km^2) is the largest catchment tested so far with the ML-RR-MI toolkit.

R3.4

[Comment] Line 271. "As MIKA-SHA rely on GP, ..." should be "As MIKA-SHA relies on GP, ..."

[Response] The sentence is corrected in the revised manuscript. – Page 9, Line 269

R3.5

[Comment] Line 411. "To prevent overfitting, the optimal model selection process considers performances of both calibration and validation periods." What detailed method did you use to consider both periods? In machine learning and deep learning, we normally simply used the one with the lowest validation error.

[Response] Generally, in Genetic Programming (GP), the algorithm drives its total population towards the solution. On top of that, in MIKA-SHA, a multi-objective optimization framework is utilized based on the Pareto-optimality concept. Therefore, at the end of GP optimization (training period), a set of non-dominated models in terms of the multi-objective criterion used is identified as the potential solutions (in the context of rainfall-runoff modelling, these are different models). In the MIKA-SHA framework, the performances of all these models are evaluated on the validation period using the same multi-objective criterion. Then, the Pareto-optimal models are reidentified using the validation fitness values. This way, the models that perform better only in the training period are screened out, and the optimal model is identified from the reidentified Pareto-optimal models by using the optimal model selection strategy of MIKA-SHA (2.1.3 Model Selection).

R3.6

[Comment] Line 711. "One of the major issues with machine learning models is the overfitting of the model to its training dataset. The consistent performances over the calibration, validation and testing periods of all selected optimal models through MIKA-SHA show no such issues in this case." This is not appropriate to say so since the calibration, validation, and testing periods are independent time, the model efficiency values are not meaningful to compare directly. You could address it by comparing the results of two different models at the same period. Your ML-RR-MI_SUPERFLEX has KGEs of 0.88/ 0.82/ 0.65, while MIKA-SHA_SUPERFLEX has KGEs of 0.83/ 0.82/ 0.83. Comparing these two results, we can see that ML-RR-MI has higher model efficiency at training but a much lower testing efficiency than MIKA-SHA. This is a signal that MIKE-SHA has reduced overfitting issues than the ML-RR-MI models.

[Response] Yes, we understand your comment here. Performing consistently in different time frames does not necessarily mean that models are not overfitted. Therefore, the sentences about overfitting are rephrased accordingly in the revised manuscript. – Page 30, Line 710 - 712

R3.7

[Comment] Other suggestions in further studies: For the study of distributed modelling, a larger catchment with more sub-catchments would be much more helpful.

[Response] We gladly accept your suggestion here. Indeed, a large catchment with more sub-catchments may provide more meaningful insights into the runoff dynamics of the watershed.

R3.8

[Comment] I understand that this is a machine learning modelling paper with the Induction Knowledge Augmented to make it done automatically. However, a lot of technologies in deep learning have been used to make the progress automatically, and we do not need to test them manually as you mentioned in Line 255. "...which makes it almost impossible to test them manually." I am very interested to see the results of comparing interpretable or physical-informed deep learning models in the future if you are interested in it. For example, some physical-informed deep learning models (Rao et al., 2020) predict the parameters first and then applied to physical equations to generate the final results. Deep learning models have natural advantages in adjusting model parameters and selecting models. Hope this would help. Rao, C., Sun, H., & Liu, Y. (2020). Physics-informed deep learning for incompressible laminar flows. Theoretical and Applied Mechanics Letters, 10(3), 207-212.

[Response] We highly appreciate mentioning the exciting research work of Rao et al. (2020). Indeed, we are also highly interested in comparing physics-informed deep learning findings with physics-informed genetic programming findings. In fact, this is one of our major future research interests as a research group. As you have stated here, deep learning models have their unique advantages over the other machine learning techniques in the context of physics-informed machine learning. On the other hand, incorporating theoretical knowledge with learning algorithms is more explicit with GP. Therefore, comparing both in the context of theory-guided data science would undoubtedly be very impactful.

Reviewer 4:

[Comment] I thank the authors for effectively addressing most of my comments and suggestions within short period of time especially the ones that require time demanding revisions but critical for quality of the manuscript. I am satisfied with the overall responses and thorough revisions of the manuscript.

[Response] We are pleased with your decision. We would like to express our gratitude for your time and effort dedicated to providing valuable feedback to help in improving this journal paper. Indeed, your recommendations immensely helped to improve the quality of our manuscript.