Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-247-SC1, 2018

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Rainfall-Runoff modelling using Long-Short-Term-Memory (LSTM) networks" *by* Frederik Kratzert et al.

S. Mylevaganam

mylevaganamsivarajah@gmail.com

Received and published: 15 May 2018

1) After decades of entangled and forbidden understanding in the field of hydrology, more systematic research and effort to better understand the underlying processes and the components that form a rainfallâĂŤrunoff model has made to lease a transition from regression based models to more process oriented rainfall-runoff models (see P-1 LN-1:20). In other words, the quest for more process oriented models prevails to better understand a hydrological system of interest. Otherwise, the efforts to better understand the underlying processes and the components that form a rainfallâĂŤrunoff model over the past few decades become futile if more regression based models are used instead of process oriented models. What was the reason to spend many decades to understand the components that form a hydrological system? What

was the reason to spend many decades to seek a transition to a process based model? Therefore, the reasons for reversing the gear that leads to old school of using regressed equation in lieu of processed oriented models are left unfound.

2) As per the authors, the computational requirements and high computational costs are some of the striking factors that force to use conceptual models for operational purposes in lieu of physically based models which have been formulated after many decades of scientific efforts and findings to wade off the lack of understanding of the underlying processes of a hydrological system (see P-1 LN-21; P-2 LN-3). What are those computational requirements and computational costs? It would be more appropriate to list all the computational requirements and computational costs to solve the intended tasks. It would be more appropriate to plot the graph of computational speed with years and technology development to show the current status of technological development to support the hydrological system. As per the authors, the technology development is not at an appreciable level to meet the computational requirement and the computational cost to solve a simple hydrological problem that is explained in the current version of the manuscript.

3) As per the authors, more conceptual based models are used in operational purposes (see P-2 LN-4). From the reader's point of view, the statement of this nature needs more understanding on the purposes that these conceptual models are used. A conceptual model may perfectly suffice an operational need if the need is well governed by the conceptual model. In other words, the selection of models should be based on the need and the problem to be solved. For example, if the need is about the peal flow, a conceptual model may (depending on the consequences of incorrect estimation of flow magnitude) suit the operational need. On the other hand, if the operational need is about the timing of the peak flow, a conceptual model may not meet the operational purpose. Therefore, is it wise to conclude that conceptual models are applied to meet the operational needs? The manuscripts that are cited need to be thoroughly scrutinized to understand the purposes for which those conceptual models are used in lieu

of process based models.

4) As per the authors, the CAMELS dataset (i.e., freely available dataset of 671 catchments with minimal human disturbances across the contiguous United States) contains time series of simulated discharge from the calibrated Snow-17 models coupled with the Sacramento Soil Moisture Accounting Model (see P-9 LN-20). Were the developers of the dataset constrained by high computational requirements and computational costs? How did they develop the time series of simulated discharge for this dataset that represents the contiguous "United States"? What is accomplished by the authors in the manuscript using the proposed LSTM is pending further clarification.

5) As per the authors, process based models are more data intensive (see P-1 LN-18; P-1 LN-21). Therefore, the temptation to adopt data driven models rooted based on ANN and its branches is fast becoming common. However, as per the authors, the data driven models also heavily rely on extensive data for proper training and validation (see P-10 L-17; P-10 L-20). Without proper training with adequate good quality data, data driven models (e.g., ANN) that reveal no physical meaning of the underlying processes of a hydrological system are also not feasible. Therefore, the authors' statements need more clarification.

6) As per the authors, the output (discharge) for a specific time step is predicted from the input x = [x-n, ..., x0] consisting of the last n consecutive time steps of independent variables (daily precipitation, min/max temperature, solar radiation and vapor pressure) and is processed sequentially(see P-4 LN-6). In other words, as per the authors, the rainfall-runoff modeling is represented by the selected independent variables (i.e., daily precipitation, min/max temperature, solar radiation and vapor pressure). Among the selected independent variables, what is the variable that best explains the "infiltration"?

