EGU21-2866
https://doi.org/10.5194/egusphere-egu21-2866
EGU General Assembly 2021

Hybrid modeling using multivariate, discrete probability distributions

Uwe Ehret
Karlsruhe Institute of Technology KIT, Institute of Water Resources and River Basin Management, Karlsruhe, Germany (uwe.ehret@kit.edu)

In this contribution, I will suggest an approach to build models as ordered and connected collections of multivariate, discrete probability distributions (dpd's). This approach can be seen as a Machine-Learning (ML) approach as it allows very flexible learning from data (almost) without prior constraints. Models can be built on dpd's only (fully data-based model), but they can also be included into existing process-based models at places where relations among data are not well-known (hybrid model). This provides flexibility for learning similar to including other ML approaches - e.g. Neural Networks - into process-based models, with the advantage that the dpd's can be investigated and interpreted by the modeler as long as their dimensionality remains low. Models based on dpd's are fundamentally probabilistic, and model responses for out-of-sample situations can be assured by dynamically coarse-graining the dpd's: The farther a predictive situation is from the learning situations, the coarser/more uncertain the prediction will be, and vice versa.

I will present the main elements and steps of such dpd-based modeling at the example of several systems, ranging from simple deterministic (ideal spring) to complex (hydrological system), and will discuss the influence of i) the size of the available training data set, ii) choice of the dpd priors, and iii) binning choices on the models' predictive power.