Ecological Applications, 21(5), 2011, pp. 1429–1442

Ecological forecasting and data assimilation in a data-rich era

Yiqi Luo,[1,7] Kiona Ogle,[2,3] Colin Tucker,[2] Shenfeng Fei,[1] Chao Gao,[1] Shannon LaDeau,[4] James S. Clark,[5] and David S. Schimel[6]

[1]*Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma 73019 USA*
[2]*Department of Botany, University of Wyoming, Laramie, Wyoming 82071 USA*
[3]*Department of Statistics, University of Wyoming, Laramie, Wyoming 82071 USA*
[4]*Cary Institute of Ecosystem Studies, Millbrook, New York 12545 USA*
[5]*Department of Biology, Duke University, Durham, North Carolina 27708 USA*
[6]*NEON, Boulder, Colorado 80301 USA*

Abstract. Several forces are converging to transform ecological research and increase its emphasis on quantitative forecasting. These forces include (1) dramatically increased volumes of data from observational and experimental networks, (2) increases in computational power, (3) advances in ecological models and related statistical and optimization methodologies, and most importantly, (4) societal needs to develop better strategies for natural resource management in a world of ongoing global change. Traditionally, ecological forecasting has been based on process-oriented models, informed by data in largely ad hoc ways. Although most ecological models incorporate some representation of mechanistic processes, today's models are generally not adequate to quantify real-world dynamics and provide reliable forecasts with accompanying estimates of uncertainty. A key tool to improve ecological forecasting and estimates of uncertainty is data assimilation (DA), which uses data to inform initial conditions and model parameters, thereby constraining a model during simulation to yield results that approximate reality as closely as possible.

This paper discusses the meaning and history of DA in ecological research and highlights its role in refining inference and generating forecasts. DA can advance ecological forecasting by (1) improving estimates of model parameters and state variables, (2) facilitating selection of alternative model structures, and (3) quantifying uncertainties arising from observations, models, and their interactions. However, DA may not improve forecasts when ecological processes are not well understood or never observed. Overall, we suggest that DA is a key technique for converting raw data into ecologically meaningful products, which is especially important in this era of dramatically increased availability of data from observational and experimental networks.

Key words: *data assimilation; data–model fusion; ecological forecasting; inverse analysis; optimization; predictions; prognosis; projections.*

The Need for Ecological Forecasting

The capability to forecast the impacts of environmental change on our living environment and natural resources is critical to decision making in a world where the past is no longer a clear guide to the future (Clark et al. 2001). We are living in a period marked by rapid climate change (Solomon et al. 2007), profound alteration of biogeochemical cycles (Vitousek et al. 1997), unsustainable depletion of natural resources (Heinz Report 2008), proliferation of exotic species (D'Antonio and Vitousek 1992, Liao et al. 2008) and infectious disease (Smith et al. 2005), and deterioration of air and water quality (Gleick 2002, Akimoto 2003). Human populations are increasing at an alarming rate, and society is dependent on the extraction and utilization of natural resources to support regional and global economies. Predictable and increasing supplies of energy, food, fiber, freshwater, and clean air are necessary to maintain healthy human societies. To effectively mitigate and adapt to climate change, we need to develop robust methods to apply data and current knowledge to the problem of anticipating future states of ecosystems and then to assess resilience and, potentially, collapse of ecosystem services.

Nascent ecological forecast models are in use in some areas. For example, ecosystem and biogeochemical cycling models have been incorporated into earth-system models to project terrestrial carbon sinks and sources and their feedback to climate change in the 21st century (Cox et al. 2000, Friedlingstein et al. 2006). Those model predictions have been incorporated into the assessment

Manuscript received 15 July 2009; revised 19 July 2010; accepted 4 August 2010. Corresponding Editor: S. K. Collinge. For reprints of this Invited Feature, see footnote 1, p. 1427.
[7] E-mail: yluo@ou.edu

TABLE 1. Distinguished characteristics of terms: forecasting, prediction, projection, and prognosis

Term	Characteristics
Forecasting	probabilistic statement on future states of an ecological system after data are assimilated into a model
Prediction	future states of an ecological system based on logical consequences of model structure
Projection	future states of an ecological system conditioned upon scenarios
Prognosis	subjective judgment of future states of an ecological system

reports of the Intergovernmental Panel on Climate Change (IPCC) to guide mitigation efforts by governments and public (Solomon et al. 2007). At local and regional scales, ecosystem models have been used to forecast changes in natural resources for improved management (Hood et al. 2007, Kirilenko et al. 2007). Ecological models have also played a role in forecasting the timing and intensity of infectious diseases and in the development of control strategies (Smith et al. 2005). Inference from these types of models has led to successes in early warning of disease risk and vaccination strategies (Glass et al. 2000, Ferguson et al. 2001, Keeling et al. 2001, 2003, Smith et al. 2005, Chaves and Pascual 2007). Although forecasting has happened in some areas of ecology, it is generally done without formally integrating data to constantly improve models and assess uncertainties of their forecasts.

With the advent of new measurement techniques and observatory networks over the past decades, experimental and observational data are increasingly abundant, offering a tremendous potential to improve ecological forecasting. For example, FLUXNET is a worldwide network with over 400 tower sites operating on a long-term and continuous basis to measure ecosystem gas exchange, supplemented with data on vegetation, soil, hydrologic, and meteorological characteristics at the tower sites (Baldocchi et al. 2001). Satellite observations provide remote measurements of climate, ocean circulation, terrestrial vegetation and phytoplankton, and hydrology at multiple scales, which can be used to inform ecological models about large-scale processes. In response to the need for long-term data on ecological responses to changes in land use, biological invasions and climate, the U.S. National Science Foundation (NSF) is establishing a National Ecological Observatory Network (NEON), which is a continental-scale research platform to gather such data. However, the ultimate value of the diverse and abundant data will depend on how well the data can be integrated with the best available understanding of biological processes.

Integrative analysis that can combine data sources into models that explicitly acknowledge sources of uncertainty will be critical to advances in ecological forecasting (Clark et al. 2004, 2007, 2010, Weng and Luo 2011). This paper examines the potential to advance ecological forecasting by combining data with models using data assimilation (DA) techniques. While it has been widely used in other scientific disciplines (Evensen 2007), DA has not been often employed in ecology. Traditionally, DA uses data to constrain a model during simulation to yield results that approximate reality as closely as possible before it is applied to forecast the most likely future state of an ecosystem. This is a more stringent set of requirements than applied to models for producing plausible futures generally consistent with theory. By analogy with weather forecasting, running an atmospheric model many times, initialized with average summer conditions, will produce a number of realizations, but only by chance would one (or the average) resemble the actual conditions on a particular day. Initializing the same model with a best estimate of today's weather should produce a better forecast of the next day's weather than the former experiment. Ecological models have generally been used in the former mode, and it is our contention that today's environmental management issues also require the latter capability, to forecast actual ecological futures and their likelihood.

In this paper, we first review the definition of ecological forecasting to illustrate key approaches and major elements of forecasting. We then offer a historical perspective on uses of data, model, and their integration toward a predictive understanding of ecological systems. Research methods have evolved from simple theoretical models to a point nowadays when improvement of complex simulation models using advanced optimization tools makes ecological forecasting a realizable goal. We also evaluate the current uses of simulation models for ecological forecasting and prediction. These models are useful but, by themselves, not adequate to provide realistic forecasts. The models have to be iteratively improved against data using DA techniques before they can be effectively used for forecasting. We briefly describe DA techniques and explore their applications to model improvement and ecological forecasting. Although DA is promising to improve ecological forecasting, ecologists are just beginning to use it in research. In the last section, we offer a vision on future research opportunities and challenges as we enter a data-rich era.

DEFINITION OF ECOLOGICAL FORECASTING

Ecological forecasting has several synonyms, including prediction, projection, and prognosis (Table 1), all of which are used to describe a process of estimating future unknown situations. Prediction may be viewed as a more general term to anticipate that something is likely to happen in the future, sometimes under given conditions, and implies a quantitative result. It is important to recognize that prediction is not limited to

the future but could be generated for data already observed (e.g., cross-validation), observations that could not be obtained (e.g., gap-filling), and current states that have not been observed, but can be predicted (e.g., kriging; Clark et al. 2011). Prediction, even if not accurate, provides a formal mechanism for evaluating the state of current knowledge as reflected by both models and data (Clark et al. 2001). Projection is an evaluation of future states of an ecological system in response to changes in key driving forces under different scenarios, which are based on assumptions concerning future socioeconomic and technological developments that may or may not be realized and are therefore subject to substantial uncertainty. Prognosis or prognostic analysis has been used for environmental assessments with an emphasis on a scientist's view about the state of an ecosystem and is generally more subjective. These terms are often used interchangeably to express the idea of estimating a future state of the ecological system. Here we focus primarily on the concept of forecasting.

