Universität Stuttgart

Cluster of Excellence SimTech
Stuttgart Center for Simulation Science

Clustering catchments for hydrological modelling

M.Sc. Topic

For regional hydrological modelling, in recent years data-driven approaches (such as LSTM networks) have become the state-of-the-art approach[1] due to their ability to assimilate large datasets, exemplified by the 671 catchments with extensive time series data in the CAMELS-US dataset[2]. While previous studies have explored the effect of training on smaller datasets[3], an aspect remains unexplored: the potential of harnessing the diversity inherently present in these datasets.

The goal of this M.Sc. thesis project is to address the following research questions:

- Can the diversity of the CAMELS-US dataset be reduced into a characteristic space?
- Can this characteristic space be exploited to choose training points and train an equally well performing model as the state-of-the-art using fewer data?
- Would this lead to an optimal training strategy for regional hydrological models?

Tasks

- Literature research on regional hydrological models, catchment characterization and hydrological signatures [4], clustering techniques, model training.
- Implementation of a clustering strategy to group catchments in the CAMELS-US dataset by similarity.
- Evaluation of the effect of using a smaller set of very diverse catchments on the performance of a LSTM-based regional hydrological model.
- Writing the thesis and presentation at a colloquium.

Desireable Skills

- Curiosity.
- Knowledge of hydrology, statistics, and/or machine learning.
- Some experience in Python for data analysis and machine learning.

Literature/ Resources

[1] 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. *Hydrology and Earth System Sciences, 23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

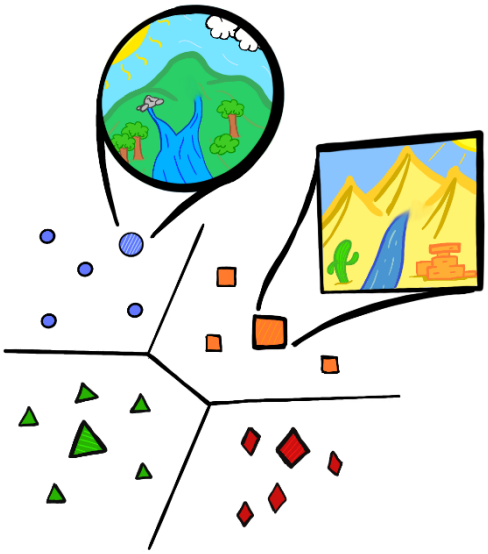
[2] Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences, 21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

[3]* Gauch, M., Mai, J., & Lin, J. (2021). The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. *Environmental Modelling & Software, 135*, 104926. https://doi.org/10.1016/j.envsoft.2020.104926

[4] Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water Resources Research, 54*(11), 8792–8812. https://doi.org/10.1029/2018WR022606

Advisors/Examiners

- Manuel Álvarez Chaves, M. Sc.
- Dr. Anneli Guthke (Independent Research Group Leader for Statistical Model-Data Integration, SimTech)



Apply now!

manuel.alvarez-chaves@simtech.uni-stuttgart.de
anneli.guthke@simtech.uni-stuttgar.de

www.simtech.uni-stuttgart.de