A climate similarity-based transfer learning using global Caravan dataset for enhancing streamflow prediction in Ouémé River Basin (Benin, West Africa)

(Oral)

Jérôme Enagnon Ahouandjinou[1,2]*, Aymar Yaovi Bossa[1], Jean Hounkpe[1], Riccardo Taormina[3]

[1]*University of Abomey-Calavi/National Water Institute (INE), Abomey-Calavi, Benin*
[2]*Université of Abomey-Calavi/International Chair in Mathematical Physics and Applications (ICMPA-UNESCO Chair), Abomey-Calavi, Benin*
[3]*Delft University of Technology/Faculty of Civil Engineering and Geosciences, Department of Water Management, Delft, Netherlands*
**Corresponding author: jeromeahouandj@gmail.com*

ABSTRACT

Reliable streamflow forecasting is a fundamental component of flood risk management. However, the accuracy of such forecasts in data-scarce river basins remains one of the pressing challenges in most African catchment. While Long Short-Term Memory (LSTM) networks have demonstrated their effectiveness in rainfall-runoff modelling, their data-intensive requirements limit their direct applicability in poorly gauged catchments across sub-Saharan Africa. To address this issue, we propose a climate similarity-based transfer learning framework using the global Caravan hydrological dataset to improve local streamflow forecasting in Ouémé River Basin (ORB). Specifically, a tropical-climate catchments are first identified from Caravan global dataset using the Köppen-Geiger climate classification. A composite climate similarity index (IC) is constructed from two normalized climate indices: annual mean precipitation and seasonality index. We selected 200 basins and ranked by IC. A fixed validation and test set of 50 basins was drawn by stratified sampling across IC quartiles from the 200 ranked basins. The remaining 150 basins were then used to define three cumulative pre-training subsets of increasing size (50, 100, and 150 b), ranked by IC. Therefore, the LSTM models are pre-trained on each subset and fine-tuned on the ORB streamflow. The developed transfer learning models are evaluated against the LSTM model trained exclusively on local data as baseline model. Results indicate that our approach enhances forecasting accuracy and outperforms the baseline models, with the best model achieving KGE value of 0.88(50 basins), 0.83(100 basins) and 0.81(150 basins) compared to 0.78 of LSTM trained exclusively on local data. Also, the results indicate that climate similarity offers a practical basis for donor catchment selection, and that transfer learning with global datasets can meaningfully improve flood prediction in data-scarce basins across sub-Saharan Africa particularly in ORB.

Keywords: Transfer learning; LSTM; Flood prediction; Caravan; Ouémé River Basin.