Enabling understanding and quantitative prediction of complex ecological dynamics through modeling and observation

David Schimel

NEON Inc

Boulder Colorado 80305

Abstract

The design of ecological studies, whatever their size, requires identifying appropriate questions and hypotheses, identifying the information that is needed to address the questions, and then determining the methods to obtain that information. Ecologists are faced with a grand intellectual and practical challenge: understanding how living systems will respond on a human-dominated planet. The changes triggered by human domination of climate, fragmentation and global species distributions challenge the theories, experimental approaches and culture of ecology. How can we enable understanding and forecasting of the impacts of climate change, land use change and invasive species on continental-scale ecology by providing infrastructure and consistent methodologies to support research and education in these areas.

Introduction

Successful design of an ecological study, whatever its size, requires identifying appropriate questions, determining the information that is needed to address the questions, and then determining the methods whereby that information will be obtained (Lindenmayer and Likens 2010). In the new millennium, ecologists are faced with a grand intellectual and practical challenge: understanding how living systems will respond on a human-dominated planet (Turner II et al. 1993). The changes triggered by human domination of climate, fragmentation and global species distributions test the theories, experimental approaches and culture of a field whose theory is deeply rooted in equilibrium concepts and just making the transition to a more dynamic view (Golley 1996, Gunderson et al. 2006). This challenge requires studies that cross levels of biological organization, understand processes at scales larger than traditional field sites, recognize the importance of direct and indirect human influences, and capture the nonlinear and chaotic dynamics of ecological systems. Many of these new ecological questions require answers that integrate over huge areas (Pacala et al. 2001)(Schimel et al 2001), or involve transport of organisms or material over long distances (Chadwick et al. 1999).

Ecological systems as complex dynamical systems

Forecasting the future behavior of ecological systems poses many challenges because of the largest being the complex, contingent and nonlinear behavior of the environment (McNaughton 1983, Hobbs et al. 1991). Ecological systems are known to exhibit nonlinear behavior, possess alternative states and may exhibit complex, nonperiodic behavior and chaotic dynamics (Pastor and Cohen 1991, Scheffer et al. 2001). Complex dynamics were proposed for the atmosphere by Ed Lorenz (1963) and demonstrated in ecological models by May (2001). Ecological systems are strongly driven by the deterministic chaos of the geophysical system, while also being influenced and constrained by the complex structure of the evolutionary landscape (Kauffman and Johnsen 1991). Ecological systems are subject to multiple stresses (Field et al. 2007) and so the number of variables whose interactions can produce complex dynamics is very high (Schimel et al. 2005)

Prediction is complex systems requires specialized mathematical approaches. In ecology, complex behavior is often probed with experiments, and while experiments are crucial, they are strongly complemented by another emerging paradigm, that of ecological forecasting. Ecological forecasting has some features in common with weather forecasting. In weather forecasting, a mechanistic physical model of the chaotic atmosphere is initialized with observations, integrated forward and then compared to the next set of observations, allowing the model skill score to be computed from the resemblance between the model and observations, averaged over space.

Through the forecast-analysis cycle, forecast models have been evaluated under thousands of regimes (El Nino, La Nina, droughty, humid, dusty, affected by volcanoes, and so on) and experience has been accumulated about its behavior under these conditions. Each prediction step constitutes an evaluations of the hypotheses embodied in the model. While each forecast-analysis cycle does not constitute a definitive test, the discipline of regular forecasting results in accumulating many, in the case of meteorology thousands to millions, of hypothesis tests. This is a vital methodology for problems that occur at scales too large to capture in manipulative experiments. The type of learning and cumulative hypothesis testing of Earth System-scale problems is for the most part unavailable to ecologists, who are usually compelled to generalize from small-scale experiments, and short-term, localized or partial data.

Strategies for observing and forecasting the behavior of the atmosphere are guided by theory. Understanding the mathematical nature of chaotic behavior in the atmosphere has allowed the development of observing and analysis strategies that allow efficient prediction in the presence of chaos (Kalnay 2002, Lewis et al. 2006). Chaotic systems, by definition, are ones where small differences in initial conditions can cause exponential diverging trajectories. Since observations can never be perfect, very small errors in observations can cause large prediction errors. In meteorology this has led to the development of data assimilation techniques where simulated state variables are regularly updated using observations to correct the drift inevitable in a chaotic system. Ecologists have begun to use such approaches (Raupach et al. 2005).

Ecology may also rely on forecasting approaches other than classical prediction. In ecological systems attention has focused also on developing models of qualitative behavior preceding impending regime transitions (grass to shrub, population collapses and so on). Often such transitions are preceded by changes in variability (Scheffer et al 2009). Detecting changes in variability in time and space requires adequate sampling: sampling adequate to observe central tendencies may not be adequate to characterize variability. As will be discussed below,

Manipulative experiments designed to probe the dynamical behavior of ecological systems

Conceptual model for observing systems

The NEON design is strongly guided by conceptual models, in addition to the overarching complex systems perspective. Two basic linked conceptual models have guided NEON design. The first model deals with change in time and space, and the second with integration of information across scales.