We refer to forecasting as the process of predicting some future event or condition usually as a result of study and analysis of available pertinent data. Weather is the most familiar format for forecasting. Weather forecasting attempts to predict the state of a weather system for a future time (typically short term of 10 days or less) at a given location. The weather is a chaotic system and reliable forecasts require accurate measurements of initial conditions and a model representing atmospheric processes. The improvement of skill (i.e., the performance of a given model relative to some baseline) in the weather forecast (Fig. 1) has resulted from advances in both models and observations of the current state of the atmosphere and, critically, requires their integration. In contrast, climate models are used to explore the potential response of the atmosphere to changing forcing, usually via the accumulation of greenhouse gases and aerosols over time. In climate simulations, the boundary conditions (energy trapped in the atmosphere) are varied and potential responses modeled. The initial conditions are usually only loosely based on observations, and simulations often begin from climatologic or equilibrium states. Such forward simulations are not expected to accurately predict the future, but rather are viewed as sensitivity or scenario analyses to assumed future boundary conditions.

Ecological forecasting, in particular, has been described as the projection of future states "of ecosystems, ecosystem services, and natural capital, with fully specified uncertainties, and is contingent on explicit scenarios for climate, land use, human population, technologies, and economic activity" (Clark et al. 2001). We define two types of ecological forecasting: (1) classic prediction, asking the question, "What is the most likely future state of an ecological system?" (2) What-if analysis, asking, "What is the most likely future state of a system, given a decision today?" In both cases, results are often represented as a probability distribution, conditional on the many sources of uncertainty (model, observations, and so on). Classic prediction may be applied to fast-evolving systems whose dynamics are strongly governed by its own current state (for example, forecasting the spread of an infectious disease), whereas the what-if analysis comes into play when alternate management actions or scenarios are being considered (for example, forecasting the likely impacts of alternate forest fire risk mitigation practices on biodiversity or studying alternate climate change scenarios). In the latter case, the system boundaries as defined by explicit scenarios of climate, land use, human population, technology, and economic activity, which are critical to understanding ecological changes over longer time scales. Both types of forecasting have to quantify the past and current states of ecological systems as a starting point and use models to project the future dynamics.

Thus, the model is a key component for producing ecological forecasts. Of course, no model is a perfect representation of a system, but the derived equations and parameters that integrate reasoning (theory) and information (data) are only approximately correct in certain situations. The power of ecological models for deductive inference is usually limited partly because ecological responses may depend on the current state of

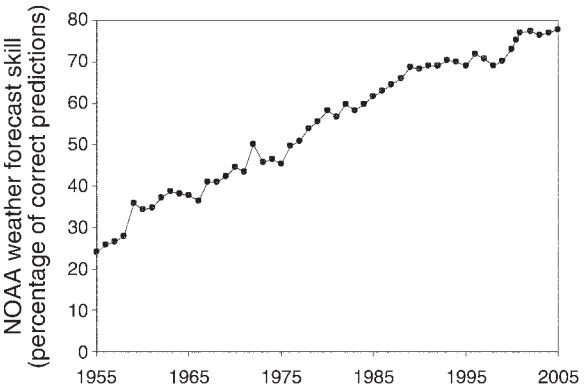

FIG. 1. Improvement in the skill of the NOAA weather forecast from 1955 to 2005 as measured by predicting the atmospheric pressure field 36 hours in advance in the U.S. operational forecast model. The overall trend is surprisingly steady given the changes to satellite observations, computing power, and advances in knowledge over the decades between the 1950s and 2000s. This highlights the need to understand that quantitative models cannot be evaluated in a binary fashion (right vs. wrong). Errors must be measured, assigned to weaknesses in theory, simulations, or observations (or some of each) and targeted efforts made to correct the identified problems. Analogously, the skill of ecological forecasts will evolve over time as fundamental theory advances, as techniques for estimation of states and parameters improve, and as system behavior is observed under a wider and wider range of conditions to characterize more parameter space. The figure is modified from the following NOAA source: ⟨http://www.vos.noaa.gov/MWL/dec_07/Images/Figure3-WeatherPrediction.jpg⟩.

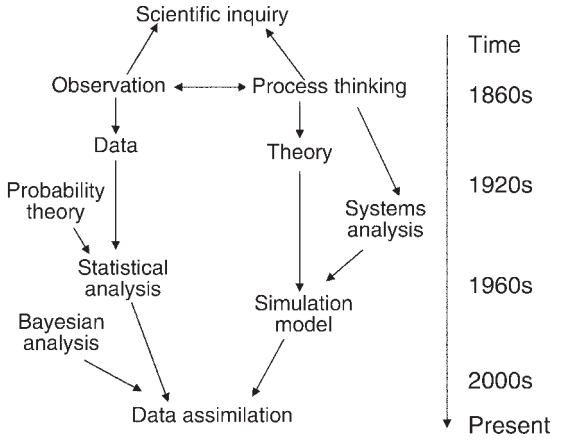

FIG. 2. Evolution of various research approaches to scientific inquiry, all converging toward data assimilation (DA) to improve predictive understanding of ecological systems. Observation from natural ecosystems and experiments generates data, which were not rigorously analyzed until the 1960s using frequentist methods (Fisher 1959). Process thinking is a mental activity to figure out processes that operate in a system and is formally expressed in a model and theory. Process thinking, assisted by systems analysis, has led to the development of simulation models. DA combines process thinking in the form of numerical models with data that record the states of a system, usually under a Bayesian framework, to improve model forecasts.

the system, as evidenced by empirical studies. For example, Knapp and Smith (2001) showed that productivity responds conditionally to rainfall variability depending on the mean value of rainfall at a site. Temperature sensitivity of soil respiration depends on temperature, soil moisture, and substrate supply because different microbial populations are activated under different ecosystem conditions (Luo and Zhou 2006, Monson et al. 2006). Thus, simple extrapolation of such ecosystem processes using models often fails if such state-dependent mechanisms are not incorporated into the model.

While no amount of data can constrain model responses to conditions that have never been observed, comparing models and observations over a wide range of conditions increases the chance of capturing important nonlinearities and complex or contingent responses that may control future behavior. This notion of comparing models to data is fundamental to DA, which uses the model structure and parameters as prior information to represent the state of knowledge. DA improves models and their forecasts using information contained in data on the past and current states of an ecosystem through rules of probability (e.g., a Metropolis criterion) to obtain the posterior probability distributions of targeted parameters and forecasted state variables. The posterior probability density functions quantify the potential ranges and uncertainties of parameters and future states of an ecosystem.

HISTORICAL PERSPECTIVE

Two foundational approaches to scientific inquiry are observation and process thinking (Fig. 2; Ogle 2009). Observations record as data the states of ecosystems at the time when the measurements were made. The data also contain information about underlying processes, which can be revealed by quantitative analyses. Process thinking is a mental activity to figure out which processes are operating in a system. Formal expression of outcomes from process thinking occurs via development of a theory and/or model. Thus, theory and/or model represent our theoretical and/or conceptual understanding of the processes that operate in a system. A model should be continuously evaluated and improved as we advance our understanding based on accumulated empirical observations of the real system. The difference between theory and empiricism was summarized as: "Theory delineates possibilities. Empirical studies discriminate the actualities." (May 1981). Integration of models with observations using formal estimation methods rather than ad hoc tuning is a powerful new way of combining theory and empiricism (Ogle 2009). This is where DA methods become important.

Historically, the two aforementioned approaches have mainly been combined into relatively simple systems that could be described by classic models like the logistic growth equation (Verhulst 1838) or competition and prey–predator models (Lotka 1924, Volterra 1926). These models can generalize measurements of systems in some special cases and, within limits, can be used for extrapolation (Pascual and Kareiva 1996). The models possess only a few parameters and can be tested and reconstructed to provide valuable insight into system dynamics and reveal general features of model behavior. Such models, however, usually do not incorporate enough processes to realistically describe behavior of complex ecological systems and thus have a limited capability for quantitative extrapolation or forecasting.

Since the 1950s, process thinking has been facilitated by systems analysis for ecological research. Systems analysis focuses on understanding the behavior of a system as a whole. Process thinking and systems analysis are the foundation for development of the simulation modeling approach (Forrester 1961). Since the 1960s, the simulation modeling approach has been applied to ecological research mainly for (1) integration of process knowledge and data, (2) analyzing the potential behavior of ecological systems under changing conditions (climate) or stress, (3) hypothesis generation, and (4) resource management and policy development.