7) In the current version of the manuscript, the distinction between "basins" and "catchments" is not well understood. For example, as per the authors, the CAMELS dataset is freely available and includes meteorological forcing data and observed discharge for

"671 catchments" across the contiguous United States (see P-3 LN-15). However, in the subsequent statement(see P-3 LN-16), the authors state that for each basin, the CAMELS data set also includes time series of simulated discharge from the Sacramento Soil Moisture Accounting Model coupled with the Snow-17 snow model. What are basins? What are catchments?

8) As per the authors, the first 14 years of the 15-year calibration period is the training data and the last, fifteenth, year is the independent validation period (see P-10 LN-10). The selection of training period and the validation period needs more explanation. What is the impact of selecting the last year of the 15-year calibration period for validating the trained model (i.e., LSTM)? Any scientific evidence to show that this type of data selection for training and validating a LSTM model to solve a hydrology related problem works well.

9) In the current version of the manuscript, the length of the input sequence, which corresponds to the number of days of meteorological input data provided to the network for the prediction of the next discharge value is set to 365 days in order to capture at least the dynamics of a full annual cycle(see P-6 LN-22). Would this lead to a highly memorized network? What would be the status of the trained network if the length of the input sequence is set to 90 days instead of 365 days? How did you validate your trained network with one year of data (see comment-8) when the length of the input sequence is set to 365 days?

10) What is the definition of "HUC"? As per the USGS website (https://water.usgs.gov/GIS/huc.html), the HUCs contain either the drainage area of a major river, such as the Missouri region, or the combined drainage areas of a series of rivers, such as the Texas-Gulf region, which includes a number of rivers draining into the Gulf of Mexico. With this definition of HUC, is the development of a more generalized model (see P-10 LN-30) for each of the selected HUCs misleading? Moreover, are the catchments/basins in each HUC ungaged (see P-10 LN-27)? I think, the current version of the manuscript is distant from providing all these details.

11) The section 2.1 needs an example to illustrate the use of the authors' mathematical formulations of LSTM. For example, on Wednesday, May 16, 2018, if the precipitation, max temperature, min temperature, and vapor pressure are p unit, l unit, h unit, and v unit, respectively, how the reader of this manuscript uses the developed LSTM model to determine the intended output (i.e., discharges/runoff?) is needed. Otherwise, the equations that are formed and welded would lead to rely on mathematicians to decode and understand.

12) What is meant by SAC-SMA+Snow-17(see P-1 LN-11)? Is it meant to convey that the outputs of SAC-SMA and Snow-17 are added to determine the final output? What is the output of SAC-SMA? What is the output of Snow-17?

13) In the current version of the manuscript, some of the cited manuscripts are questionable. For example, citing Shen et al., 2018 to state that the "potential use and benefits" of DL approaches in the field of "hydrology and water sciences" has "only recently come into the focus of discussion"(see P-2 LN-32). Are the authors citing Shen et al., 2018 based on their relationship with Shen et al., 2018? When that manuscript (i.e., Shen et al., 2018) is under severe criticism from the esteemed referees, does it make sense to give credits for that manuscript? Moreover, the cited manuscripts to support the following statements also lead to confusion. Would it be possible for the authors to state the reason for citing these manuscripts to support the statements?

a) The transferability of model parameters (regionalization?) from catchments where meteorological and runoff data are available to ungauged or data scares basins is one of the ongoing challenges in hydrology (Buytaert and Beven, 2009; He et al., 2011; Samaniego et al., 2010).

b) The second motivation is the prediction of runoff in ungauged basins, one of the main challenges in the field of hydrology (Blöschl, 2013; Sivapalan, 2003). A regional model that performs reasonably well across all catchments within a region could potentially be a step towards the prediction of runoff for such basins.

Minor Comments: In Figure 5, would not it be appropriate to show the HUC boundaries instead of the state boundaries. In Figure 5, the precipitation values are given in mm/yr. However, on P-9(see LN-19), the precipitation values are given in mm/year. Should it be "dataset" or "data set"?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-247, 2018.