Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-487-RC2, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

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

Anonymous Referee #2

Received and published: 2 December 2020

General comments

A new model "induction" framework (labelled as MIKA-SHA, though I would assume it should be MIK-ASHA based on the full title of the framework on Line 18 "Machine Induction Knowledge-Augmented System Hydrologique Asiatique") is proposed which builds on an existing induction framework (ML-RR-MI), moving from a fully lumped conceptual model to a semi distributed model. The main driver of this research seems to stem from the well-documented problem of spatial heterogeneity of forcing data and model parameters on model calibration/performance. A watershed from Mississippi, USA is used as a case-study to demonstrate the utility of the framework.

Printer-friendly version

Discussion paper

The main focus of this paper is on the use of machine learning to inform rainfall runoff modelling. This focus is of interest to many in the field and the development and application of such a method is welcome. However, I have a number of concerns related to the paper, primarily the first 12 pages, which I think take away from the main message.

These first pages provide a very basic review of well-established topics in hydrological modelling (e.g. physics based vs data based modes, fixed v flexible, lumped v distributed, etc.). I think the authors' intentions are to provide some baseline definitions and maybe highlight some common issues under each subtheme. However, in its current form, it reads like a very "high level" summary (similar to textbook descriptions) rather than adding value or providing the current state of research in each field. The way this initial text is framed suggests that the proposed framework will collectively address many of the pitfalls found in existing data-driven methods, whereas in fact the proposed method has a very specific focus (more details on this are below). Many of the problems with hydrological models (physics or data-driven) are well-documented and it would be better if the authors focus on some recent innovations in these fields, and used these advances as a basis of comparison for their proposed method (some suggestions below). This would enable a proper or more detailed evaluation of the proposed method, which currently seems lacking. I would recommend reducing the focus on the introductory text and expanding the methods sections (Section 4 and 5) to better explain the overall method (much of it is in bullet form). Lastly, the analysis is performed on a single watershed, but the discussion of the results seems to imply that the advantages of the proposed methods are more widespread (which may indeed be the case), but I think a comparison with other watersheds may be beneficial for to support these statements. The majority of the comments below seek clarity in the text (which in general can be improved); some comments I disagree with and have offered alternative.

Specific comments

Abstract Machine learning and data sciences have been extensively researched in hydrology – I do not agree that "in general,. . .limited use in scientific fields". I think a better description would be that perhaps they are not put in practice as much as they are used for research applications. However, support for such a statement would be difficult to quantify. Additionally, "commercial fields" is a vague term that should be better described.

FUSE and SUPERFLEX are mentioned in the abstract without any contextual information: what are these frameworks?

Very limited information, other than a general description of the framework and its advantages, are presented in the abstract. It would be useful for a reader to better understand how the model works, how well the model works, and the basis for the conclusion that the model is ". . .compatible with previous findings".

Introduction

"Hydrological models play a key role in capturing the discharge signatures of watersheds." I am not sure if this statement is really necessary – in some ways it's obvious, and in others, I am not sure if it is correct: models are used to simulate the discharge in a watershed; underlying data provides the mechanism to understand the system. By discharge signatures, do the authors mean the factors driving the discharge?

"So far, no hydrological model can perform equally well over the entire range of problems." Unclear what the authors mean by problems here. But if the message is that no hydrological model can simulate a system under the full range of observed conditions, perhaps some supporting information is needed.

"This leads to different research directions seeking different hydrological models based on different modelling strategies." Again, this is a pretty vague statement and not useful in outlining the present research. Are the authors trying to state that several hydrological modelling approaches are currently being researched to solve different research problems?

“Therefore, the final goal of any successful hydrological model must be based on a physically meaningful model architecture along with a good predictive performance.” This also seems fairly obvious and standard practice in model development and selection.

L36: see comments above re: data science and “commercial” fields. One can argue that statistical methods have been successfully used in hydrology for decades, and statistics can be considered as a part of “data science”?

