Invited Feature

Model improvement via data assimilation toward ecological forecasting[1]

The field of ecology has changed to a data-rich enterprise, largely due to rapid development of measurement sensors and long-term accumulation of data from research networks. With implementation of the National Ecological Observatory Network (NEON), a network with different kinds of sensors and measurements at many locations over the United States, large volumes of ecological data will be generated every day. Thus, there will be an unprecedented demand to convert massive amounts of raw data into ecologically meaningful products using data assimilation (DA) techniques.

DA is a tool that combines observational data with ecological models to provide reliable ecological forecasting. It can help improve model parameterization, inform choices between alternative model structures, design better sensor networks and experiments for data collection, and analyze uncertainty of ecological forecasts.

Forecasting generally means a capability to project future states of a system by modeling the evolution of the system as a function of its state at an initial time. DA is an essential tool to constrain parameters and the initial system state before a model is used to project the evolution of the system. Ecology as a research community has recently explored, examined, and developed a variety of DA techniques to analyze multi-scale ecological data in space and time.

To prepare the community for major research challenges in a data-rich era, we organized eight papers in this Invited Feature to address several issues on DA and ecological forecasting. The paper by Luo et al. offers a perspective on DA, ecological forecasting, and their relationships. Ecological forecasting (or prediction or projection) has been traditionally made using process-oriented models, informed by data in largely ad hoc ways. Most of the simulation models are generally not adequate to quantify real-world dynamics so as to provide reliable forecasts. DA uses data to constrain initial conditions and model parameters to yield simulations that approximate reality as closely as possible. Models conditioned upon the best information from both process-oriented thinking and empirical knowledge should generate the best forecasts. However, forecasts cannot be improved by DA when ecological processes are not well understood or never observed.

LaDeau et al. use four case studies (Severe Acute Respiratory Syndrome [SARS], Dengue Fever, Lyme, and West Nile virus) to demonstrate that advances in disease forecasting require better understanding of the zoonotic host and vector ecology that support pathogen amplification and disease spillover into humans. The authors identify cases where additional data, greater biological understanding, and coherent treatment of spatiotemporal variability could substantially improve forecasts of disease dynamics. To do so, DA is required in a hierarchical state-space framework to (1) integrate multiple data sources across spatial scales to estimate latent parameters, (2) partition uncertainty in process and observation models, and (3) explicitly build upon existing ecological and epidemiological understanding.

The Ensemble Kalman Filter (EnKF) has been successfully used in weather forecasting to assimilate observations into models. Gao et al. examine how effectively EnKF can improve forecasts of carbon sequestration. Eight data sets from Duke Forest between 1996 and 2004 were assimilated into a terrestrial ecosystem model, which was then used to forecast changes in carbon pools from 2004 to 2012. Parameter uncertainties decreased as data were sequentially assimilated into the model. Uncertainties in forecast carbon sinks increased over time for the long-term carbon pools but remained stationary for the short-term pools. EnKF effectively assimilated multiple data sets to constrain parameters, forecast dynamics of state variables, and evaluate uncertainty.

Hill et al. assimilated data from both tower flux and atmospheric profile measurements into a coupled atmosphere–biosphere model to constrain ecosystem processes. Variations in temperature,

[1] Reprints of this 119-page Invited Feature are available for $10.00 each, either as PDF files or as hard copy. Prepayment is required. Order reprints from the Ecological Society of America, Attention: Reprint Department, 1990 M Street, N.W., Suite 700, Washington, D.C. 20036 (esaHQ@esa.org).

CO_2 concentration, and mixing ratio along profiles of the planetary boundary layer reflect dynamics of both ecosystem and atmospheric processes at a large spatial scale while eddy-flux measurements contain more direct information about ecosystems states and processes in a small area. The authors used both forward and inverse modeling approaches to demonstrate consistency between the two sets of data, which together enhance constraints of model parameters.

Model and data may contribute different amounts of information to constraints of ecological forecasting. Weng and Luo developed a Shannon information index to evaluate the relative information contributions of model vs. data to constraints of forecasts of carbon sequestration. The eight data sets from Duke Forest contributed more information than the model to constrain carbon dynamics in foliage and fine root pools over the 100-year forecasts. The model, however, contributed more than the data sets to constrain the litter, fast soil organic matter (SOM), and passive SOM pools. For woody biomass and slow SOM, the model contributed less information than the data in the first few decades and then more in the following decades.

Spadavecchia et al. partitioned sources of uncertainty in regional-scale forecasting into driving variables and parameters. Data from a ponderosa pine forest in central Oregon were used to estimate parameters of an ecosystem model. Meteorological driving variables were generated with an ensemble of geostatistical simulations conditioned on observations. Their evaluation shows that the relative contribution to forecasting uncertainty was much larger from driving variables than parameters in areas with sparse meteorological stations, but the opposite was true for areas with at least one meteorological station near the study site. Thus, when climate forecast data are used as drivers, they potentially cause large uncertainty in ecological forecasts in areas with sparse weather records.

Clark et al. demonstrate that Bayesian inference can potentially help improve network design for data collection. In an example of wireless sensor networks, the value of an observation is weighed against the cost of data collection by inferential ecosystem models. An observation is evaluated according to its contribution to estimates of state variables and parameters. Such a contribution varies with processes to be studied, the current states of the system, the uncertainty about those states, and the perceived potential for rapid change.

Training the next generation of ecologists is crucial to support ecological forecasting by NEON. Hobbs and Ogle outline a curriculum for an introductory course on DA and ecological forecasting. Key elements for a successful course include models as routes to insight, uncertainty analysis, basic probability theory, hierarchical models, likelihood and Bayes, computational methods, research design, and problem solving. Training with these combined elements can help ecologists of the future master principles of mathematics and statistics and allow them to create models for analysis that are appropriate for a diverse range of ecological problems, including those that involve complex interactions operating at large scales.

Research on DA and ecological forecasting is at an infant stage. Most studies on DA and forecasting use simple models and are applied to confined ecological systems. We have not approached the data and models from a complex systems perspective. In nonlinear dynamic systems, errors may grow as a result of errors in initial conditions and parameters. If systems are chaotic, state changes become imminent. Future research should include diagnostics of state change in addition to traditional simulations of the evolution of state variables.

—Yiqi Luo
University of Oklahoma
—David S. Schimel
NEON