Most observing systems are not built around a cause-and-effect model but instead seek to efficiently monitor a small number of either driver or response variables. For example, the NOAA Climate Reference Network (http://www.ncdc.noaa.gov/crn/) measures climate but not impacts. At the other end of the spectrum, the USFWS Breeding Bird Survey (http://www.pwrc.usgs.gov/bbs/) measures impacts in the avifauna, but does not include measurements of any hypothesized causes of change in avian communities (Faaborg et al. 2010). Very few observing systems measure any key indicators of feedbacks and interactions. Quantitative forecasting requires gaining a mechanistic understanding and dictates measuring key drivers of change, parameters of feedbacks and interactions and impacts of changing drivers. As a consequence, observe systems must address a carefully selected set of drivers of change (for example: temperature, incoming solar radiation and precipitation, nutrient deposition, land cover). Measures of interactions and feedbacks include soil moisture, functional gene expression, element stoichiometry, species changes in multiple taxa, parasite and disease burden and isotopic indicators of metabolism. Measures of impacts include species composition and diversity, invasive species abundance, primary productivity, and phenology. Of course some measurements may fit into more than one category: invasive species can be both a cause of and a response to environmental changes.

A second conceptual model focuses on how processes at different spatial scales interact. Fundamental processes in the environment occur at multiple scales. Metabolism and physiology occur in cells and organisms. Behavioral and trophic interactions occur between organisms at somewhat larger scales, up to the continental scales of migratory birds and Lepidoptera, and long-range movements of invasive species. Ecohydrology and biogeochemistry play out at these larger scales through atmospheric, hillslope and riverine transport of gases, water and suspended or dissolved substances. Coherent patters of climate variability (for example, the El Nino) and change affect huge regions in spatial patterns dictated by global processes (Wang and Schimel 2003). Multiple scales of process in the biosphere produce the stocks and fluxes and patterns of matter and patterns of abundance and diversity that we observe in nature.

An Observatory may observe nature using either site-based (stream reach, plot and tower) or spatial (remote sensing) techniques from which we must reconstruct the multiple scales and scale interactions that occur in nature. Critically, spatially extensive measurements must constrain a critical process, measured at local scales, so that process models can be extrapolated in space. Researchers can bring intensive experimental or process study data together with remote sensing (and other continental survey data, such as the Forest Inventory Analysis) using mathematical or statistical models to reconstruct the underlying organismal and ecological processes.

Sampling spatio-temporal variability: the roles of site selection and technology.

Sampling ecological processes over large regions is a long-standing challenge to ecology. Ecosystems, as dynamical systems, respond differently in different regimes (eco-climatic conditions) and so one of the main goals driving the design is to observe as much variability in conditions as possible.

Statistically efficient sampling can only go so far to address ecological heterogeneity. Ecological extrapolation is almost always limited by a paucity of appropriate spatially extensive data. To complement the stratified site selection, some long-standing issues of spatial extrapolation must be addressed using innovative technology. Airborne remote sensing can bridge from the organismal to the regional scale by providing measures of canopy nutrient content, leaf area, chlorophyll and photosynthetic capacity, total biomass and canopy height at an individual tree scale (1-2 m ground resolution) for all trees or shrubs within a 200-400 km^2 km region. Current (or in design) systems can observe structural and chemical measurements for 1 to 30 million trees per scene, and could be repeated annually. To put this number in perspective, 5 to 10 scenes will sample approximately the same number of individual trees as the entire US Forest Inventory Analysis program. Local and remotely sensed information can be combined using relatively recent statistical techniques to provide data products with quantitative estimates surfaces and corresponding uncertainty (Huang et al. 2002, Johannesson and Cressie 2004, Tzeng et al. 2005, Wikle and Berliner 2005, Cressie and Johannesson 2008).

In addition to a spatial sampling approach, researchers also needed guidelines for temporal frequency of sampling. We analyzed the requirements for detecting decadal and longer trends. In analyzing this, consider two requirements for an observatory. First, can the Observatory *detect* trends? For example, could we detect a systematic change over time, even if there was noise in the observations and measurement error? This type of analysis is fairly conventional and is described in Schimel et al (2009). Second, and more challenging, could we *attribute* observed trends to specific drivers? In other words, could we link *drivers* and *impacts* to help researchers establish hypotheses about causation? Could we measure or infer model parameters for the interactions and feedbacks that linked the drivers to the responses? This question may be addressed via simulation. These simulations can build on the the driver-interaction/feedback-impact conceptual model .

