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.

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

Literature/ Resources


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-stuttgart.de

www.simtech.uni-stuttgart.de