Simulation models have been widely applied in ecology and resource management in modes of ecological predictions, projections, forecasting, and/or prognostic analysis (Parton et al. 1987, Pacala et al. 1993, Ribbens et al. 1994). Papers on modeling from ISI's Web of Science exponentially increased from the early 1970s to approximately 600 per year for ecological

prediction(s), 100 per year for ecological forecast(s) or forecasting, and 60 for ecological projection(s) in 2008 (Fig. 3). Although many papers refer to prediction, projection, and forecasting, ecological modeling has mainly been concerned with gaining a quantitative understanding of ecological processes and only secondarily with projection or prediction. As a result, ecologists have combined models and data mainly during model development, using data to help distinguish between alternate model structures and to quantify parameter values. To the extent that models are used for forecasting today, applications are normally in a research or proof of concept mode, involve limited verification and analysis, rarely address all of the major sources of uncertainty, and make limited use of observations to reduce these uncertainties.

As predictive, quantitative understanding becomes a more important goal, modeling practices in ecology are evolving. Especially with many advanced mathematical and statistical tools available for data–model fusion, improvement of models using DA techniques makes ecological forecasting a realizable goal. Clark et al. (2001) articulated the need and feasibility of ecological forecasting and its roles in decision-making processes. More papers have been published to address uncertainty issues in ecological inference and forecasting (e.g., Clark et al. 2003). We are entering an initial period of research on a variety of issues related to ecological forecasting, and here we discuss the important role of DA in improving ecological forecasting by integrating diverse data sources and process-based models.

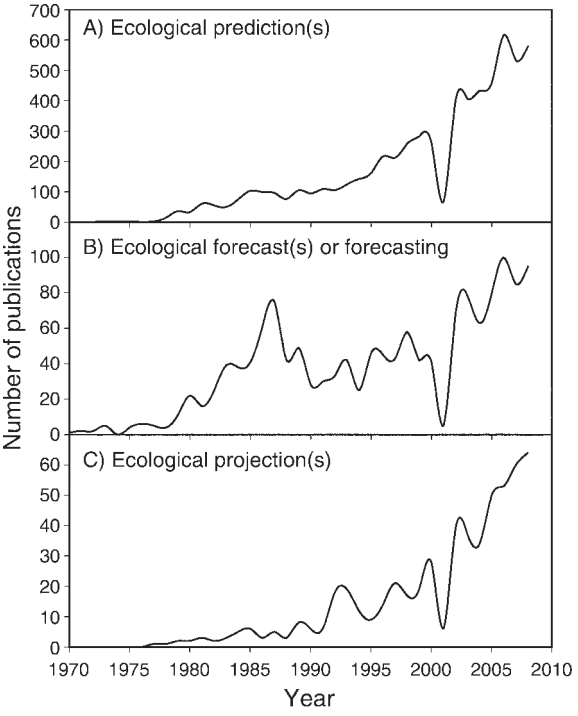

FIG. 3. Numbers of publications that can be identified by (A) ecological predictions, (B) ecological forecasts or forecasting, or (C) ecological projections in the Web of Science.

USES OF MODELS FOR ECOLOGICAL FORECASTING

To refine our understanding on the current status of ecological forecasting in the context of future climate or global change scenarios, we searched ISI's Web of Science to locate articles in the top ecology journals (*Ecological Applications*, *Ecological Monographs*, *Ecology*, *Ecology Letters*, *Ecosystems*, *Functional Ecology*, *Global Change Biology*, *Journal of Applied Ecology*, *Journal of Ecology*, *New Phytologist*, *Oecologia*, *Oikos*, *American Naturalist*, *Global Ecology and Biogeography*, *Frontiers in Ecology and the Environment*, *Conservation Biology*, *Wildlife Monographs*, *Journal of Animal Ecology*, *Journal of Biogeography*, and *Biological Conservation*) using the search string: (model* AND (forecast* OR predict* OR project*) AND ("climate change" OR "global change")). Out of 840 hits produced, we narrowed the list to articles with titles containing (forecast* OR predict* OR project*), which produced 129 hits (search conducted 4–7 May 2009). Of these 129 articles, 63 made predictions under different climate or global change scenarios using a particular model. We reviewed each of the 63 articles to determine (1) the types of models that they employed (e.g., empirical such as regression-type models vs. mechanistic or process-based models), (2) the manner in which time is incorporated (e.g., explicit such as in difference or differential equation models vs. implicit via, for example, driving variables that change over time vs. not included), and (3) methods for parameterizing the models (e.g., fixed parameter values vs. parameters are tuned so that some aspect of the model output "matched" data vs. parameters estimated via rigorous statistical methods for fitting the model to data). The third criteria allowed us to evaluate the current role that DA methods play in ecological forecasting.

The articles we reviewed could generally be divided into two categories, those focused on species distributions and richness (35/63) and those focused on ecosystem processes (21/63). A small number of articles dealt with physiological and other processes. The majority of the articles (56%) focused on forecasting (or predicting or projecting) changes in native species distributions and ranges or overall species richness/diversity under future climate scenarios. Other studies produced forecasts of soil or ecosystem carbon stocks and fluxes (~14%), habitat availability (~6%), biotic invasions (~5%), animal physiological responses to changing climate (~5%), dispersal processes (~5%), ecosystem-level phenology, carrion availability to scavengers, nitrogen cycling, and hydrology.

The models most commonly used for producing forecasts of species distributions can be classified, in general, as bioclimatic envelope models (BEM), or more broadly as ecological niche models. Of these, most were categorized as "empirical" (i.e., regression-based models), time was not explicitly included (i.e., the forecasts

are steady-state distributions based on future environmental conditions generated by other models), explicit DA methods were rarely used, and the majority produced point forecasts (i.e., a single species map for a given future scenario without quantified uncertainty). The BEMs were typically parameterized using a variety of statistical methods to correlate known species' distributions with observed climate variables (e.g., see Beaumont et al. 2005, Kearney and Porter 2009). Most often, uncertainty in predictions is generated by using different future climate scenarios or by comparing different algorithms for estimating the correlation between species distributions and climate variables (e.g., see Currie 2001, Hijmans and Graham 2006, Prasad et al. 2006, Beaumont et al. 2007). Some recent studies, however, have incorporated more mechanisms into these types of models, such as dispersal, regeneration, disturbance, and physiological limitations (e.g., Morin et al. 2008, Kearney and Porter 2009, Vallecillo et al. 2009), that produce time-varying models, which are particularly amendable to DA approaches.

Models used for producing forecasts of ecosystem responses were generally process-based dynamic global vegetation (DGVM; e.g., Beerling et al. 1997, Foley et al. 1998) or biogeochemical models (e.g., Clein et al. 2000). These models explicitly include time as a dynamic model element, typically based on difference equations. These models are often parameterized using a combination of literature data and/or experimental data, and ad hoc tuning or calibration methods may be applied to adjust the model parameters so that the model output agrees with empirical data for the system studied. Applications of process-based ecosystem models for forecasting have generally produced point estimates for each output variable at each future time point rather than distributions acknowledging explicit sources of uncertainty (e.g., Beerling et al. 1997, McGuire et al. 2000, Jones et al. 2005, Morin et al. 2008). Occasionally the articles may report the results of ensemble model runs, where each simulation run used different potential future climates. Model outputs may be averaged across ensemble runs to produce the "most likely" forecasts, and the variances of the model outputs may be reported as prediction uncertainty. In such cases, the prediction uncertainty usually could not be formally attributed to model structure (e.g., model or process error), parameters, or initial conditions.

We acknowledge that our literature search criteria may have missed many relevant papers. For example, individual-based models of ecosystem components are often grounded in a mechanistic and dynamic and/or spatially explicit framework (e.g., Ogle and Pacala 2009) that can be coupled to data and used to produce future forecasts, but such models did not appear in our search. Our search also did not return studies on forecasting disease dynamics, yet we know that such models have been used to predict the timing and intensity of infectious diseases and to inform control strategies (Smith et al. 2005). Ecological disease models involve mathematical representations of disease-status transitions in host populations (Kermack and McKendrick 1933, Anderson and May 1979), and they are generally fitted to population time series to characterize disease progression and transmission (Grenfell et al. 2001, Bjornstad et al. 2002, Ferrari et al. 2008). Simulations based on these data-informed models can then be used to develop control strategies and predict the timing of epidemics, and inference from such models has led to successes in early warning of disease risk and vaccination strategies (Glass et al. 2000, Ferguson et al. 2001, Keeling et al. 2001, 2003, Smith et al. 2005, Chaves and Pascual 2007).

The Role of Data Assimilation in Ecological Forecasting

Most data–model comparison studies indicate that the current generation of predictive models is generally adequate to simulate qualitative patterns of large-scale dynamics of ecological systems (Parton et al. 1993, Hanson et al. 2004). It appears, however, that their ability to provide a realistic forecast at a given location is limited (Schimel et al. 1997). By coupling process thinking in the form of the numerical model and information contained in data, DA is expected to improve ecological forecasting by (1) providing estimates of parameters, initial values, and state variables; (2) quantifying uncertainties with respect to parameters, initial conditions, and modeled states of an ecosystem; (3) helping to select between alternative model structures; and (4) providing a quantitative basis to evaluate sampling strategies for future experiments and observations that will enable improvements to models and forecasts (Baker et al. 2008).

