Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-487-AC1, 2020

Hydrology and Earth System Sciences Discussions

Open Access

Interactive comment on "Hydrologically Informed Machine Learning for Rainfall-Runoff Modelling: Towards Distributed Modelling" *by* Herath Mudiyanselage Viraj Vidura Herath et al.

Herath Mudiyanselage Viraj Vidura Herath et al.

e0149661@u.nus.edu

Received and published: 24 December 2020

We would like to express our gratitude towards the time and effort that the editor and all the reviewers dedicated to providing valuable feedback to help in improving this journal paper. We appreciate the insightful comments and suggestions given by the reviewers on this paper.

We are pleased that both the reviewers have accepted our proposed toolkit as a novel, valuable contribution to the field of hydrological modelling. As most of the concerns raised by the reviewers are related to the structure of the manuscript rather than the content, we believe a restructured manuscript would adequately address all comments and concerns of reviewers.

Overall, we have addressed all the concerns of the reviewers through this document. Here, we have included the point-by-point response to the reviewers' comments and concerns. The original reviewer's comment is marked by starting the line with ">", while the corresponding response is annotated with "[Response]". The line numbers in reviewers' comments refer to the original manuscript.

Reviewer 1:

GENERAL COMMENT

> There are many extremely long paragraphs (e.g. lines 70 to 90) that express multiple concepts. I would make paragraphs shorter, creating a paragraph per concept. This should help the readability of the paper.

[Response] As you have suggested we will divide the long paragraphs into shorter paragraphs in the revised manuscript (paragraph per concept).

> There are several statements without a reference justifying them.

[Response] The relevant references will be added appropriately in the revised manuscript (details are provided in the responses below).

OVERALL STRUCTURE

> I do not like the structure of the paper and the amount of content given to each paragraph. Your main message is to present MIKA-SHA but this is left to section 4, which is barely 1 page out of 42. If that was the main concept of the paper, I would give it more space.

[Response] The section on MIKA-SHA will be expanded by providing more details in the revised manuscript (details are given below).

> I like the material in sections 1 to 3 (included) but they are basically a mix of intro-

duction and a “methods” section. I would move some content from section 2 and 3 to the introduction, which can be divided in subsections. A possible structure of the introduction can be:

- Quick introduction on hydrological modelling and TGDS

- On hydrological models

* Physics-based models vs. conceptual models vs. data science models (mix of 2.1 and 2.2)

* Focusing on conceptual models, difference between fix and flexible structure (some part of 2.2)

* Lumped vs. distributed (2.3)

- On ML models

* Some generalities (3)

* ANN (3.1, maybe reduce)

* GP (3.2)

- Physics informed ML (3.3)

I would then add a “methods” part where you write about what of sections 2 and 3 you actually use: SUPERFLEX, FUSE, GP. I would also add to this the metrics that you later use in section 5.

[Response] We find the structure that you have proposed is more organised than the original order. Hence, the proposed structure will be used to organise the content in the revised manuscript.

> Section 5 and 6 are about the case study and I think this should represent a minor part of the paper (which objective is to present how MIKA-SHA works). To this end, I would more or less keep the same content and put it in a single section divided in:

- Presentation of the case study

- Settings

- Results that you get

- Meaningful discussion, potentially showing that MIKA-SHA works. I would therefore move some aspects of this section elsewhere:

- Metrics to a “methods” section

- Further explanation of MIKA-SHA functioning to the section that presents the model

- General implications on the goodness of the approach to a general “discussion and conclusion” section.

[Response] A single section on results will be added as proposed here in the revised manuscript. Further, some parts of the results section in the original paper will be moved to other sections as proposed.

PAPER CONTENT

> Keeping in mind that I do not have a deep knowledge of GP, I find quite difficult to understand what MIKA-SHA actually does and my difficulty can be motivated by the following reasons:

- Use of jargon from GP, that may be not common in the hydrological community (e.g. model induction vs. model selection)

- Assuming good familiarity of the reader with GP

- Assuming that the reader knows ML-RR-MI: you do not have to re-write here that paper but at least explain here the concepts that are necessary. I find it difficult to follow something that says that MIKA-SHA is basically ML-RR-MI plus something else.

[Response] We accept the concern you made here. Hence, in the revised manuscript, a separate paragraph about the ML-RR-MI will be added when MIKA-SHA is introduced (e.g. paragraph below). Further, the workflow of the MIKA-SHA will be explained in a more general way of avoiding jargon specific terms.

