Review hess-2021-511

TITLE

How can regime characteristics of catchments help in training of local and regional LSTM-based runoff models?

RECOMMENDATION

Accept after minor corrections

REVIEWER

John Quilty

GENERAL COMMENTS

This paper carefully studies long short-term memory networks (LSTM) for rainfall-runoff prediction, using a large-sample of catchments in France. The key focus is on exploring local and regional models as well as the impact of the 'lookback' period, an important hyper-parameter of LSTM, with respect to predictive performance and physical understanding of the model results. The authors include well-thought out experiments to identify the impact of the lookback period and cases where local and regional LSTM models are best suited. The authors also benchmark LSTM with GR4J, due to its useful ability to capture ground water exchanges with aquifers and/or between catchments.

The authors spend a considerable amount of effort on tying the performance of LSTM, locally and regionally, to a physical understanding of the results. Some examples include the comparison between local and regional LSTM models with GR4J in terms of a water balance exercise in Section 5.3 as well as the ability of LSTM to predict runoff in controlled catchments at a higher degree of accuracy than GR4J (in Section 5.4). This paper also presents findings (e.g., LSTM does not necessarily outperform simple conceptual rainfall-runoff models) that are counter to other recent studies on LSTM (Gauch et al., 2021; Kratzert et al., 2019; Lees et al., 2021); in all such cases, the authors take the time to carefully describe potential reasons for these differences.

Overall, this paper is very strong and I could not find much to criticize. The methodology seems correct. The figures are very nice and easy to interpret and I did not find any of the content, tables, or figures to be superfluous.

I suspect this paper will be very useful to other researchers interested in exploiting the generality of machine learning for hydrological modelling and rainfall-runoff prediction, in particular. I think the paper only needs some very minor corrections and some additional brief explanations (as noted below). Afterwards, the paper could be published.

SPECIFIC COMMENTS

1. Line 32: does LSTM also help mitigate against exploding gradients? If so, this would be good to mention as well.

2. Fig 1: it would be good to include a description of the acronyms HP and FR in the figure caption (since it is unclear what these acronyms represent).

3. Grammatical corrections: for the most part, the paper is well-written but there are a number of grammatical errors. I stopped correcting such errors around line 154. I recommend that a very carefully read through the paper be completed before re-submission.

4. The sentence on line 165 can be moved to the last sentence of the same sub-section. A short sentence, 'The main equations used in LSTM are as follows (Ref, XXX):' can be used in it's place.

5. L205-206: is this sort of standardization the most appropriate for LSTM? Since sigmoid and tanh activation functions are used, should not the data be scaled to [0,1] or [-1,1] as these ranges match the output ranges of the (previously mentioned) activation functions? Perhaps others have adopted the form of standardization adopted here, if so, can the authors indicate this?

6. What were the hyper-parameters (alpha, beta_1, beta_2) set to in the Adam algorithm?

7. Equation 17: what does epsilon represent?

TECHNICAL CORRECTIONS

- Line 16: '…**a** catchment…'
- L35: 'significantly depends'
- L57: 'It is therefore tried to…' Unclear, please re-write this sentence.
- L61: 'should' instead of 'can'.
- L73: 'popular' instead of 'acknowledged'?
- L135: 'The pattern of runoff…'
- L148: 'that have' instead of 'and has'.
- L149: '...**are** unknown lags between the response of the system **and** a continuous input to it.'
- L150: 'signal**s**'
- L154: 'In an ordinary RNN**,** cell information sharing is **achieved** through a feedback connection…'
- Fig 6 caption: '…are computed **by** taking…'

REFERENCES

Gauch, M., Mai, J., Lin, J., 2021. The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. Environ. Model. Softw. 135, 104926. https://doi.org/10.1016/j.envsoft.2020.104926

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol. Earth Syst. Sci. 23, 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., Dadson, S.J., 2021. Benchmarking data-driven rainfall--runoff models in Great Britain: a comparison of long short-term memory

(LSTM)-based models with four lumped conceptual models. Hydrol. Earth Syst. Sci. 25, 5517–5534. https://doi.org/10.5194/hess-25-5517-2021