Brief description of data assimilation techniques

Modern DA methods combine data with model by updating model parameters and/or selecting alternative model structures (i.e., target variables) using optimization techniques or posterior simulations. Optimization procedures involve a cost function that quantifies the deviation (the vector **e**) between modeled and observed values as

$$\mathbf{e}(t) = \mathbf{Z}(t) - \varphi\mathbf{X}(t) \quad (1)$$

where $\mathbf{Z}(t)$ is an observation and $\varphi\mathbf{X}(t)$ is the modeled value at time t. The modeled value is usually related to state variables of the model, $\mathbf{X}(t)$, at time t using a mapping function φ. The mapping function relates the modeled variable to its observed counterpart. When DA is applied to multiple data sets, Eq. 1 represents a vector as

$$\mathbf{e}(t) = [e_1(t), e_2(t), \ldots e_i(t), \ldots, e_m(t)]^\top. \quad (2)$$

Corresponding to ith data set, Z_i, $i = 1, 2, \ldots, m$, there is

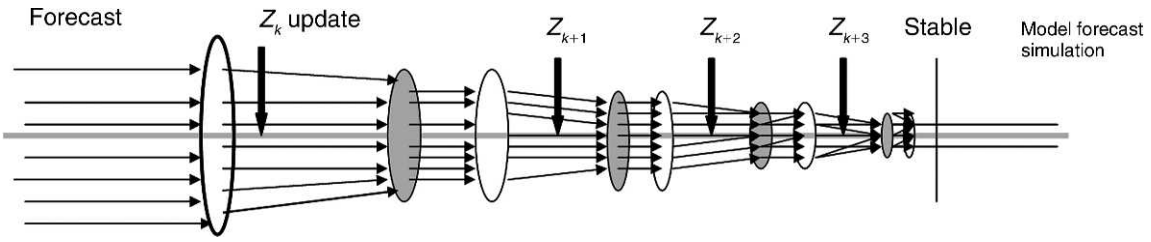

FIG. 4. Schematic illustration of an ensemble Kalman filter (EnKF). EnKF uses a Monte Carlo technique to generate an ensemble of estimated parameter values and forecasted state variables for sequential DA. Observations (Z, at time k) help to correct the forecast trajectories and reduce the spreads of parameter values and modeled state variables.

one random error component $e_i(t) = Z_i(t) - \varphi_i \mathbf{X}(t)$ with φ_i being a mapping function for the ith data set.

The deviation is usually termed an error, resulting from inaccurate observations, an imperfect model, or both. Most DA studies assume, for simplicity, that $\mathbf{e}(t)$ follows a Gaussian distribution with a zero mean (Braswell et al. 2005, Raupach et al. 2005), but this assumption is not required.

Optimization methods are used to obtain model structures and/or parameter values that minimize the deviations between modeled and observed values. The parameters or model structures (usually a set of difference functions) to be optimized by DA are termed target variables, c. Thus, the deviation becomes a function of c as

$$\mathbf{e}(t) = \mathbf{Z}(t) - \Phi\mathbf{X}(c, t). \quad (3)$$

By adjusting the target variables, c, the modeled value $\Phi\mathbf{X}(c, t)$ and consequent deviation (i.e., error), $\mathbf{e}(t)$, vary. Thus, we can define a cost function, $J(c)$, with multiple data sets as

$$J(c) = [\mathbf{Z}(t) - \Phi\mathbf{X}(c,t)]^{\top} \mathbf{cov}(e_t)^{-1} [\mathbf{Z}(t) - \Phi\mathbf{X}(c,t)] \quad (4)$$

where $\mathbf{cov}(e_t)$ is a covariance matrix for vector $\mathbf{e}(t)$. The non-diagonal elements in the matrix $\mathbf{cov}(e_t)$ represent correlations between different error components, while the diagonal elements specify variances of the components of $\mathbf{e}(t)$, which can be estimated from observations (Luo et al. 2003).

In most DA studies, the cost function is formulated using a least squares approach. The least squares method is equivalent to the maximum likelihood estimation for optimization when a Gaussian distribution is assumed for $\mathbf{e}(t)$ (Todling 2000). When the errors do not follow a Gaussian distribution or are not independent, other forms of the cost function may be used, such as the sum of absolute deviations, which gives maximum likelihood estimates for the Laplace error distribution (Liu et al. 2009).

The search for optimal target variables leading to minimal deviations between model predictions and observations (Eqs. 2 and 3) is generally accomplished using optimization techniques. There are two general types of optimization techniques: batch and sequential methods. Batch methods assimilate all the data within a time interval at once and treat the cost function as a single function to be minimized over that window. Sequential methods assimilate data one time step at a time. There are many batch techniques, such as the variational or adjoint method (Vukicevic et al. 2001), Levenburg-Marquardt (Luo et al. 2003), and genetic algorithms (Zhou and Luo 2008). A frequently used method is the Markov chain Monte Carlo (MCMC) technique, after the foundational work of Metropolis et al. (1953) and Hastings (1970). The basic idea is to use a Markov chain with Gibbs sampling and/or Metropolis-Hastings (M-H) algorithm to sample the target variables. Once the chain has been simulated for a sufficiently long period so that the distributions of target variables follow stationary states, samples from the simulations are collected to approximate the distributions of the target variables. When the DA approach is done within a Bayesian framework, the cost function is used in the Metropolis criterion to evaluate the full joint distribution of target variables (Xu et al. 2006). The posterior distributions of the target variables generated by MCMC algorithms can be used to determine most probable values, mean values, quantiles, and other summaries of uncertainty.

One of the most popular sequential methods is the Kalman filter (KF), which is a recursive DA algorithm for estimating initial conditions, parameters, and state variables of a system at each time using a state-space model from a series of heterogeneous, intermittent measurements (Kalman, 1960, Gelb 1974). The KF iteratively repeats two steps: forecast and update. The forecast step evolves the state vector forward in time, using a process-based or statistical model. The update step adjusts target parameters by a Kalman gain matrix when data are assimilated. The Kalman gain is calculated by minimizing the deviations between observations and model forecasts. The two-step procedure is repeated until the last datum is assimilated (Fig. 4). The ensemble Kalman filter (EnKF) is one variant of KF and uses a Monte Carlo technique to generate an ensemble of models for sequential DA (Evensen 2007). The covariance among parameters can be computed directly from the ensemble instead of from the explicit solution in the KF and its nonlinear extensions. KF and EnKF can estimate both parameter and state variables

in a dynamic ecological system simultaneously (Gao et al. 2011). State-space models are often implemented in a Bayesian framework, where current understanding of the state is a prior to be updated with each new observation (e.g., Clark and Bjornstad 2004).

Applications of data assimilation in ecological research

DA methods are being employed to develop, calibrate, and evaluate model accuracy and parameter uncertainty (Williams et al. 2009). Raupach et al. (2005) reviewed the use of DA methodologies in the analysis of terrestrial carbon observations, though they do not specifically address forecasting. They discussed the importance of methods for assimilating diverse data sources and separating observational and model errors, with the goal of producing more accurate analyses (or forecasts) of the global carbon cycle. Within this field, applications of DA range from global inversions (e.g., Lokupitiya et al. 2008, Ricciuto et al. 2008*b*) to individual forest stands (e.g., Reichstein et al. 2003, Clark et al. 2004, 2007, 2010, Klemedtsson et al. 2008, Moore et al. 2008). Several studies employed Bayesian methods to estimate state variables and parameters in process-based models to analyze ecosystem carbon exchange and its process controls (e.g., Braswell et al. 2005, Moore et al. 2008, Ricciuto et al. 2008*a*, *b*, Tang and Zhuang 2008), with some focusing on carbon and/or water exchange (e.g., Chen et al. 2008*b*, Klemedtsson et al. 2008, Moore et al. 2008, Svensson et al. 2008, Clark et al. 2011). Several studies employed the Kalman filter or ensemble Kalman filter directly (e.g., Williams et al. 2005, Chen et al. 2008*a*, Mo et al. 2008, Quaife et al. 2008) or other maximum likelihood or least squares optimization approaches (e.g., Reichstein et al. 2003, Lokupitiya et al. 2008, Prihodko et al. 2008).

DA methods have also been used extensively to calibrate marine or aquatic ecosystem models (Spitz et al. 1998, Dowd 2007). Techniques have included the use of Bayesian calibration approaches (Borsuk et al. 2004, Arhonditsis et al. 2008, Law et al. 2009), various versions of the Kalman filter (Ourmieres et al. 2009), as well as variational or adjoint methods (Zhao and Lu 2008). Arhonditsis et al. (2008) and Borsuk et al. (2004) also note that such methods, especially Bayesian approaches, are important for forecasting aquatic ecosystem responses in the context of environmental management where estimates of uncertainty are critical to making well informed decisions.