"Machine Learning Rainfall-Runoff Model Induction toolkit (ML-RR-MI) (Chadalawada et al., 2020) is a hydrologically informed lumped rainfall-runoff model induction framework based on GP. The unique feature of ML-RR-MI is that it utilizes the existing body of hydrological knowledge to govern the GP algorithm to induce physically sound and consistent models with high prediction accuracies. The building blocks of two flexible modelling frameworks, namely SUPERFLEX (Fenicia et al., 2011; Kavetski and Fenicia, 2011) and FUSE (Clark et al., 2008) are used as the elements of incorporated hydrological knowledge of ML-RR-MI. The model building components of SUPERFLEX consist of reservoir units, lag functions and constitutive functions to represent storage-discharge relationships and characteristics of lag functions (Fenicia et al., 2016). In the FUSE framework, building blocks include the selection of upper and lower layer architectures, flux equations to represent surface runoff, percolation, evaporation and the presence of interflow and flow routing. These building blocks are incorporated as special functions into the function set of ML-RR-MI along with basic mathematical functions. ML-RR-MI optimizes both model structure and model parameters simultaneously and selects an optimal model for the catchment of interest without any direct human involvement."

DATA AND CODE AVAILABILITY

> I do not know if HESS forces the sharing of the source code but I believe that, potentially, MIKA-SHA can be a valuable tool for the hydrological community and, therefore, I invite you to make it publically available.

[Response] We are pleased with your comment. Currently, we are adding several other building blocks to the toolkit where the hydrologists may have more flexibility when they use MIKA-SHA. We are planning to make the toolkit public in the coming year (2021) with a user manual. Further, we need to gain permissions from some other authors before making MIKA-SHA public as we are using their findings as the elements of hydrological knowledge incorporated within the toolkit.

FURTHER COMMENTS

> L18: MIKA-SHA - do you want to make it look like MIKE-SHE? If yes, motivate; if no, keep this in mind

[Response] Yes, our name was inspired by the popular distributed hydrological model MIKE-SHE.

> L19: meaning of induces? "creates"??

[Response] Yes, it means creates or generates.

> L21: What do you mean with "internally coherent"

[Response] we use the term “internally coherent collection of building blocks” to refer the building blocks themselves connect or follow in a reasonable way and each part of them are carefully considered within the framework. A good example of this would be the model-building components of a flexible modelling framework.

> L27: "signatures" has a specific meaning in hydrology, which may be different from your intentions

[Response] Term “signatures” will be changed to “dynamics” in the revised manuscript.

> L30: hydrological model - hydrological model structure

[Response] Will be corrected in the revised manuscript.

> L44: induction - consider changing with "selection" since more common in hydrology. Here and afterwards

[Response] In this manuscript, we specifically chose the term induction instead of selection to state that the algorithm itself builds the model structure within its optimization process rather than selecting an already available model structure. As you pointed out the term selection would be more common in hydrology however, the term induction is frequently used in GP applications in hydrology.

> L65: vs. (lower case and with the dot at the end) (here and elsewhere)

[Response] Will be corrected in the revised manuscript.

> L66: this definition applies both to physical based (in the sense of models that are intended to represent the phenomena happening in the catchment) and conceptual models (reservoir models). Clarify better what you mean.

[Response] Yes, this definition applies to the conceptual models as well. Since this section in the original manuscript differentiates physics-based models and data science models, we used this definition only for physics-based models. However, as we are planning to use the structure you proposed for the revised manuscript where the section heading would be physics-based models vs. conceptual models vs. data science models, we can use the same definition on physics-based and conceptual models.

> L78-79: elaborate more. also on the role of conceptual models. In theory if a physics-based model has only pars that can be related to physical measurable properties then you have a different data requirement compared to conceptual models that calibrate their parameters based on model outputs.

[Response] As you correctly said, in this context, physics-based models and conceptual models may have different data requirements (e.g. physics-based models require measured physical quantities to use as model parameter values while conceptual models need measured catchment responses to calibrate the model parameters). Hence, the differences in data requirements between the conceptual and physics-based models will be more elaborated in the revised manuscript.

> L86: was – has?

[Response] Will be corrected in the revised manuscript.