Ecological trends can be modeled with five main components:

1. The magnitude of the trend.
2. The intrinsic variability of the trend
3. The relationship between the forcing and the response. The response may be more or less sensitive, and the form of the response may be linear or nonlinear, and may vary in space and as a function of other variables.
4. The error of the measurement. This includes the accuracy and precision of the measurement technique and adequacy of sampling in time and space.
5. The number of measurement locations (replication) and how correlated drivers or responses are between locations.

By simulating responses under varying levels of the factors above that influence observations, potential opportunities and weaknesses can be identified. We made assumptions about the magnitude of trends, amount of interannual variability and structure of spatial correlations among locations and assessed the network sensitivity using annual time-scale information because quantifying long-term changes is a fundamental requirement. Within the network of expected ranges for magnitude of trend, interannual variability, and correlation among locations, we analyzed simulation results for bounding levels of measurement error. In this case, measurement error includes instrumental or observer accuracy and precision, sampling or representativeness error and errors associated with data processing algorithms.

We simulated the relationship between a hypothesized forcing and an ecological response, and again created simulations to test the ability of a spatially-distibuted network to 1) detect a trend in an ecological response, 2) identify whether the relationship between forcing and response is linear or nonlinear, and 3) determine the ability of the network to estimate the parameters of the relationship between forcing and response (eg, for *response* = a□□□b $e^{kforcing}$ where a□□b and k are parameters). These analyses are substantially more complex. The results are encouraging for the ability of the network to detect and determine the form of complex non-linear relationships, or to *attribute* changes to specific processes (Duffy et al submitted). Quantitatively retrieving the parameters of ecological relationships is challenging and highlights the need for process studies and experiments linked to time-series observations.

Summary

While theory and observational science are often separate, or linked in the analysis phase, addressing the grand challenges of ecological sustainability requires a new approach to designing observations, on the one hand, and on the other hand, designing theory and models that reflect the nature of the actual (not idealized) system. This is because simulation and prediction in complex systems requires integrated mathematical and observing approaches. The observing systems must address the key driver, process and outcome variables, and the models must recognize the impact of uncertainty in observations on the predictability of the system. Moving towards a future where the predictability of ecological systems (what can be predicted, how well can it be predicted, what information is needed for a prediction, what is the intrinsic time horizon of the prediction) is understood will require a new partnership between mathematicians, ecological theoreticians and modelers and observational scientists.

References cited

Brock, W. A. and S. R. Carpenter. 2006. Variance as a leading indicator of regime shift in ecosystem services. Ecology and Society **11**:9 [online].

Chadwick, O. A., L. A. Derry, P. M. Vitousek, B. J. Huebert, and L. O. Hedin. 1999. Changing sources of nutrients during four million years of ecosystem development. Nature **397**.

Cressie, N. and G. Johannesson. 2008. Fixed rank kriging for very large spatial data sets. Journal of the Royal Statistical Society: Series b (Statistical Methodology) **70**:209-226.

Faaborg, J. R., R. T. Holmes, A. D. Anders, K. L. Bildstein, K. M. Dugger, S. A. Gauthreaux Jr., P. Heglund, K. A. Hobson, A. E. Jahn, D. H. Johnson, S. C. Latta, D. J. Levey, P. P. Marra, C. L. Merkord, E. Nol, S. I. Rothstein, T. W. Sherry, T. S. Sillett, F. R. Thompson III, and N. Warnock. 2010. Recent advances in understanding migration systems of New World land birds. Ecological Monographs **80**:3-48.

Field, C. B., D. B. Lobell, H. A. Peters, and N. R. Chiariello. 2007. Feedbacks of terrestrial ecosystems to climate change. Annual Review of Environment and Resources **32**:1-29.

Golley, F. B. 1996. A history of the ecosystem concept in ecology: More than the sum of the parts. Vail-Ballou Press, Binghamton, New York, USA.

Gunderson, L. H., C. R. Allen, and C. S. Holling. 2006. Foundations of ecological resilience. Island Press, Washington, DC, USA.

Hargrove, W. W. and F. M. Hoffman. 1999. Using multivariate clustering to characterize ecoregion borders. Computers in Science and Engineering: Special Issue on Scientific Visualization of Massive Data Sets **1**:18-25.

Hargrove, W. W. and F. M. Hoffman. 2004. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environmental Management **34**:S39-S60.

Hobbs, N. T., D. S. Schimel, C. E. Owensby, and D. S. Ojima. 1991. Fire and grazing in the tallgrass prairie: Contingent effects on nitrogen budgets. Ecology **72**:1374-1382.

Huang, H.-C., N. Cressie, and J. Gabrosek. 2002. Fast, resolution-consistent spatial prediction of global processes from satellite data. Journal of Computational and Graphical Statistics **11**:63-88.