“data-driven models are often performing better in terms of predictive capabilities”: I think the authors should provide some references to direct comparisons between physics-based and data-driven hydrological models to support this statement. While I agree that data-driven models can perform exceptionally well in predicting a given output, very little research compares directly with physics-based models. Of course any comparison would also have to include issues related to complexity, computational expense, etc. along with common model performance metrics.

On a related note: “they may contribute little towards the advancement of scientific discovery due to the lack of interpretability of the model configurations.”: I think the idea of treating data-driven or machine learning methods as a “black box” is not entirely fair. Sufficient information can be gleaned from a simple ANN, as an example, to understand the effects of the forcing on the output, the sensitivity, and so on.

My comments above are not to undermine the central thesis of the paper, I do think there is an interest and a need to study “physics informed machine learning”, but I do not think some of the statements made by the authors are necessary to make this claim. The premise ought to be clarified and perhaps toned down. In addition, I think the authors can cite appropriate literature to better highlight the need for their proposed framework: are their many studies that demonstrate the inability of standard ML approaches to not capture physical phenomenon correctly? Similarly, with respect to model complexity (objective 2), some reference to current literature that explore complexity for model selection should be highlighted (the “simpler the better paradigm” is not the only paradigm currently being used).

Fundamental Approaches in Hydrological Modelling

Section 2.1: I don’t think this discussion is really necessary. These concepts have been explored in numerous research papers over the years, and in my opinion are well-understood. The review of current research presented here is short, and no new information or discussion is presented, that isn’t well-understood. I think the manuscript would benefit from reducing length of the text and many of the introductory comments here can be removed, save for presenting the central thesis of TGDS (which to some extent is already presented in the Introduction).

“The first reported physics-based model was introduced by Freeze and Harlan (1969).” Perhaps a caveat should be added that this refers to a digitally-simulated hydrological model. Physics based models in general are not that recent – whether a computer was used or not is a different question.

“This dichotomy led to the evolution of two major communities in water resources engineering: those who work with physics-based modelling and those who deal with machine learning techniques, which appear to be working quite separately.” – Perhaps this is a generalisation? Many work in both communities simultaneously. This distinction and statement is not necessary for the central thesis of the paper.

“The key concept behind this approach is to incorporate the existing body of scientific knowledge into learning algorithms to come up with physically meaningful models with good predictive power.” Repeated text from Introduction.

Section 2.2: As per my comment in Section 2.1, much of this discussion will be well-known to readers of HESS. Conceptual modelling is well defined and understood. I don’t think it is necessary in this paper.

L100: equifinality: other reasons for equifinality may also exist than those cited by the authors (e.g., measurement uncertainty, lumping).

L119: to “customize model structure” alone cannot be the difference between fixed and flexible?

“Hence, a hydrologist with novice knowledge would require to test many model structures beforehand selecting an optimal model which is time demanding and computationally intensive, in consequence, hinders the opportunity to use the flexible modelling frameworks in their full potential. Further, the selection of a model configuration without testing a large number of possible combinations may introduce a high level of subjectivity into the model building phase. Therefore, we find a requirement to automate the model building phase to remove the subjectivity and consider many configurations without direct human involvement.” As per my previous comments, much of the preceding text in the manuscript is not necessary, and not properly supported by references to existing literature, to arrive at the goal quoted above. Model selection based on “subjectivity”, or experience, isn’t necessarily bad on its own. I think the authors are simply trying to state model selection and model configuration is an on-going issue in hydrological modelling research (whether physical or data-based), and there is a need to develop a more independent framework to do this.

Section 2.3: Again, I don’t think the elementary definitions of lumped v distributed v semi-distributed models are necessary. Paper length can be reduced by removing these sections in the review.

L176 and 178: citation to appropriate literature are need to support these statements. L182: what are the effects of over-parameterisation? I assume the authors are alluding to difficulty in calibration, but this should be explicitly stated.