DA techniques have been used to explicitly estimate uncertainties of parameters and state variables. The uncertainty analysis has been applied to ecosystem phenology (Cook et al. 2005), carbon dynamics (Xu et al. 2006), salmon life cycles (Crozier et al. 2008), migration (Schwartz et al. 2001), and long-distance dispersal (Clark et al. 2003). Uncertainty may derive from many sources, such as (1) data uncertainties, due to random and systematic errors in observations; (2) model structural uncertainties arising from overall architecture (i.e., the way to connect components and processes), relationships among processes and drivers, and functional forms of equations to describe individual processes; (3) parameter uncertainties, typically due to inadequate or conflicting information about the parameters; (4) uncertainties in boundary conditions, such as scenarios of human behavior, climate, land use, and social and economic activities; and (5) uncertainties from the statistical method used to combine the model and data in the DA system. It is still challenging to separate such sources of uncertainty and assess their relative contributions to the forecast uncertainty, but new and incoming data and advances in DA methods are expected to facilitate this partitioning.

Ecological forecasting and data assimilation

DA methods are just beginning to be used in ecological forecasting. For example, Xu et al. (2006) applied a Bayesian probability inversion approach to estimate parameters and then forecast state variables (i.e., pool sizes) in the Duke forest, suggesting that, the ecosystem would store 3190 g C/m^2 more at elevated CO_2 by the year of 2010 than at ambient CO_2 with 95% confidence. Ricciuto et al. (2008*b*) used the posterior distributions for parameters (obtained by a Bayesian fitting procedure) in a global carbon cycle model to produce probabilistic forecasts. They obtained predictions of carbon sources and sinks under a future IPCC emission scenario, and then characterize the uncertainty in these predictions via posterior predictive distributions. Gao et al. (2011) used an ensemble Kalman filter (EnKF) system to assimilate eight sets of data from 1996 to 2004 into a terrestrial ecosystem model as a basis for forecasting carbon pools (state variables) daily from 2004 to 2012.

While most DA studies have been focused on parameter estimation, quantitative forecasts in complex dynamical systems require estimates of initial conditions. Initial conditions (e.g., abundance and age distribution in demographical models, biomass and pool sizes in biogeochemical models) are critical and sometimes govern the subsequent trajectory of systems. In a chaotic system, infinitesimal differences in initial conditions can lead to exponential divergence between trajectories (May 2001). For such a system, very complex DA procedures may be required to stabilize forecasts as developed for weather forecasting (Kalnay 2003). In carbon cycle modeling, initial values of pool sizes determine the directions and magnitudes of carbon sequestration (Carvalhais et al., 2008). Estimation of the initial pool sizes using DA methods is essential for quantifying the rate of carbon sequestration in an ecosystem.

Ecological system dynamics are so influenced by weather and climate that rarely are ecological models used to make a forecast alone. More often what-if analyses are conducted, using assumed future climate conditions as boundary or forcing conditions. For near-

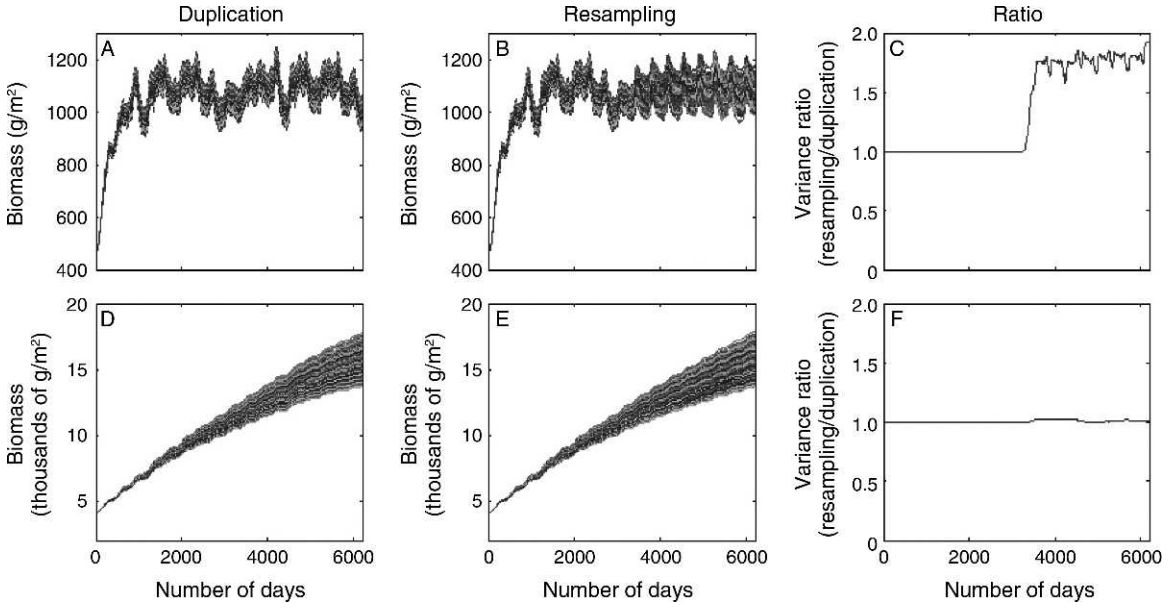

FIG. 5. Daily analysis from 1996 to 2004 and daily forecast from 2004 to 2012 of carbon sink dynamics in (A, B) leaf and (D, E) wood pools at Duke Forest using the Markov chain Monte Carlo (MCMC) method with the M-H criterion. The gray area is composed of many lines that are very dense and diverge over time. Each line in the plot shows the time series of the biomass value, simulated and forecasted using each accepted parameter set from data assimilation. Panels A and D present results of the daily analysis and forecast with repeated weather data as in Xu et al. (2006). The daily forecast in panels B and E was done with a resampling method. One data point of temperature and its corresponding moisture and GPP values in that day were randomly resampled from a pool of nine data points, lumped from 1996 to 2004, in each day of a year as input forcing variables to drive forecasting. Panels C and F present the ratio of variances from the resampling series divided by that from the original methods (i.e., variances in panel A divided by variances in panel B; variances in panel D devided by those in panel E).

term forecasts, past weather data are often repeated to drive model forecasts of future dynamics (e.g., Xu et al. 2006, Gao et al. 2011). When weather data are resampled to account for potential weather variability in influencing model forecasts, uncertainty in forcing variables strongly influences variability of forecasted state variables of short-term processes but less so for the long-term state variables (Fig. 5).

Potential limitations of data assimilation in ecological forecasting

It should be noted that DA methods may not improve forecasting when the associated ecosystem models do not include all key processes that potentially influence the system's behavior. For example, many ecosystem models incorporate carbon fixation, allocation, and decomposition processes but usually not disturbances, such as fire, land uses, and human activities. On the other hand, models that focus on disturbances often have minimal representation of physiology. Evolution of ecological systems over longer time scales and larger spatial scales depends on the interaction of physiological processes with population and community processes. All those processes are further influenced by stochastic or episodic disturbance events that are driven by processes occurring on different and broader time and space scales (for example, scales linked to severe weather, drought, or the time and space scales of dispersal and establishment of invasive species and pests). Linking scales is a frontier for process-based models in all disciplines and the opportunity exists for ecologists to lead in developing innovative techniques for forecasting the evolution of complex systems on global scales. Until these processes and disturbance events at different scales are integrated into models, DA will have limited utility given the limitations of existing models.

Moreover, DA may not improve forecasting when ecological processes have never been observed or understood, again resulting in process models that do not sufficiently describe the system of interest. For example, a coupled global carbon–climate model predicts that climate warming may drive ecosystems in tropical regions past critical thresholds, leading to forest dieback (Cox et al. 2004). However, in situ experiments of tropical forest response to elevated CO_2 and warming are not available. Indeed, we lack empirical knowledge on nonlinear responses, thresholds, and tipping points of ecosystems in a future climate-changed world (Williams et al. 2008). Thus, it is important to design experimental and observational studies to obtain critical data that will enable the development of models that are capable of capturing potentially novel conditions. The value of emerging DA techniques comes from the capacity to coherently integrate information from many sources, which is not the same problem as lack of information (Clark 2005). In other words, DA cannot fully substitute

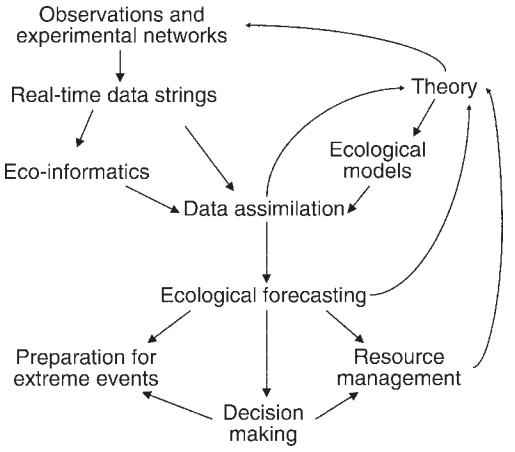

FIG. 6. Relationships among measurement networks, ecoinformatics, ecological theory and models, DA, decision making, and resource management in a data-rich era.

for lack of information, but rather allows one to fully exploit the information that is available.

