An information perspective on hydrological learning and prediction

Uwe Ehret (1), Paul Darscheid (1), Grey Nearing (2), and Hoshin Gupta (3)

(1) Karlsruhe Institute of Technology KIT, Institute of Water Resources and River Basin Management, Karlsruhe, Germany (uwe.ehret@kit.edu), (2) Department of Geological Sciences, The University of Alabama, Tuscaloosa, USA, (3) Department of Hydrology & Atmospheric Sciences, The University of Arizona, Tucson, USA

Hydrological systems are often complex and hydrological problems are often underdetermined: Complex due to a multitude of factors influencing hydrological function, acting across a wide range of spatial and temporal scales, underdetermined as we usually lack exhaustive measurements of system properties, initial and boundary conditions, such that identification of system properties or model parameters from local measurements can be afflicted by limited data.

This unfortunate situation is mitigated by the fact that no hydrological system and related problem setting is truly unique, and insight gained in other, similar systems and problems can be used to inform the problem at hand. This is typically done by applying model structures developed in systems deemed similar to the system at hand, and sometimes also by applying model parameters from models calibrated in systems deemed similar.

Seen from the perspective of information theory, taking these steps means that we combine different sources of information, however without explicitly keeping track of the particular uncertainties associated with each of them (e.g. uncertainties due to limited observations or due to only partial agreement of the chosen model structure and the system at hand). Tracking sources of information, or uncertainty, is often further hampered by the use of deterministic models, which offer no direct way to account for uncertainty.

Here we present an approach based on concepts from information theory and by representing relations among data directly by their joint empirical distributions. With ten years of hydrometeorological observations (rainfall, runoff, temperature, snow height) from a mesoscale alpine catchment in Austria, we use Entropy, Conditional Entropy, Kullback-Leibler Divergence and Cross Entropy to quantify and compare
• The information content of the data about the discharge at the catchment outlet
• The information loss if only subsets of the data are available
• The antagonistic effects when adding more predictors to a model: Gain of information due to more predictors, loss of information due to the increased curse of dimensionality
• The information loss when applying transferred relations instead of relations gained from local data
• The information loss when compressing empirical, probabilistic data-relations to deterministic functional forms and applying the latter instead of the first for predictions.

The advantage of taking an information perspective is that it offers a very general language and framework, which allows to explicitly calculate and compare information from various sources such as data or models in a single currency, bit.