Machine Learning in Water Resources

Much of the information presented until Section 3.3 are not really necessary – these are well established within hydrological models. As with previous sections, the current review is not thorough enough (if the intention is to provide a full review), while on the other hand, not necessary to arrive at the main objective of the paper. Comments below require attention, but my recommendation is to drastically reduce this background information.

“Another major advantage of a machine learning model is that it requires much less effort to develop and calibrate than a physics-based model.” Perhaps a reference to support this statement can be included?

L216: perhaps replace scientific theories with physical hydrology?

L247: “several output nodes”: should be “one or more” . . .there is not requirement to have more than one node in the output layer.

L253: feed forward is a type, back propagation is a training algorithm.

“ANN can handle incomplete or erroneous data, highly complex and interdependent parameters.” I don’t agree that “erroneous” data can be “handled” by ANNs per se.

“One of the key disadvantages of using ANN for data modelling is the fact that it produces overfitting results which make it difficult to extrapolate beyond witnessed train data.” This is contradictory to the statement about ANN handling “noisy data”. Models that are prone to overfitting do not deal well with overfitting. Next, there are several well established methods for preventing overfitting (i.e., regularisation, drop-out, stop-training, cross-validation, etc.). Finally, extrapolation on data beyond the training domain is a distinct issue from overfitting and model generalisation. A generalised model should not be expected to perform well on data outside of the training domain.

“determining the efficient network architecture and tuning hyperparameters make it hard for the user to completely understand how the model makes its predictions.” There is considerable research in this field both in and out of hydrology.

L300: “This modelling paradigm aims to simultaneously address the limitations of data science and physics-based models and induce more generalizable and physically consistent models": This is the major premise of the proposed research. However, I do not think the authors adequately cover existing research efforts in addressing these issues. A fairer comparison would be to simply highlight known issues with ML methods (e.g., "black box", over fitting, uncertainty, generalisability, model selection) and then highlight current efforts to address these in hydrology. Similarly, known issues with physics-based modelling (e.g., spatial-temporal issues, data resolution, uncertainty, model selection) can be highlighted along with recent efforts to address these issues. Following this, the proposed framework can be described as an alternative way to collectively address these issues. However, the proposed framework is not attempting to address all the limits of both data-centric and physics-based models, rather MIKA-SHA seems to address issues related to spatial heterogeneities and a model selection only.

L313: "ANNs suffer the most severe consequences of lack of interpretability of resulted models" Please provide some evidence for this statement.

Results

"Results of this study, such as achieving high efficiency values for the absolute performance measures and obtaining a good visual equivalent between measured and modelled hydrographs suggest that topography of the catchment may have a strong impact on runoff generation." It is not clear to me how the model performance metrics can indicate the latter – please expand.

"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 perhaps an over generalisation: (1) Table 8 shows the testing performance is lower for each metric; (2) to conclusively say this, a cross-validation approach is recommended (where the testing dataset is iteratively changed, and the performance under each training scenario is calculated).

With the results in general, I think the authors can make a stronger attempt to connect to their central thesis. How does MIKE SHA address all the comments and limitations that the authors explored earlier in the paper with respect to data-driven and physics-based model. A key limitation, for me, was difficulty in understanding the proposed method in Section 5 (and to some extent in Section 4). Some claims, as commented on above, are difficult to confirm if the results are based on one catchment, under one training-validation-testing split. The question is if this is repeatable under different conditions?

Conclusions

“may contribute to the development of accurate yet pointless models with severe difficulties with interpretation may not serve towards the advancement of hydrological knowledge” I disagree with the “pointless” claim. . .I think it is OK to highlight some of the limitations of data-driven models, but calling them pointless collectively is not fair.

Technical corrections

L122: his/her => their

L211: rewrite for clarity “potential to apprehend the noise complexity”

L242: transpired should be inspired? Not limited to human brains, by the way.

L426: “greater extend.” Should be extent?

L428: “a bunch of”: perhaps “set” is more appropriate than “bunch”

Table 4: “ln” is defined at the bottom of the table, but the equation uses “log”; unsure if the log is ln or base10. Please clarify.

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