EGU24-20170, updated on 20 May 2024
https://doi.org/10.5194/egusphere-egu24-20170
EGU General Assembly 2024

Advances and prospects in hydrological (error) modelling

Carlo Albert
Eawag, Siam, Dübendorf, Switzerland (carlo.albert@eawag.ch)

In recent years, Machine Learning (ML) models have led to a substantial improvement in hydrological predictions. It appears these models can distill information from catchment properties that is relevant for the relationship between meteorological drivers and streamflow, which has so far eluded hydrologists.
In the first part of this talk, I shall demonstrate some of our attempts towards understanding these improvements. Utilising Autoencoders and intrinsic dimension estimators, we have shown that the wealth of available catchment properties can effectively be summarised into merely three features, insofar as they are relevant for streamflow prediction. Hybrid models, which combine the flexibility of ML models with mechanistic mass-balance models, are equally adept at predicting as pure ML models but come with only a few interpretable interior states. Combining these findings will, hopefully, bring us closer to understanding what these ML models seem to have 'grasped'.
In the second part of the talk, I will address the issue of uncertainty quantification. I contend that error modelling should not be attempted on the residuals. Rather, we should model the errors where they originate, i.e., on the inputs, model states, and/or parameters. Such stochastic models are more adept at expressing the intricate distributions exhibited by real data. However, they come at the cost of a very large number of unobserved latent variables and thus pose a high-dimensional inference problem. This is particularly pertinent when our models include ML components. Fortunately, advances in inference algorithms and parallel computing infrastructure continue to extend the limits on the number of variables that can be inferred within a reasonable timeframe. I will present a straightforward example of a stochastic hydrological model with input uncertainty, where Hamiltonian Monte Carlo enables a comprehensive Bayesian inference of model parameters and the actual rain time-series simultaneously.