OPPORTUNITIES AND CHALLENGES IN A DATA-RICH ERA

The field of ecology is changing from a data-limited to a data-rich scientific endeavor due to the accumulation of research data from networks such as FLUXNET together with the integration of information from space-borne remote sensing systems like MODIS, LANDSAT, and IKONOS, and the contribution of data from an enormous number of individual or small groups of investigators. Additionally, the incipient National Ecological Observatory Network (NEON) is designed to acquire measurements at many locations, multiple scales, and from both observations and experiments across the nation. NEON, beginning in about 2012 and expanding to completion in 2016, will generate large amounts of ecological data every day (terabytes to petabytes a year). There will be an unprecedented demand to convert the raw data from networks such as NEON into meaningful ecological information products, with the aim of accelerating advances in our fundamental knowledge of ecological processes, testing ecological theory, forecasting changes in ecological services, educating teachers and students, and supporting decision making.

In this data-rich era, DA will be an essential tool of ecological research and will assist the transformation of ecological research (Fig. 6). Major research activities may shift focus from measurement and data collection to data processing, analysis and interpretation, providing a way forward for improving ecological theory, models, and forecasts (Table 2). In the data-limited era, data could be stored and processed with simple tools, whereas data storage and analysis in a data-rich era must be accomplished with more sophisticated tools. In the data-rich era, data collection and analysis will often be designed by communities to infer broad-scale patterns and support decision-making processes. The designed observational and experimental programs will support a range of research goals synergistically. Continental or global long-term data sets obtained from those programs make it possible to develop novel diagnostics of ecosystem processes and to discover hidden features of data arrays, mutual relationships between variables, and couplings between ecological and geophysical systems. Ecosystem models then can confront these observations via DA to improve models for forecasting.

Data processing and assimilation in a data-rich era require new ecoinformatic and analysis tools (Fig. 6). Continental or global, long-term data could directly feed into DA systems to produce analyses and forecasts. Ecoinformatics involves more than data acquisition, curation, and dissemination with metadata. It also includes complex analyses and forecasts (including DA) to generate standard derived products. For instance, high-level data products, such as gross primary productivity, leaf area index, and land cover have been generated from NASA's MODIS sensors instead of just simple reflectance indices (e.g., Nemani et al. 2003). Second, we need to develop technologies to streamline data acquisition from sensors, observers, and laboratories, and to improve quality assessment and quality control, DA, and data products to support credible and timely forecasting and analysis. Thus, tools of workflows from observation or experimental networks to ecological forecasting need to be developed as a part of next-generation ecoinformatics and applied on a regular basis. As regularly computed data products from a research mode prove their values to science, education, or management, data production and forecasting may transition to an appropriate institutional home for sustained operation. Third, our experience indicates that ecological forecasting will be important for supporting decision-making processes to prepare society for complex ecological situations, including disturbances like invasions, pest outbreaks, and responses to wildfire, severe weather, and drought.

Additions to the data acquisition and curation components of ecoinformatics may include components of (1) core computational algorithms (e.g., ecological models) that are specifically designed to analyze and

TABLE 2. Characteristics of research in data-rich and data-limited eras.

Characteristics	Data-limited era	Data-rich era
Motive	curiosity driven	decision making
Primary activity	data collection	data analysis
Experimental design	individual researcher	research community
Focus	measurements	theory test and development
Data analysis	spreadsheet	eco-informatics
Objective	discovery	forecasting
Service to society	long-term	near- or real-time

forecast ecological change, (2) appropriate optimization techniques for DA, and (3) data that will support these models and algorithms from ecological measurements. This system needs to both support the volume and complexity of the primary and derived data and, perhaps more challenging, provide environmental information to diverse stakeholders, researchers, students and teachers, citizens, and decision-makers. We believe that individual investigators and groups will be able to make use of the observational and forecast data resources to more creatively address larger-scale questions than they could if they collected all the data themselves.

To meet the challenges in the data-rich era, we need continuous development of DA techniques. Most DA studies have used a single model, such as TECO (Luo et al. 2003, Xu et al. 2006), SIPNET (Braswell et al. 2005), and REFLEX (Fox et al. 2009). But ecosystems in the real world have to be described by much more comprehensive models that are capable of prediction under novel scenarios. Application of multiple, competing models, however, can be more important in the context of forecasting (Dormann et al. 2008). For example, the use of multiple, highly detailed, process-based models within a rigorous DA framework (referred to as "multimodel superensembles") has greatly improved near real-time hurricane forecasting (Krishnamurti et al. 1999, Williford et al. 2003).

Conclusions

Observation and process thinking are two fundamental approaches to scientific inquiry. Observation records in data the states of ecosystems and information of underlying processes at the time when the measurements are made. Process thinking identifies which processes operate in a system, formally expressed in a theory and/or model. Both approaches provide insights into ecological systems in different but complementary ways. Combining the two approaches by assimilating empirical data into models generally provides significantly greater understanding, yet such data–model integration is underutilized. Recently, data assimilation (DA) has been actively applied to improvement of models in several areas, such as carbon cycle models, process-based dynamic vegetation models, and marine or aquatic ecosystem models. Most ecological DA studies have focused on parameter estimation, with a few studies addressing effects of initial conditions and selection of alternative model structures. While improved models via DA eventually contribute to advances in ecological forecasting, not many studies have explicitly addressed issues directly pertinent to forecasting, such as forecasting accuracy, sources of uncertainty, and usefulness of forecasting under different domains. As we enter a data-rich era, measurement networks yield vast amounts of temporally and/or spatially rich data that may be used within a DA framework to improve existing models. Computationally intensive methods are required to assimilate extensive data into ecological models and to make realistic forecasts of ecological change. We need a research agenda linking ecological process thinking, modeling, and advanced observational programs to innovative cyberinfrastucture, computational science, and applied mathematics to facilitate the necessary transformation to a data-rich paradigm for ecological research.

Acknowledgments

This study was financially supported by the National Science Foundation (NSF) under DEB 0743778, DEB 0840964, EF 0938795; by the Office of Science (BER), Department of Energy, Grants No. DE-FG02-006ER64319; and through the Midwestern Regional Center of the National Institute for Climatic Change Research at Michigan Technological University, under Award Number DE-FC02-06ER64158.

Literature Cited

Akimoto, H. 2003. Global air quality and pollution. Science 302:1716–1719.

Anderson, R. M., and R. M. May. 1979. Population biology of infectious diseases. Part 1. Nature 280:361–367.

Arhonditsis, G. B., D. Papantou, W. T. Zhang, G. Perhar, E. Massos, and M. L. Shi. 2008. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. Journal of Marine Systems 73:8–30.

Baker, D. F., H. Bösch, S. C. Doney, and D. S. Schimel. 2008. Carbon source/sink information provided by column CO_2 measurements from the Orbiting Carbon Observatory. Atmospheric Chemistry and Physics Discussions 8(6): 20051–20112.

Baldocchi, D., et al. 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society 82:2415–2434.

Beaumont, L. J., L. Hughes, and M. Poulsen. 2005. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. Ecological Modelling 186:250–269.

Beaumont, L. J., A. J. Pitman, M. Poulsen, and L. Hughes. 2007. Where will species go? Incorporating new advances in climate modelling into projections of species distributions. Global Change Biology 13:1368–1385.

Beerling, D. J., F. I. Woodward, M. Lomas, and A. J. Jenkins. 1997. Testing the responses of a dynamic global vegetation model to environmental change: a comparison of observations and predictions. Global Ecology and Biogeography 6:439–450.

Bjornstad, O. N., B. F. Finkenstadt, and T. G. Bryan. 2002. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological Monographs 72:169–184.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow. 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecological Modelling 173:219–239.

Braswell, B. H., W. J. Sacks, E. Linder, and D. S. Schimel. 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Global Change Biology 11:335–355.

Carvalhais, N., et al. 2008. Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. Global Biogeochemical Cycles 22:GB2007.

Chaves, L. F., and M. Pascual. 2007. Comparing models for early warning systems of neglected tropical diseases. PLoS Neglected Tropical Diseases 1:e33.

Chen, M., S. Liu, L. L. Tieszen, and D. Y. Hollinger. 2008*a*. An improved state-parameter analysis of ecosystem models using data assimilation. Ecological Modelling 219:317–326.

