FORUM

Applying Information Theory in the Geosciences to Quantify Process Uncertainty, Feedback, Scale

PAGE 56

The geosciences are increasingly utilizing a systems approach to quantify spatial and temporal dynamics among multiple subsystems, their couplings, and their feedbacks. This systems approach demands novel strategies for experimentation and observation in the "natural laboratory" rather than in simple controlled experiments and thus relies heavily on Earth system observations and observation networks. Current and forthcoming examples of Earth system observatories include the Critical Zone Observatories (CZOs), the National Ecological Observatory Network (NEON), EarthScope, FLUXNET, National Water Information System/National Water-Quality Assessment (NWIS/NAWQA), and others. These networks are designed to observe complex processes across a wide range of temporal and spatial scales to synthesize scientific understanding of the fundamental interactions across the interfaces of society, hydrology, ecology, atmospheric sciences, and geosciences.

Using the hydrosphere as an example, *James* [2012] advocates an explicit shift in focus to multi-scale interactions and feedbacks between hillslope, watershed, and global scales, including interactions between geophysical and anthropogenic processes. The need to address issues at the interface between the geosciences and society is motivating geoscientists to adopt systems analysis concepts developed in physics, engineering, sociology, and ecology. Some mature systems analysis approaches include statistical thermodynamics, network theory, and hierarchy theory, all of which are related to information theory [*Shannon,* 1948]. The concepts of information theory add value to the analysis of the flow and feedback of mass, energy, and information between multi-scalar processes in geoscience.

Information theory quantifies the information content and structure of messages. It was developed by Claude Shannon in his seminal paper "A mathematical theory of communication" [*Shannon*, 1948]. Shannon's entropy measures the uncertainty of a random variable by quantifying an expected value of information encoded in some message. In other words, Shannon's entropy is a probabilistic concept describing the average uncertainty in an observer's knowledge of a system's state. Data or models that reduce this uncertainty are defined as providing "information." Information theory is part of the family of Bayesian methods, which are unique among statistical and probabilistic methods in that they provide for quantitative statistical inference based on fundamental first principles.

Two broad categories of techniques from information theory have found use in Earth science: (1) signal analysis, model diagnostics, and pattern discovery in dynamic systems using information statistics, e.g. mutual information, relative entropy, and transfer entropy [*Shannon*, 1948; *Schreiber*, 2000], and (2) maximum Shannon entropy estimation, often called "MaxEnt," applied to estimate the most likely state taken by a system given the available empirical information and constraints [*Jaynes*, 1957]. It is important to understand that Shannon Entropic methods and MaxEnt techniques are used to quantify information transfer and are not to be confused with the thermodynamic concept of entropy maximization under the second law of thermodynamics. This theory and the mathematical applications are now mature and are well documented for physical and dynamical geoscience applications (e.g., reviews by *DelSole* [2004] and *Kleeman* [2011]).

Information theory enables a variety of important applications for geoscientists. Previous applications of information theory optimize spatio-temporal observational sampling strategies [*Stoy et al.*, 2009], identify physical process scales in land-surface processes [*Brunsell and Anderson*, 2011], estimate small scale fractal structure based on large-scale properties [*Nieves et al.*, 2011], assimilate data selectively into dynamical forecast models [*Zupanski et al.*, 2007], quantify the probabilistic relationships between remotely sensed climate variables [*Knuth et al.*, 2005], identify teleconnections and climate networks [*Abramov et al.*, 2005; *Donges et al.*, 2009], and use MaxEnt techniques that identify most-probable system states for energy fluxes [*Wang and Bras*, 2011]. Important emerging applications include quantifying sources of error in model outputs compared with observations, delineating climate or ecosystem, and identifying missing predictive information in observations of dynamic systems.

To illustrate a detailed application, *Ruddell and Kumar* [2009] applied information statistics to formally resolve feedback between an agricultural terrestrial ecosystem and the system's meteorological climate forcings (Figure 1 in Additional Supporting Information in the online version of this article). For example, a two-way feedback of information is observed between precipitation and solar radiation processes. This information feedback characterizes the physics of cloud formation and precipitation recycling in the atmospheric boundary layer. The information flows between system processes are dynamic in time and delineate a "process network" describing the constantly changing couplings and feedbacks within the system. In this example, information feedbacks in the atmospheric boundary layer are reduced or severed when the system enters a drought state, as it did during July 2005. Drought can therefore be uniquely described with information flow statistics as a system state characterized by a significant reduction of information feedback associated with precipitation and clouds when a certain threshold of moisture availability is crossed.

