Learning with Models & Data: A Maximum Entropy Approach

Hoshin V Gupta

Department of Hydrology and Atmospheric Sciences
The University of Arizona, Tucson, AZ 85721
hoshin.gupta@hwr.arizona.edu

The Robert E Horton Lecture presented at the 2017 AMS Meeting
(Seattle, WA, Jan 2017)

ABSTRACT

In this talk, we will discuss how scientists learn about the nature of dynamical environmental systems via the adaptive process of encoding *Information* of various kinds (based in physics, contained in data, and provided via assumptions) into model hypotheses. In particular, we are interested in models that characterize a physics-based understanding of the world (although many of our comments also apply to databased models), and on the important issues of inferring *System Architecture* and *Process Parameterization*, rather than the comparatively trivial problem of parameter value selection.

The talk will consist of two main parts. In Part I, we discuss the concept of *Information* and its relationship to *Uncertainty*, *Data*, *Models* and the process of *Learning*. We discuss how *Information* is encoded into models as a hierarchy of structural decisions, the kinds of problems that can arise due to making unjustifiably strong hypotheses, and the consequent need for an *informationally justifiable* approach to model development and inference based in the *Principle of Maximum Entropy* (MaxEnt). The resulting model structural hypothesis can then be quantitatively assessed in terms of its *Likelihood* and *Information Content.* Specifically, we show how alternative model architectures of varying complexity can be assessed without the need to make precise statements about the specific forms of the process parameterization equations. In Part II, we show how *Information Theoretic* concepts can be used to assimilate new information (from data) into a *MaxEnt* model hypothesis, and how the information content of the data (in regard to the model) can be meaningfully assessed. Specifically, we show how the likely mathematical forms for the process parameterization relationships can be inferred via a process of *Bayesian Inference*.

To facilitate clarity of communication, the ideas will be illustrated via a relatively simple catchment-scale hydrological modeling study. If time permits, we will also discuss the relationship between the *Learning Problem* and the need for *Operational Implementation.* While computational issues and matters of detail remain to be addressed, the theory provides a robust approach to model development that accurately represents what we actually know (and know that we don't know) about an environmental system, enables a balance between complexity and parsimony, and facilitates ongoing learning.