> L93: I would consider introducing conceptual models in the previous paragraph, in contrast with physically based and data driven. Here you can narrow the focus on conceptual and bring the differentiation between fixed and flexible

[Response] Will be changed as proposed in the revised manuscript.

> L111-113: bring references to this statement

[Response] Following reference will be added in the revised manuscript.

"A simple illustration of the practical limitations of a fixed model structure is the need to add specialized modules for specific catchment conditions. For example, in many models, the simulation of snowmelt requires the addition of an external snow module, already implying that the overall model structure requires customization for a specific climatic region." – Fenicia et al. (2011)

> L115-116: (Knoben, 2019) - paper 2020 wrr on applying the models on the CAMELS

[Response] The mentioned recent paper (Knoben et al., 2020) will also be cited in the revised manuscript.

> L130: subjectivity - motivate more on this. usually the best model solution is selected testing different options which, while being subjective, are designed to cover a wide range of possibilities, in order to exclude subjectivity from the process.

[Response] We use the term subjectivity here to refer the model structure selection or development based on personal judgement, preference and experience. We recognise that the model selection or development based on expert's knowledge can be as good as those through an automative process (can even be better than models through an automative process). However, expert knowledge might not be available and may be expensive. In such situations, we believe an automative model building algorithm would be more appropriate. As you have proposed we will elaborate more on subjectivity in the revised manuscript.

Printer-friendly version

Discussion paper

> L169-171: maybe a ref on this

[Response] Following reference will be added in the revised manuscript.

“It is now more than 40 years since Freeze and Harlan published their seminal blueprint for a physically-based digitally-simulated hydrologic response model”. “At the time, implementation of those process descriptions was severely limited by the computer power available. The expansion of computer power in the last 40 years has, however, greatly relaxed this constraint and it is possible now to apply such models with a fine discretisation to both small and large catchments.” – Beven (2012)

> L177: will – would?

[Response] Will be corrected in the revised manuscript.

> L205-206: kratzert 2017 hess and subsequent papers

[Response] We appreciate pointing out the related work of Kratzert. Related citations (Kratzert et al., 2018; 2019a,b) will be added to the revised manuscript.

> L213: process-based - keep consistency. it was physics based before

[Response] Will be corrected in the revised manuscript.

> L214: much less effort - depends how you define effort. human effort -> yes computational effort -> no (probably)

[Response] Yes, as you said it depends on how we define it. We used the term to refer to human effort. We will clearly define this in the revised manuscript.

> L216: scientific theories - maybe better domain knowledge, although data science without domain knowledge is the perfect recipe for a disaster

[Response] Yes, domain knowledge would be the more suitable term here. Will be corrected in the revised manuscript.

> L231: which – that?

[Response] Will be corrected in the revised manuscript.

> L241: Overall, I don't like the structure. I would write sth like:

- AAN is a subset of ML

- describe briefly AAN and their applications in hydrology

- move to RNN and LSTM, explaining why they are useful to model RR

[Response] Will be restructured as you proposed here in the revised manuscript.

> L255-261: provide some references for the points you are addressing here

[Response] Following references will be added here in the revised manuscript.

"The literature shows that ANNs suffer from some apparent drawbacks and limitations, which are local minima, slow learning speed, over-fitting problem and trivial human intervention such as learning rate, learning epochs and stopping criteria." – Yaseen et al. (2015)

"The fact that there is no standardized way of selecting network architecture also receives criticism. The choice of network architecture, training algorithm, and definition of error are usually determined by the user's past experience and preference, rather than the physical aspects of the problem." – Govindaraju (2000)

> L272: you are missing all the kratzert-nearing part

[Response] We find the related work of Kratzert and Nearing are highly appropriate for the content of our paper. Hence, the related work will be cited in the revised manuscript. Rainfall-runoff modelling: Kratzert et al. (2018, 2019a,b); Nearing et al. (2020)

> L286: It is not clear what differentiates GP from ANN. To my understanding, it looks like ANN have predefined functions and you tune the parameters in training while GP writes the functions.

[Response] Yes, your understanding is correct. In the context of this application, there is no need to predefine the model structure for GP. GP itself develops (induce) the model structure using the available building blocks. GP is capable of optimizing both model structure and parameters simultaneously. Further, the capability to produce explicit mathematical input-output relationships makes GP distinctly different from the rest of the machine learning techniques.