Chen, X. Y., Y. Rubin, S. Y. Ma, and D. Baldocchi. 2008*b*. Observations and stochastic modeling of soil moisture control on evapotranspiration in a Californian oak savanna. Water Resources Research 44.

Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8:2–14.

Clark, J. S., P. Agarwal, D. M. Bell, P. G. Flikkema, A. Gelfand, X. Nguyen, E. Ward, and J. Yang. 2011. Inferential ecosystem models, from network data to prediction. Ecological Applications 21:1523–1536.

Clark, J. S., et al. 2010. High-dimensional coexistence based on individual variation: a synthesis of evidence. Ecological Monographs 80:569–608.

Clark, J. S., and O. Bjornstad. 2004. Population time series: process variability, observation errors, missing values, lags, and hidden states. Ecology 85:3140–3150.

Clark, J. S., et al. 2001. Ecological forecast: an emerging imperative. Science 293:657–660.

Clark, J. S., S. LaDeau, and I. Ibanez. 2004. Fecundity of trees and the colonization–competition hypothesis. Ecological Monographs 74:415–442.

Clark, J. S., M. Lewis, J. S. McLachlan, and J. HilleRisLambers. 2003. Estimating population spread: What can we forecast and how well? Ecology 84:1979–1988.

Clark, J. S., M. Wolosin, M. Dietze, I. Ibanez, S. LaDeau, M. Welsh, and B. Kloeppel. 2007. Tree growth inference and prediction from diameter censuses and ring widths. Ecological Applications 17:1942–1953.

Clein, J. S., B. L. Kwiatkowski, A. D. McGuire, J. E. Hobbie, E. B. Rastetter, J. M. Melillo, and D. W. Kicklighter. 2000. Modelling carbon responses of tundra ecosystems to historical and projected climate: a comparison of a plot and a global-scale ecosystem model to identify process-based uncertainties. Global Change Biology 6:127–140.

Cook, B. I., T. M. Smith, and M. E. Mann. 2005. The North Atlantic Oscillation and regional phenology prediction over Europe. Global Change Biology 11:919–926.

Cox, P. M., R. A. Betts, M. Collins, P. P. Harris, C. Huntingford, and C. D. Jones. 2004. Amazonian forest dieback under climate-carbon cycle projections for the 21st century. 78:137–156.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell. 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408:184–187.

Crozier, L. G., R. W. Zabel, and A. F. Hamlett. 2008. Predicting differential effects of climate change at the population level with life-cycle models of spring Chinook salmon. Global Change Biology 14:236–249.

Currie, D. J. 2001. Projected effects of climate change on patterns of vertebrate and tree species richness in the conterminous United States. Ecosystems 4:216–225.

D'Antonio, C. M., and P. M. Vitousek. 1992. Biological invasions by exotic grasses, the grass fire cycle, and global change. Annual Review of Ecology and Systematics 23:63–87.

Dormann, C. F., et al. 2008. Prediction uncertainty of environmental change effects on temperate European biodiversity. Ecology Letters 11:235–244.

Dowd, M. 2007. Bayesian statistical data assimilation for ecosystem models using Markov Chain Monte Carlo. Journal of Marine Systems 68:439–456.

Evensen, G. 2007. Data assimilation: the ensemble Kalman filter. Springer, Berlin, Germany.

Ferguson, N. M., C. A. Donnelly, and R. M. Anderson. 2001. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. Nature 413:542–528.

Ferrari, M. J., R. F. Grais, N. Bharti, J. K. Conlan, O. N. Bjornstad, L. J. Wolfson, P. J. Guerin, A. Djibo, and B. T. Grenfell. 2008. The dynamics of measles in sub-Saharan Africa. Nature 451:679–684.

Fisher, R. A. 1959. Statistical methods and scientific inference. Second edition. Oliver and Boyd, Edinburgh, UK.

Foley, J. A., S. Levis, I. C. Prentice, D. Pollard, and S. L. Thompson. 1998. Coupling dynamic models of climate and vegetation. Global Change Biology 4:561–579.

Forrester, J. W. 1961. Industry dynamics. MIT Press. Cambridge, Massachusetts, USA.

Fox, A. M., M. Williams, A. D. Richardson, D. Cameron, J. Gove, M. Reichstein, T. Quaife, D. Ricciuto, E. Tomelleri, C. M. Trudinger, and M. T. Van Wijk. 2009. The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. Agricultural and Forest Meteorology 149:1597–1615.

Friedlingstein, P., et al. 2006. Climate–carbon cycle feedback analysis: results from the C^4MIP model intercomparison. Journal of Climate 19:3337–3353.

Gao, C., H. Wang, E. Weng, S. Lakshmivarahan, Y. Zhang, and Y. Luo. 2011. Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics. Ecological Applications 21:1461–1473.

Gelb, A. 1974. Applied optimal estimation. MIT Press. Cambridge, Massachusetts, USA.

Glass, M., M. Dragunow, and R. L. M. Faull. 2000. The pattern of neurodegeneration in Huntington's disease: a comparative study of cannabinoid, dopamine, adenosine and GABA(A) receptor alterations in the human basal ganglia in Huntington's disease. Neuroscience 97:505–519.

Gleick, P. H. 2002. The world's water 2002–2003: the biennial report on freshwater resources. Island Press, Washington, D.C., USA.

Grenfell, B. T., O. N. Bjornstad, and J. Kappey. 2001. Travelling waves and spatial hierarchies in measles epidemics. Nature 414:716–723.

Hanson, P. J., et al. 2004. Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data. Ecological Monographs 74:443–489.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chain and their applications. Biometrika 57:97–109.

Heinz Report. 2008. The state of the nation's ecosystems. Measuring the land, waters, and living resources of the United States. Island Press, Washington, D.C., USA.

Hijmans, R. J., and C. H. Graham. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology 12:2272–2281.

Hood, S. M., C. W. Mchugh, K. C. Ryan, E. Reinhardt, and S. L. Smith. 2007. Evaluation of a post-fire tree mortality model for western USA conifers. International Journal of Wildland Fire 16:679–689.

Jones, C., C. McConnell, K. Coleman, P. Cox, P. Falloon, D. Jenkinson, and D. Powlson. 2005. Global climate change and soil carbon stocks; predictions from two contrasting models for the turnover of organic carbon in soil. Global Change Biology 11:154–166.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82:35–45.

Kalnay, E. 2003. Atmospheric modeling, data assimilation and predictability. Cambridge University Press, Cambridge, UK.

Kearney, M., and W. Porter. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecology Letters 12:334–350.

Keeling, M. J., M. E. J. Woolhouse, R. M. May, G. Davies, and B. T. Grenfell. 2003. Modelling vaccination strategies against foot-and-mouth disease. Nature 421:136–142.

Keeling, M. J., M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294:813–817.

Kermack, W. O., and A. G. McKendrick. 1933. Contributions to the mathematical theory of epidemics. Part III. Proceedings of the Royal Society A 141:94–122.

Kirilenko, A., B. Chivoiu, J. Crick, A. Ross-Davis, K. Schaaf, G. Shao, V. Singhania, and R. Swihart. 2007. An internet-based decision support tool for non-industrial private forest landowners. Environmental Modeling and Software 22:1498–1508.

Klemedtsson, L., P. E. Jansson, D. Gustafsson, L. Karlberg, P. Weslien, K. von Arnold, M. Ernfors, O. Langvall, and A. Lindroth. 2008. Bayesian calibration method used to elucidate carbon turnover in forest on drained organic soil. Biogeochemistry 89:61–79.

Knapp, A. K., and M. D. Smith. 2001. Variation among biomes in temporal dynamics of aboveground primary production. Science 291:481–484.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran. 1999. Improved weather and seasonal climate forecasts from multimodel superensemble. Science 285:1548–1550.

Law, T., W. T. Zhang, J. Y. Zhao, and G. B. Arhonditsis. 2009. Structural changes in lake functioning induced from nutrient loading and climate variability. Ecological Modelling 220:979–997.

Liao, C. Z., R. H. Peng, Y. Q. Luo, X. H. Zhou, X. W. Wu, C. M. Fang, J. K. Chen, and B. Li. 2008. Altered ecosystem carbon and nitrogen cycles by plant invasion: a meta-analysis. New Phytologist 177:706–714.

Liu, M., H. L. He, G. R. Yu, Y. Q. Luo, X. M. Sun, and H. M. Wang. 2009. Uncertainty analysis of CO_2 flux components in subtropical evergreen coniferous plantation. Science in China Series D: Earth Sciences 52:257–268.

Lokupitiya, R. S., D. Zupanski, A. S. Denning, S. R. Kawa, K. R. Gurney, and M. Zupanski. 2008. Estimation of global CO_2 fluxes at regional scale using the maximum likelihood ensemble filter. Journal of Geophysical Research—Atmospheres 113:D20110.

