Hierarchical Bayesian Modeling in the Environmental Sciences

L. Mark Berliner

Ohio State University

Understanding the Earth's ecological systems and prediction of their responses to environmental variations require a systems approach encompassing information ranging from the climate and weather sciences to ecology and biology. Challenges to achieving these goals include (i) quantification of the complexities of the processes involved (feedbacks, interactions at various space-time scales); (ii) massive, diverse observational datasets; (iii) intricate, typically nonlinear models representing scientific knowledge of aspects of Earth subsystems and their interactions; and (iv) uncertainties associated with both observational data and models. These issues mandate the development of modeling and computational strategies for combining observational data and model-based information, in a way that accounts for uncertainty.

Climate and Weather Prediction.

The Earth's atmosphere and oceans are fluids whose circulations are in response to energy influxes from the sun, and controlled by gravity and the rotation of the planet. Modeling is made difficult because major subsystems of the earth interact. In general, modeling of the planetary system relies on physical and chemical models for the atmosphere and oceans and their interactions, as well as models for planetary ice, land forms, and the biosystem. While computational advances have recently enabled scientists to form global numerical models for the system, these models are highly nonlinear (subject to so-called chaotic effects) and replete with approximations and uncertainties.

Estimation of various features of these models, especially initial conditions, involves the use of observational data. Through applications of methods generally known as *data assimilation*, vast amounts of data are processed to produce inputs for numerical global climate/weather models. Current methods have strong ties to statistical techniques (kriging, Kalman filtering, etc.). However, associating meaningful uncertainties with the resulting forecasts is difficult and the subject of much research. Further, while prediction of weather based on atmospheric models has become quite skillful for a few days, important challenges remain regarding longer-term (seasonal, interannual, interdecadal, and longer) prediction. A familiar example is the El Niño Southern Oscillation phenomenon. This phenomenon is an important controlling influence on climate and weather on global scales. Its behavior and impacts involve complex air-sea interactions. Generally, to improve long-lead forecasts, better forecasts of ocean behavior and better modeling of ocean-atmosphere interaction are needed. The issue of uncertainty arises with a vengeance at such time scales. Most scientists recognize that meaningful climate-scale predictions are not really point predictions in the traditional sense, but rather the construction of probability distributions of future weather. Almost by definition, climate science is a statistics problem!

Environmental Prediction.

A second critical area is the modeling of ecology-climate interactions. This issue includes all the difficulties of climate and weather prediction, but adds the need for modeling the impacts of climatic variation on habitat, and then in turn on the constituent biological inhabitants and their interactions. The challenges are intense; sciences involved range from climatology to ecosystem sciences and biology. Further, our interests involve prediction of environmental features impacting human activities (e.g., agriculture) and health. This element of environmental science is particularly challenging in that humans (like beavers) both respond to and impact their environment in significant ways.

Environmental Policy Making.

While the science I have outlined is fascinating, it has critical, societal motivations. Forecasting short-term weather events (e.g. storms, tornadoes) are of obvious importance. Longer-term environmental prediction is also of clear value. Further, climate change ("global warming") is of obvious interest to us all. Thinking about such issues adds another dimension to the environmental sciences, a dimension that reminds us of the often cited definition of the discipline of statistics: namely, "decision making under uncertainty." What information do we need to make good decisions? Can we even make "optimal ones"? How do we provide intra- and interannual predictive analyses with the intent of performing decision analyses regarding agricultural strategies? Natural resource management? At long-time scales, policy questions relating to our responses (including remediation and associated costs) of climate change offer extreme challenges. Quantitative answers require intricate combinations of all the issues and sources of uncertainty raised above, augmented by inputs from the medical, social, and behavioral sciences.

A Paradigm for the Environmental Sciences: Hierarchical Bayesian Analysis.

Our knowledge about the environment accrues both from scientific ideas and models and from observations. However, both sources of knowledge are complex and subject to errors. Bayesian statistical modeling is suggested as a mechanism for extracting the information in both sources in a fashion that manages our uncertainties: (i) since the Bayesian viewpoint provides an effective approach to decision analysis, it provides a framework for treating environmental policy making; (ii) Bayesian analysis is well-suited for prediction based on the combination of relevant scientific understanding and data; and, (iii) the hierarchical Bayesian viewpoint offers a natural structure for dealing with complexities arising in the environmental sciences.

Point (ii) warrants further clarifications. Bayesian analysis is typically touted as an approach that facilitates the incorporation of prior information. In the environmental sciences, this attribute is critical. Bayesian modeling includes a broad spectrum of models in which scientific understanding may be incorporated in a variety of ways ranging from qualitative construction of models and priors in concert with scientific notions to quantitative use of numerical models both in the construction of priors and as components embedded in stochastic models.

Regarding Point (iii), note that my presentation of the issues in environmental science was itself hierarchical. I began with review of climate and weather modeling. From there, we turn to modeling climate impacts on habitat, and resulting impacts on the behavior of organisms. I can envision writing down a corresponding sequence of conditional probability models, say the marginal distribution of large-scale climate variables; the conditional distribution of regional weather variables conditional on large-scale climate; the conditional distribution of habitat given weather; and the distribution of organisms (or at least selected properties of organism populations) conditional on habitat. Probability theory provides formulas for the combination of these distributions yielding predictive distributions for ecological variables based on climatic modeling.

Readers will surely join me in finding the suggestions here to be ambitious, to say the least. On the other hand, scientists do seek such ambitious prediction analyses; I know of no other quantitative approach that could achieve the requirements of combining information sources efficiently, while reflecting their uncertainties. However, implementation of Bayesian strategies is not without its challenges. In particular, without the recent advances and the promise of continued advances in Bayesian computation (e.g., Markov chain Monte Carlo), the suggestions discussed here would be mere pipe dreams. I believe that progress in the application of hierarchical Bayesian analysis in large-scale environmental problems is not only desirable, but also achievable.