

EGU22-3281

<https://doi.org/10.5194/egusphere-egu22-3281>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Assessment of Transfer Learning Techniques to Improve Streamflow Predictions in Data-Sparse Regions

Yegane Khoshkalam¹, Farshid Rahmani², Alain N. Rousseau¹, Kian Abbasnezhadi³, Chaopeng Shen², and Etienne Foulon¹

¹Institut National de la Recherche Scientifique - Eau Terre Environnement (INRS-ETE), Québec City, QC, Canada

(yegane.khoshkalam@inrs.ca)

²Pennsylvania State University Main Campus, Department of Civil and Environmental Engineering, University Park, PA, United States

³Environment and Climate Change Canada, Toronto, ON, Canada

Reliable streamflow predictions are critical for managing water resources for flood warning, agricultural irrigation apportionment, hydroelectric production, to name a few. However, there are geographical heterogeneities in available observed streamflow data, river basin geophysical attributes, and meteorological data to support such predictions. Moreover, in data-sparse regions, both process-based and data-driven models have difficulties in being sufficiently calibrated or trained; increasing the difficulty to achieve satisfactory predictions. That being mentioned, it is possible to transfer knowledge from regions with dense and available measured data to data-sparse regions. In earlier work, we have shown that transfer learning based on a long short-term memory (LSTM) network, pre-trained over the conterminous United States, could improve daily streamflow prediction in Quebec (Canada) when compared to a semi-distributed hydrological model (HYDROTEL). The dataset used for pre-training (source dataset) was the Catchment Attributes and Meteorology for Large-sample Studies (CAMELS), while the data for the basins located at the target locations (local dataset) were extracted from the Hydrometeorological Sandbox-École de Technologie Supérieure (HYSETS). Both datasets provide access to various types of information with different spatial resolutions. While HYSETS is generally spanning from 1950 to 2018, the temporal interval for most of the basins reported in CAMELS goes back to 1980. The types of data included in both CAMELS and HYSETS include daily meteorological data (precipitation, temperature, etc.), streamflow observations, and basins physiographic attributes (i.e., considered time-invariant or static). In this work, the techniques applied to further improve streamflow simulations included the use of: (i) streamflow observations and simulated flows from HYDROTEL as input to the LSTM model, (ii) different forcing (meteorological data) and static attribute data from the source and the local datasets, and (iii) additional basins from HYSETS with similar climatological features for model training. The ultimate goal was to improve the accuracy of the predicted hydrographs with an emphasis on enhancing the prediction of peak flows by transfer learning while using the Kling-Gupta efficiency (KGE) and Nash-Sutcliffe efficiency (NSE) metrics. This investigation has revealed the benefits of using transfer learning techniques based on deep learning models to improve streamflow predictions when compared to the application of

a distributed hydrological models in data-sparse regions.