Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/hess-2021-423-AC2, 2021

Reply on RC1

Jonathan M. Frame et al.

Author comment on "Deep learning rainfall–runoff predictions of extreme events" by Jonathan M. Frame et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-423-AC2, 2021

Thank you for your thoughtful review.

You ask a few questions that I think might be rhetorical, but I will answer them assuming that they are genuine.

RC1: ["in ML we are hoping to program computers by telling them what we want to achieve without having to explicitly instruct them how to achieve such goals."]

The algorithms in ML are explicit instructions.

RC1: ["what is it that we want from those programs?"]

We want these programs to simulate the physical watershed processes resulting in streamflow during extreme runoff events.

RC1: ["Do they just need to be accurate or should we also be able to interpret them?"]

Accuracy and interpretation are not competing interests. ML models are just as interpretable as numerical solutions to partial differential equations (PDEs). As a matter of fact, the interpretation of PDEs and ML happens in the same way, through sensitivity analysis and visualization of the 1) relationship between input and diagnostic variable, and 2) the phase-space of model states.

RC1: ["In the case of two equally good approximations of the same data set: the one which blindly fits the data and the other which, in addition to the fit, also respects the background knowledge, we should be biased toward the latter one. "]

If background knowledge is desirable, and it does not hinder the value of the products, then it is surely better to have background knowledge. But I do not understand your statement of "blindly fits the data". Is this some assumption that one can make good predictions, but not understand how they are making good predictions? If so, this is not applicable, as we do in fact know how LSTM makes good predictions of streamflow during extreme events, and that is: The LSTM was trained to simulate the hydrological processes of a watershed by finding model parameters to represent dynamic relationships between static watershed attributes, dynamic atmospheric forcings and streamflow response. Exactly the same way as any other dynamic hydrological model.

RC1: ["The fashion in which we express knowledge about the processes and make it available to the learning machine remains rather unclear."]

We express knowledge about the processes (I assume you mean hydrological processes) by setting up the input data (static watershed attributes and atmospheric forcings) that we assume will be informative to predict the target (streamflow). The "learning machine" has available to it a sample of the input and target data. We have set up our experiment here to test the hypothesis that the sample data is sufficient to learn a relationship that is suitable for extremely high runoff events.

RC1: ["One can insist on strict adherence to the background knowledge principles - such as 100% mass balance accuracy. We declare this desire and hence this is referred to as declarative bias. "]

We do not have access to enough measurement data to attempt a 100% mass balance accuracy. We cannot distinguish between losses, from our watershed-scale control volume, to the atmosphere or the ground. We can only attempt to parameterize these losses through our model architecture.

RC1: ["Alternatively, one can treat the bias as an additional objective that should be treated simultaneously with the goodness of fit in the learning process. This is referred to as preferential bias. "]

Preferential bias is "preference for one class of concepts over another class". The LSTM, MC-LSTM and SAC-SMA are trained and calibrated to prefer a higher NSE score. We then test these models on their performance predicting the peak flow to address our hypothesis.

RC1: ["Declarative bias reduces search space but results in so-called broken ergodicity. Preferential bias results in a Pareto-optimal set of solutions."]

And according to Keijzer and Babovic, reducing the search space does not help finding better solutions faster. We address this in our paper on lines 236-242.

RC1: ["In the present paper, it would appear that authors prefer declarative treatment of background knowledge. However, I would appreciate further analysis, comparison, and, if that is not possible, at least a discussion on preferential vs. declarative bias in the case studies described in theirs work."]

No, the authors do not prefer declarative treatment of background knowledge. According to Keijzer and Babovic, reducing the search space "does not help finding better solutions faster. In fact, for the class of scientific discovery problems the opposite seems to be the case." **We come to the same conclusion in our paper (lines 234-242)**: "It is important to understand that there is only one type of situation in which adding any type of constraint (physically-based or otherwise) to a data-driven model can add value: if constraints help optimization. Helping optimization is meant here in a very general sense, which might include processes such as smoothing the loss surface, casting the optimization into a convex problem, restricting the search space, etc. Neural networks (and recurrent neural networks) can emulate large classes of functions (Hornik et al., 1989; Schäfer and Zimmermann, 2007), and by adding constraints to this type of model we can only restrict (not expand) the space of possible functions that the network can emulate. This form of regularization is valuable only if it helps locate a better (in some general sense) local minimum on the optimization response surface (Mitchell, 1980). And it is only in this sense that constraints imposed by physical theory can add information relative to what is available purely from data." We do not believe that further discussion is required.

RC1: ["Bias Variance Tradeoff. Arguably incorporation of the knowledge bias affects model variance. In this case, bias denotes the difference between the average prediction of a model and the correct value which it is trying to predict. Variance is the variability of model prediction for a given data point or a value that tells us the spread of our data."]

This is well covered in our paper. We train/calibrate our models using the Nash-Suttcliffe Efficiency (NSE), which when decomposed includes a term for bias (Ratio of the means of observed and simulated flow) and a term for variance (Ratio of standard deviations of observed and simulated flow). There is certainly a bias-variance tradeoff in our trained/calibrated models, but the NSE as a loss function is a good way to include these two terms. In Table 2 we present the results of both the Apha-NSE (The variance term), and the Beta-NSE (The bias term).

According to Elements of Statistical Learning "to trade bias off with variance in such a way as to minimize the test error." The bias-variance tradeoff is an analysis to see how a model generalizes to be used on data that is not part of the training set. We show in our results that the LSTM model generalizes from a training set without extremely large runoff events to low probability, high flow events, that are not included in the training set.

High bias can cause an algorithm to miss the relevant relations between features and target outputs. High variance may result from an algorithm modeling the random noise in the training data

RC1: ["LSTM-type of ML models are extremely good at forecasting. The authors have eloquently argued in favour of the approach in this (as well as in previous) published research works. At the same time, one must consider if such a n ML approaches induce models or forecasters."]

I am not sure of the question here. I think there was some typos, and I am not able to make out the last sentence. The Herath et al. paper is good, but not applicable to this study, which has the specific hypothesis of deep learning predictions in extremely large runoff events that are not included in the training.