Information theory is an important tool for use in geoscientific applications to quantify how and when the subsystems and scales of the Earth system are coupled, to improve models of complex and feedback-based systems, to statistically estimate system states in the presence of process and observational uncertainty, and to quantify the extent to which we can—or cannot—predict, model, and observe the behavior of the dynamic Earth system. We believe that the geoscience community can make better use of the tools and concepts of information theory and that these tools and concepts will add significant value to Earth-observational research.

Acknowledgments

N. A. Brunsell and P. C. Stoy acknowledge funding from the National Science Foundation ("Scaling ecosystem function: Novel approaches from MaxEnt and Multiresolution," DBI #1021095). B. L. Ruddell acknowledges valuable prior collaboration with Praveen Kumar of the University of Illinois and funding from NASA (ESS #NNX06AF71H).

References

Abramov, R., A. Majda, and R. Kleeman (2005), Information theory and predictability for low-frequency variability, *J. Atmos. Sci.*, *62*, 65–87, doi:10.1175/JAS-3373.1.

Brunsell, N. A., and M. C. Anderson (2011), Characterizing the multi-scale spatial structure of remotely sensed evapotranspiration with information theory, *Biogeosciences*, *8*(8), 2269–2280, doi:10.5194/bg-8-2269-2011.

DelSole, T. (2004), Predictability and information theory. Part I: Measures of predictability, *J. Atmos. Sci.*, 61, 2425–2440.

Donges, J. F., Y. Zou, N. Marwan, and J. Kurths (2009), Complex networks in climate dynamics, *Eur. Phys. J. Spec. Top.*, *174*, 157–179, doi:10.1140/epjst/e2009-01098-2.

James, L. (2012), Observatories for integrated water basin science, *Eos Trans. AGU*, *93*(10), 109, doi:10.1029/2012EO100006.

Jaynes, E. (1957). Information theory and statistical mechanics, *Phys. Rev.*, *106*(4), 620–630.

Kleeman, R. (2011), Information theory and dynamical system predictability, *Entropy*, *13*(3), 612–649.

Knuth, K. H., A. Gotera, C. T. Curry, K. A. Huyser, K. R. Wheeler, and W. B. Rossow (2005), Revealing relationships among relevant climate variables with information theory, paper presented at Earth-Sun System Technology Conference 2005, NASA, Adelphi, Md.

Nieves, V., J. Wang, and R. L. Bras (2011), Statistics of multifractal processes using the maximum entropy method, *Geophys. Res. Lett.*, *38*, L17405, doi:10.1029/2011GL048716.

Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 1. Identification. *Water Resour. Res.*, *45*, W03419, doi:10.1029/2008WR007279.

Schreiber, T. (2000), Measuring information transfer, *Phys. Rev. Lett.*, *85*, 461–464, doi:10.1103/PhysRevLett.85.461.

Shannon, C. (1948), A mathematical theory of communication, *Bell Syst. Tech. J.*, *27*, 379–423.

Stoy, P. C., M. Williams, M. Disney, A. Prieto-Blanco, B. Huntley, R. Baxter, and P. Lewis (2009), Upscaling as ecological information transfer: A simple framework with application to Arctic ecosystem carbon exchange, *Landscape Ecol.*, *24*(7), 971–986, doi:10.1007/s10980-009-9367-3.

Wang, J., and R. L. Bras (2011), A model of evapotranspiration based on the theory of maximum entropy production, *Water Resour. Res., 47*, W03521, doi:10.1029/2010WR009392.

Zupanski, D., A. Y. Hou, S. Q. Zhang, M. Zupanski, C. D. Kummerow, and S. H. Cheung (2007), Applications of information theory in ensemble data assimilation, *Q. J. R. Meteorol. Soc.*, *133*, 1533–1545, doi:10.1002/qj.123.

—Benjamin L. Ruddell, College of Technology and Innovation, Arizona State University, Mesa; E-mail: bruddell@asu.edu; Nathaniel A. Brunsell, Department of Geography, University of Kansas, Lawrence; Paul Stoy, Department of Land Resources and Environmental Sciences, Montana State University, Bozeman