



Description length as a performance metric for explaining hydrological data

S.V. Weijs (1), N. van de Giesen (2), and M.B. Parlange (1)

(1) School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland (steven.weijs@epfl.ch), (2) Section Water Resources, Delft University of Technology, Delft, The Netherlands

An explanation for observed hydrological data needs to describe the data correctly and be concise. A good hydrological model can reproduce the output data from input data with small errors, while not being overly complex. When the input, the model and the errors are stored, we do not need the output data anymore and can therefore achieve a shorter description. Good models are thus data compressors.

This is formalized in algorithmic information theory, and can be linked to a Bayesian framework, in which parsimony is used to define the prior probability of a model. Basically, the shortest program that can reproduce the data on some reference universal Turing machine is the best description of the data, containing a codification of the patterns observed in the data, which should hopefully look something like a hydrological model.

Although fundamentally incomputable with finite resources, we can approach the shortest description by exploring data compression techniques. In this research we explored the compression of hydrological data. We present the results and discuss the analogy between modeling and data compression, as well as description length as ultimate performance metric.