Lotka, A. J. 1924. Elements of physical biology. Williams and Wilkins, New York, New York, USA.

Luo, Y., L. White, J. Canadell, E. DeLucia, D. Ellsworth, A. Finzi, J. Lichter, and W. Shclesinger. 2003. Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with inversion approach. Global Biogeochemical Cycles 17(1):1021.

Luo, Y., and X. Zhou. 2006. Soil respiration and the environment. Academic Press/Elsevier, San Diego, California, USA.

May, R. M. 1981. Theoretical ecology: principles and applications. Sinauer, Sunderland, Massachusetts, USA.

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

McGuire, A. D., J. S. Clein, J. M. Melillo, D. W. Kicklighter, R. A. Meier, C. J. Vorosmarty, and M. C. Serreze. 2000. Modelling carbon responses of tundra ecosystems to historical and projected climate: sensitivity of pan-Arctic carbon storage to temporal and spatial variation in climate. Global Change Biology 6:141–159.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculation by fast computer machines. Journal of Chemical Physics 21:1087–1092.

Mo, X. G., J. M. Chen, W. M. Ju, and T. A. Black. 2008. Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter. Ecological Modelling 217:157–173.

Monson, R. K., D. L. Lipson, S. P. Burns, A. A. Turnipseed, A. C. Delany, M. W. Williams, and S. K. Schmidt. 2006. Winter forest soil respiration controlled by climate and microbial community composition. Nature 439:711–714.

Moore, D. J. P., J. Hu, W. J. Sacks, D. S. Schimel, and R. K. Monson. 2008. Estimating transpiration and the sensitivity of carbon uptake to water availability in a subalpine forest using a simple ecosystem process model informed by measured net CO_2 and H_2O fluxes. Agricultural and Forest Meteorology 148:1467–1477.

Morin, X., D. Viner, and I. Chuine. 2008. Tree species range shifts at a continental scale: new predictive insights from a process-based model. Journal of Ecology 96:784–794.

Nemani, R. R., C. D. Keeling, H. Hashimoto, W. M. Jolly, S. C. Piper, C. J. Tucker, R. B. Myneni, and S. W. Running. 2003. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science 300:1560–1563.

Ogle, K. 2009. Hierarchical Bayesian statistics: merging experimental and modeling approaches in ecology. Ecological Applications 19:577–581.

Ogle, K., and S. W. Pacala. 2009. A modeling framework for inferring tree growth and allocation from physiological, morphological, and allometric traits. Tree Physiology 29:587–605.

Ourmieres, Y., P. Brasseur, M. Levy, J. M. Brankart, and J. Verron. 2009. On the key role of nutrient data to constrain a coupled physical-biogeochemical assimilative model of the North Atlantic Ocean. Journal of Marine Systems 75:100–115.

Pacala, S. W., C. D. Canham, and J. A. Silander. 1993. Forest models defined by field-measurements. 1. The design of a northeastern forest simulator. Canadian Journal of Forest Research 23:1980–1988.

Parton, W. J., D. S. Schimel, C. V. Cole, and D. S. Ojima. 1987. Analysis of factors controlling soil organic-matter levels in Great Plains grasslands. Soil Science Society of America Journal 51:1173–1179.

Parton, W. J., J. M. O. Scurlock, D. S. Ojima, T. G. Gilmanov, R. J. Scholes, D. S. Schimel, T. Kirchner, J.-C. Menaut, T. Seastedt, E. Garcia Moya, Apinan Kamnalrut, and J. L. Kinyamario. 1993. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. Global Biogeochemical Cycles 7:785–809.

Pascual, M. A., and P. Kareiva. 1996. Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. Ecology 77:337–349.

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199.

Prihodko, L., A. S. Denning, N. P. Hannan, I. Baker, and K. Davis. 2008. Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. Agricultural and Forest Meteorology 148:268–287.

Quaife, T., P. Lewis, M. De Kauwe, M. Williams, B. E. Law, M. Disney, and P. Bowyer. 2008. Assimilating canopy reflectance data into an ecosystem model with an ensemble Kalman filter. Remote Sensing of Environment 112:1347–1364.

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 uncertainty specifications. Global Change Biology 11:378–397.

Reichstein, M., J. Tenhunen, O. Roupsard, J. M. Ourcival, S. Rambal, F. Miglietta, A. Peressotti, M. Pecchiari, G. Tirone,

and R. Valentini. 2003. Inverse modeling of seasonal drought effects on canopy CO_2/H_2O exchange in three Mediterranean ecosystems. Journal of Geophysical Research—Atmospheres 108:4726.

Ribbens, E., J. A. Silander, and S. W. Pacala. 1994. Seedling recruitment in forests: calibrating models to predict patterns of tree seedling dispersion. Ecology 75:1794–1806.

Ricciuto, D. M., M. P. Butler, K. J. Davis, B. D. Cook, P. S. Bakwin, A. Andrews, and R. M. Teclaw. 2008*a*. Causes of interannual variability in ecosystem–atmosphere CO_2 exchange in a northern Wisconsin forest using a Bayesian model calibration. Agricultural and Forest Meteorology 148:309–327.

Ricciuto, D. M., K. J. Davis, and K. Keller. 2008*b*. A Bayesian calibration of a simple carbon cycle model: the role of observations in estimating and reducing uncertainty. Global Biogeochemical Cycles 22:GB2030.

Schimel, D. S., et al. 1997. Continental scale variability in ecosystem processes: Models, data, and the role of disturbance. Ecological Monographs 67:251–271.

Schwartz, M. W., L. R. Iverson, and A. M. Prasad. 2001. Predicting the potential future distribution of four tree species in Ohio using current habitat availability and climatic forcing. Ecosystems 4:568–581.

Smith, K. F., A. P. Dobson, F. E. McKenzie, L. A. Real, D. L. Smith, and M. L. Wilson, 2005. Ecological theory to enhance infectious disease control and public health policy. Frontiers in Ecology and the Environment 3:29–37.

Solomon, S., et al. 2007. Technical summary. *In* Climate change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. S Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. Cambridge University Press, Cambridge, UK.

Spitz, Y. H., J. R. Moisan, M. R. Abbott, and J. G. Richman. 1998. Data assimilation and a pelagic ecosystem model: parameterization using time series observations. Journal of Marine Systems 16:51–68.

Svensson, M., P. E. Jansson, D. Gustafsson, D. B. Kleja, O. Langvall, and A. Lindroth. 2008. Bayesian calibration of a model describing carbon, water and heat fluxes for a Swedish boreal forest stand. Ecological Modelling 213:331–344.

Tang, J. Y., and Q. L. Zhuang. 2008. Equifinality in parameterization of process-based biogeochemistry models: a significant uncertainty source to the estimation of regional carbon dynamics. Journal of Geophysical Research—Biogeosciences. 113:G04010.

Todling, R. 2000. Estimation theory and atmospheric data assimilation. Pages 49–65 *in* P. S. Kasibhatla, editor. Inverse methods in global biogeochemical cycles. American Geophysical Monograph 114. American Geophysical Union, Washington, D.C., USA.

Vallecillo, S., L. Brotons, and W. Thuiller. 2009. Dangers of predicting bird species distributions in response to land-cover changes. Ecological Applications 19:538–549.

Verhulst, P. F. 1838. Notice sur la loi que la population poursuit dans son accroissement. Correspondance mathématique et physique 10:113–121.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human domination of Earth's ecosystems. Science 277:494–499.

Volterra, V. 1926. Fluctuations in the abundance of a species considered mathematically. Nature 118:558–560.

Vukicevic, T., B. Braswell, and D. S. Schimel. 2001. A diagnostic study of temperature controls on global terrestrial carbon exchange. Tellus B 53:150–170.

Weng, E., and Y. Luo. 2011. Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. Ecological Applications 21:1490–1505.

Williams, M., et al. 2009. Improving land surface models with FLUXNET data. Biogeosciences 6:2785–2835.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius. 2005. An improved analysis of forest carbon dynamics using data assimilation. Global Change Biology 11:89–105.

Williams, S. E., L. P. Shoo, J. L. Isaac, A. A. Hoffmann, and G. Langham. 2008. Towards an integrated framework for assessing the vulnerability of species to climate change. PLoS Biology 6(12):e325.

Williford, C. E., T. N. Krishnamurti, R. C. Torres, S. Cocke, Z. Christidis, and T. S. V. Kumar. 2003. Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. Monthly Weather Review 131:1878–1894.

Xu, T., L. White, D. Hui, and Y. Luo. 2006. Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. Global Biogeochemical Cycles 20:GB2007.

Zhao, Q., and X. Lu. 2008. Parameter estimation in a three-dimensional marine ecosystem model using the adjoint technique. Journal of Marine Systems 74:443–452.

Zhou, T., and Y. Luo. 2008. Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous United States. Global Biogeochemical Cycles 22:GB3032.