Geophysical Research Abstracts
Vol. 21, EGU2019-13795, 2019
EGU General Assembly 2019

Using large data sets towards generating a catchment aware hydrological model for global applications

Frederik Kratzert (1), Daniel Klotz (1), Mathew Herrnegger (2), Sepp Hochreiter (1), and Günter Klambauer (1)

(1) Institute for Machine Learning, Johannes Kepler University, Linz, Austria (kratzert@ml.jku.at), (2) Institute of Hydrology and Water Management, University of Natural Resources and Life Sciences, Vienna, Austria

One reason for the success of deep learning across many applications (e.g. image classification or machine translation) are large and publicly available data sets. In Hydrology, more specifically in the area of rainfall-runoff modelling, such publicly available data sets are still rare but are increasingly emerging from the community (e.g. CAMELS [1, 2] and CAMELS-CL [3]). Having such data sets allows to test new data-driven approaches for old but still outstanding hydrological problems.

In this study we build upon the idea of Kratzert et al. [4] using LSTM (Long Short-Term Memory) networks as a regional hydrological model. However, substantial changes are made: (a) We train a single model on hundreds of catchments across the USA, not only for a single HUC (hydrological unit) as in [4] (b) We do not only use meteorological time series as inputs but also static catchment characteristics through an adaption of the original LSTM architecture and therefore allow the LSTM to learn a catchment aware hydrological model. Here, we only use catchment characteristics of [2] that do not rely on discharge observation. With this setting we can therefore also investigate, if LSTMs can be used for prediction in ungauged basins. Concretely, we perform 6-fold cross-validation on 531 basins of the CAMELS data set. Out of the 6 splits, 5 are used for training and validation and a single split (approximately 89 basins) as test set for the final evaluation. This mimics an ungauged basin setting by not using any data from the basins in the test set during the entire training period.

Our results show that the simulations of the LSTM in the ungauged setting exhibit similar performances to simulations of established hydrological models, with the difference that they were explicitly calibrated for each of the basins. Another benefit of our LSTM variant is that an explicit clustering of the basins is obtained due to the way the basin characteristic are used internally.
Furthermore, if trained on even larger, potentially even global, data sets, we hypothesize that this approach could result in a sound global hydrological model with strong regionalization capacities for many ungauged catchments all over the world.

References:

[1] A. Newman; K. Sampson; M. P. Clark; A. Bock; R. J. Viger; D. Blodgett, 2014. A large-sample watershed-scale hydrometeorological dataset for the contiguous USA. Boulder, CO: UCAR/NCAR. https://dx.doi.org/10.5065/D6MW2F4D.

[2] N. Addor, A. Newman, M. Mizukami, and M. P. Clark, 2017. Catchment attributes for large-sample studies. Boulder, CO: UCAR/NCAR. https://doi.org/10.5065/D6G73C3Q

[3] Alvarez-Garreton, Camila; Mendoza, Pablo A; Boisier, Juan Pablo; Addor, Nans; Galleguillos, Mauricio; Zambrano-Bigiarini, Mauricio; Lara, Antonio; Puelma, Cristóbal; Cortes, Gonzalo; Garreaud, Rene; McPhee, James; Ayala, Alvaro (2018): The CAMELS-CL dataset - links to files. PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.894885

[4] 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.