> L326: Declarative bias and preferential bias - explain this in simple words. The audience may not know GP

[Response] The two terms will be explained in simple words in the revised manuscript. In the related work (Keijzer and Babovic, 2002), authors have introduced declarative and preferential bias into the GP algorithm to induce dimensionally correct equations. At the initialization stage, declarative bias forces to sample only the dimensionally correct solutions (a hard constraint on dimensional correctness). On the other hand, preferential bias guides the algorithm towards the dimensionally correct solution (a soft constraint on dimensional correctness) while allowing all solutions to induce.

> L344: I would make a separate section out of this

[Response] As proposed a separate section on the content of this paragraph will be added under the method section in the revised manuscript.

> L364: I would extend this section a lot, maybe including elements of the previous sections. Really difficult to read for somebody that does not know GP. Try to explain with less technical jargon. Plus you are explaining the model like "it is ML-RR-MI + sth" which requires to read about ML-RR-MI.

[Response] As you have stated here and in some earlier comments, we will restructure (expand) this section (where MIKA-SHA is introduced) by including all the relevant parts from other sections. As mentioned above, we will avoid any jargon specific terms and explain with less technical terms in the revised manuscript. Further, we will add a paragraph explaining the basics of our prior toolkit ML-RR-MI in this section.

> L381: multi-objective optimization scheme - did you make a package? Public?

[Response] We have used the approach proposed in NSGA-II (Deb et al., 2002) as the multi-objective optimization scheme in our toolkit. NSGA-II is an augmented version of the popular Genetic Algorithm (GA) which facilitates multi-objective optimization based on Pareto-optimality concept. NSGA-II is a publicly available package for GA. However, the evolution process in both GA and GP are much similar. Hence, we developed our optimization scheme based on the concepts in NSGA-II.

> L381: NSGA-II - overall, the procedure is not very clear to me. Maybe it requires large knowledge of GP or the other framework but I don't think you should rely on it.

[Response] As mentioned in the earlier response, NSGA-II is a well established multi-objective optimisation method not only in GP but also in other evolutionary computation techniques like genetic algorithm (GA). The approach is based on the Pareto-optimality concept. As NSGA-II has been used extensively in many previous studies and the package is publicly available, we avoided explaining it deeper in our paper. However, in our revised manuscript we will briefly explain how our multi-objective optimization scheme operates based on NSGA-II (e.g. paragraph below).

"Each individual (candidate solution) in the population is evaluated on each objective function separately (in the current study we use four objective functions). Based on the objective function values, each individual is assigned a non-domination rank and a crowding distance value. The ranks are identified based on the Pareto-optimality concept. For example, all the individuals with non-domination rank 2 are dominated by individuals with rank 1. However, individuals with rank 2 are not dominated by any other individuals with a higher rank (lower the rank better the individual). On the other hand, crowding distance measures how an individual located relative to the other individuals in the same rank (more the distance better the individual – more diversity). Therefore, in the selection stage, when two individuals are randomly selected and they have different ranks, the individual with the lower rank is selected. If both of them have the same rank, then the individual with higher crowding distance is selected."

> L433: objective functions - this should go into a methods section.

[Response] Will be moved to the method section in the revised manuscript.

> L450: SIS - shouldn't this go into methods?

[Response] Will be moved to the method section in the revised manuscript.

> L464: 4. - is this a continuation of the previous list?

[Response] Yes, it is a continuation.

> L520: Results - if sect 5 is on the case study, shouldn't this be a subsection of 5?

[Response] Yes, it would be more suitable as a subsection of the case study section. Therefore, the results section will be moved as a subsection of the case study section in the revised manuscript.

> L522: induce - use of the word "induce" unclear

[Response] As earlier the term induce is used to refer create, develop, or generate.

> L563: This look like a discussion of the model results, which is nice to have. I would make it more systematic comparing what the model says to your hydrological knowledge (e.g. the model suggests this structure for this HRU and we believe is correct because **) Try to give it a structure..not a single paragraph...

[Response] As you proposed earlier, we will move the discussion part here to a separate section under "Discussion and Conclusions" in the revised manuscript.

> L576: 34 model parameters - it's a lot. I'm surprised it does not overfit

[Response] As our toolkit uses semi-distributed modelling concepts to induce distributed rainfall-runoff models, it assigns a separate model structure to each HRU which may result in high model parameters. However, we use following steps within our framework to remove overfitted models.