Johannesson, G. and N. Cressie. 2004. Finding large-scale trends in massive, global, environmental datasets. Environmetrics **15**:1-44.

Kalnay, E. 2002. Atmospheric modeling: Data assimilation and predictability. Cambridge University Press, New York, New York, USA.

Kauffman, S. A. and S. Johnsen. 1991. Coevolution to the edge of chaos: Coupled fitness landscapes, poised states, and coevolutionary avalanches. Journal of Theoretical Biology **149**:467-505.

Keller, M., D. S. Schimel, W. W. Hargrove, and F. M. Hoffman. 2008. A continental strategy for the National Ecological Observatory Network. Frontiers in Ecology and the Environment **6**:282-284.

Lewis, J. M., S. Lakshmivarahan, and S. K. Dhall. 2006. Dynamic data assimilation: A least squares approach. Cambridge University Press, New York, New York, USA.

Lindenmayer, D. B. and G. E. Likens. 2010. Adaptive monitoring: A new paradigm for long-term research and monitoring. Trends in Ecology & Evolution **24**:482-486.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences **20**.

Lowman, M., C. D'Avanzo, and C. Brewer. 2009. A national ecological network for research and science. Science **323**:1172-1173.

May, R. M. 2001. Stability and complexity in model ecosystems. Princeton University Press, Princeton, New Jersey, USA.

McNaughton, S. J. 1983. Serengeti grassland ecology: The role of composite environmental factors and contingency in community organization. Ecological Monographs **53**:291-320.

National-Research-Council. 2003. NEON: Addressing the nation's environmental challenges. National Research Council (NRC), Washington, DC, USA.

Pacala, S. W., G. C. Hurtt, D. Baker, P. Peylin, R. A. Houghton, R. A. Birdsey, L. Heath, E. T. Sundquist, R. F. Stallard, P. Ciasis, P. Moorcroft, J. P. Caspersen, E. Shevliakova, B. Moore, G. Kohlmaier, E. Holland, M. Gloor, M. E. Harmon, S.-M. Fan, J. L. Sarmiento, C. L. Goodale, D. S. Schimel, and C. B. Field. 2001. Consistent land- and atmosphere based U. S. carbon sink estimates. Science **292**:2316-2320.

Pastor, J. and Y. Cohen. 1991. The responses of a forest model to serial correlations of global warming. Ecology **72**:1161-1165.

Raupach, M. R., P. J. Rayner, D. J. Barrett, R. S. DeFries, M. Heimann, D. S. Ojima, S. Quegan, and C. C. Schmullius. 2005. Model data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertaintly specifications. Global Change Biology **11**:378-397.

Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. Nature **413**:591-596.

Scheffer, M. and S. R. Carpenter. 2003. Catastrophic regime shifts in ecosystems: Linking theory to observation. Trends in Ecology & Evolution **18**:648-656.

Schimel, D., G. Churkina, B. H. Braswell, and J. Trenbath. 2005. Remembrance of weather past: Ecosystem responses to climate variability. Pages 350-369 *in* J. R. Ehleringer, T. E. Cerling, and M. D. Dearing, editors. A history of atmospheric CO2 and its effects on plants, animals, and ecosystems. Springer, New York, New York, USA.

Schimel, D., M. Keller, and P. Duffy. 2009. The NEON Strategy: Enabling continental scale ecological forecasting. NEON, Inc., Boulder, CO, USA.

Schimel, D. S., VEMAP-Participants, and B. H. Braswell. 1997. Continental scale variability in ecosystem processes: Models, data, and the role of disturbance. Ecological Monographs **67**:251-271.

Slavik, K., B. J. Peterson, L. A. Deegan, W. B. Bowden, A. E. Hershey, and J. E. Hobbie. 2004. Long-term responses of the Kuparuk river ecosystem to phosphorus fertilization. Ecology **85**:939-954.

The-National-Science-Foundation. 2009. Large Facilities Manual, NSF 10-012. The National Science Foundation (NSF), Arlington, Virginia, USA.

Turner II, B. L., W. C. Clark, R. W. Kates, J. F. Richards, J. T. Mathews, and W. B. Meyer. 1993. The Earth as transformed by human action: Global and regional changes in the biosphere over the past 300 years. Cambridge University Press, New York, New York, USA.

Tzeng, S., H.-C. Huang, and N. Cressie. 2005. A fast, optimal spatial-prediction method for massive datasets. Journal of the American Statistical Association **100**:1343-1357.

Wang, G. and D. Schimel. 2003. Climate change, climate modes, and climate impacts. Annual Review of Environment and Resources **28**:1-28.

Wikle, C. K. and L. M. Berliner. 2005. Combining information across spatial scales. Technometrics **47**:80-91.