(1) Once the Pareto-optimal models are identified based on training fitness values (calibration), their fitness values are evaluated on the validation period using the same multi-objective criterion. Then the Pareto-optimal models are reidentified using both calibration and validation fitness values. Through this, toolkit removes the models which perform better only for the calibration period.

(2) Model parsimony based on the number of model parameters is used as a selection criterion in the optimal model selection stage.

(3) Once the optimal model is selected its performance is evaluated on out of sample dataset (testing period).

(4) In the present study, we use the building blocks of two flexible modelling frameworks as the incorporated hydrological knowledge to guide the learning algorithm (as special functions). These model building components themselves follow certain physical laws within their original frameworks (internally coherent). Hence, we expect the models induced using these special functions to be less susceptible to overfitting than models induced using just mathematical functions.

> Figure 8: Floodplain - not clear if the division in hill/floodplain/plateau is done by the model or predefined by the user

[Response] The division is predefined by the user. The model assigns sperate model structure to each HRU and calculates the total runoff as per semi-distributed rainfall-runoff paradigm.

> L623: as before, make more systematic

[Response] As mentioned above we will move the discussions here to a separate section (Discussion and Conclusions) in the revised manuscript.

> L632: since there are similarities (and probably differences) between FUSE-TOPO-

M1 and SUPERFLEX-TOPO-M2 it would be nice to have a paragraph analyzing this. Having a consistent, but slightly different, model structure selection between the two configurations would be a strong point in favor of MIKA-SHA

[Response] Indeed this is a strong characteristic in favour of MIKA-SHA toolkit. Hence, we will discuss the similarities and differences between the two optimal models derived using the model components of two libraries in a separate paragraph in the Discussion and Conclusions section in the revised manuscript.

> L634: this looks like conclusions of the case study. Re-consider the division in sections.

[Response] Will move to the Discussion and Conclusion section in the revised manuscript.

> L685: Data availability - what about the code of the model?

[Response] As answered earlier, we are working on making the MIKA-SHA code public.

REFERENCES

Beven, K.: Rainfall-runoff modelling: the primer, Wiley-Blackwell, West Sussex, United Kingdom, 2012.

Chadalawada, J., Herath, H. M. V. V., and Babovic, V.: Hydrologically informed machine learning for rainfall-runoff modeling: A genetic programming-based toolkit for automatic model induction, Water Resources Research, 56, https://doi.org/10.1029/2019WR026933, 2020.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resources Research, 44, W00B02, https:// doi.org/10.1029/2007WR006735, 2008.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6, 182–197, 2002.

Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resources Research, 47, W11510, https://doi.org/10.1029/2010WR010174, 2011.

Fenicia, F., Kavetski, D., Savenije, H. H., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modelling decisions, Water Resources Research, 52, 954–989, doi:10.1002/2015WR017398, 2016.

Govindaraju, R. S.: Artificial neural networks in hydrology. II: Hydrologic applications, Journal of Hydrologic Engineering, 5, 124–137, 2000.

Kavetski, D., and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resources Research, 47, W11511, https://doi.org/10.1029/2011WR010748, 2011.

Keijzer, M., and Babovic, V.: Declarative and preferential bias in GP-based scientific discovery, Genetic Programming and Evolvable Machines, 3, 41–79, 2002.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., and Woods, R. A.: A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments, Water Resources Research, 56, e2019WR025975. https://doi.org/10.1029/2019WR025975, 2020.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resources Research, 55, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019a.

Kratzert, F., Klotz, D., Schulz, K., Klambauer, G., Hochreiter, S., and Nearing, G.: Benchmarking a Catchment-Aware Long Short-Term Memory Network (LSTM) for Large-Scale Hydrological Modeling, Hydrol. Earth Syst. Sci. Dis., https://doi.org/10.5194/hess-2019-368, 2019b.

Nearing, G. S., Sampson, A. K., Kratzert, F., and Frame, J. M.: Post-Processing a Conceptual Rainfall-Runoff Model with an LSTM, EarthArXiv preprint, https://doi.org/10.31223/osf.io/53te4, 2020.

Yaseen, Z.M., El-shafie, A., Jaafar, O., and Afan, H.A.: Artificial intelligence based models for stream-flow forecasting: 2000-2015, Journal of Hydrology, 530, 829-844, http://dx.doi.org/10.1016/j.jhydrol.2015.10.038, 2015.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-487, 2020.