The Utility of Simple Models in Ecosystem Science

Michael L. Pace

Summary

Ecosystem models are often envisioned as complex, multicompartment simulations. This vision, however, misses a large class of models that are "simple" in the sense that they consist of only one or a few equations. Three uses of simple models in ecosystem research are comparisons of empirical relationships, evaluation of alternate states, and prediction. Comparisons of ecosystems often reveal empirical patterns. For example, general patterns of the sinking flux of organic matter as a function of primary production differ between lakes and ocean systems. An evaluation of the underlying relationships reveals a distinction between oligotrophic systems that suggests fundamental differences between lake and ocean primary producers. Simple dynamic models represent a second type of modeling approach that is particularly useful for exploring conditions that promote shifts in ecosystem states. An example of these alternate states is the shift from turbid to clear-water conditions in shallow lakes. Simple models also provide predictions and may outperform more complex models in forecasting. A recent evaluation of models that forecast the dynamics of the 1997–1998 El Niño reveals that well-developed dynamic and statistical models were no better, and in some cases much worse, than a simple empirical model based on historical data.

Simple models have a wide utility in ecosystem science. They are useful for exploring general questions and depicting ideas as well as for evaluating specific research problems. Simple models are pathways to a better understanding of the patterns and underlying mechanisms that structure ecosystems, and they can aid communication with other scientists, managers, and the public.
Introduction

Imagine asking a group of ecologists to envision a model of an ecosystem. What would they see? My guess is that regardless of whether this model was of a grassland, forest, ocean, or stream, most would think of an ecosystem in a box-and-arrow format consisting of numerous compartments, flows, and feedbacks. These types of modest to highly complex models have been the dominant approach to the analysis of ecosystems since the initial studies of the International Biosphere Program (McIntosh 1985; Lauenroth et al. 1998). Current models of this genre include well-known terrestrial ecosystem models such as CENTURY (Parton et al. 1987), PnET (Aber and Federer 1992), and GEM (Rastetter et al. 1991). A common feature of these models is that they are relatively complex: they are based on more than a few equations and require that a number of parameters be estimated.

This vision of ecosystem modeling, however, is incomplete. It misses a large class of models that are “simple” in the sense that they consist of only one or a few equations. Simple models most often describe some process or aspect of ecosystems. This mathematical description might be strictly empirical (correlative) or more mechanistically based, for example, on equations that conserve mass flow. The purpose and goals of simple models are highly diverse, encompassing many aspects of ecosystem science. These models are ubiquitous tools and, hence, deserve consideration in any discussion of ecosystem modeling.

This chapter considers simple models by discussing their utility in ecosystem science, but rather than providing a comprehensive review, it instead emphasizes a few general points. First, simple empirical models are useful for integrating the results of comparative studies and can lead to important, novel insights about ecosystems. Second, simple dynamic models aid the understanding of alternate conditions in ecosystems and the possible mechanisms that promote or inhibit shifts between states. Third, simple models make useful predictions, and these predictions may be better than those derived from more complex models. Each point is illustrated by a specific example, but the specifics serve only as a means toward a larger goal of assessing the uses, advantages, and contributions of simple models.

There is no specific definition of simple models, but here I will consider them to be models with only a few equations. This discussion of the benefits of simple models should not be interpreted as a criticism of more complex models based on many equations. The advantages and shortcomings of large ecosystem models is a separate subject, and the advocacy for the use of simple models made here does not imply a lesser role for more complex models.

Simple Empirical Models and Comparisons of Ecosystems

Simple empirical models such as linear regressions are often used to integrate data from studies of numerous ecosystems. For example, many relationships
between soil respiration and temperature often are established from a variety of ecosystems with different soils and vegetation (e.g., Raich and Schlesinger 1992, Lloyd and Taylor 1994). These relationships set a standard for evaluating new data from different systems, serve as equations in models that require estimates of soil respiration, and summarize the overall pattern of how respiration changes with temperature.

In addition to summarizing major patterns, comparative studies can also lead to unexpected insight about similarities and differences of ecosystems. The comparison by Baines et al. (1994) of the sinking flux of organic carbon in ocean and lake ecosystems provides an example. Oceanographers have long been interested in the losses of carbon (C) from the photic, surface layer to the deep ocean, because this flux of C is important in the global carbon cycle. The transfer of organic carbon from surface to deep waters is known as the "biological pump" (Volk and Hoffert 1985). Inorganic carbon is taken up by phytoplankton. The resulting organic carbon has three possible fates that are mediated by a combination of nutrient, phytoplankton, and food web interactions. Most of the phytoplankton carbon is eventually respired, some is converted to dissolved organic carbon (DOC), and the balance sinks as particulate organic matter (POC). The sinking of POC, either passively as dead organic matter or actively by organism migration, and net losses of DOC during deep mixing are the two means whereby the biological pump moves carbon from surface to deep water.

The efficiency of upper-ocean systems is often assessed by comparing the sinking flux of POC (measured in sediment traps) out of the photic zone relative to primary production (Wassman 1998). The sinking flux ratio (SFR) tends to increase with primary production in oceanic ecosystems—a pattern observed in a number of different studies (Figure 4.1a). However, measurements of SFR in lakes reveal a very different pattern (Baines and Pace 1994): SFR tends to decrease with primary production (Figure 4.1a).

Why do these patterns differ so strikingly? One way to answer the question is to decompose SFR into its components and ask if there are fundamental differences between oceanic systems and lakes. The sinking flux ratio (SFR) is a function of three variables, sinking flux, SF (mg C m$^{-2}$ d$^{-1}$); volumetric primary production, PP (mg C m$^{-3}$ d$^{-1}$); and depth, D (m).

$$SFR = SF / (PP \times D)$$

All of these variables relate to the concentration of phytoplankton biomass as measured by chlorophyll a. Baines et al. (1994) derived independent, empirical equations for lakes and oceans relating chlorophyll to SF, PP, and D. The relationships between chlorophyll and SF as well as chlorophyll and D were sufficiently similar for lakes and oceans that these could not explain the difference. There was, however, a strong difference between the relationships of chlorophyll and PP (Figure 4.1b). There is much greater production per unit biomass in the ocean relative to lakes at low concentrations of chlorophyll (< 1
Figure 4.1. (A) The relationship between primary production and percent sinking flux of particulate organic matter. Solid line is for lakes and represents a linear regression fit to data assembled by Baines et al. (1994). Dashed line is an approximation of several ocean models; see Baines et al. 1994 for details. (B) Relationships for chlorophyll versus primary production for lake (solid line) and ocean (dashed line) systems. For both panels A and B, confidence intervals have been excluded to emphasize the patterns. Details on the uncertainties of the relationships are provided in the original study. (Used with permission from Baines et al. 1994, 214 and 219.)
mg m\(^{-3}\)). There are many possible reasons for this, one of which is the very small size of oceanic primary producers in the low biomass systems of the most oligotrophic regions of the ocean. These areas are dominated by phytoplankton such as *Prochlorococcus* (Chisholm et al. 1988), and this group has not been observed in lakes.

Whatever the reason, the differences of the relationships for chlorophyll and PP are sufficient to account for the distinctions between lakes and oceans in SFR across productivity gradients. Substituting the individual empirical relations based on chlorophyll (Chl) into equation 4.1 results in the following equations for oceans (4.2) and lakes (4.3):

\[
\log \text{SFR} = -0.67 + 0.30 \log \text{Chl} \quad (4.2)
\]

\[
\log \text{SFR} = -0.22 - 0.41 \log \text{Chl} + 0.27 \log \text{Chl}^2 \quad (4.3)
\]

The positive slope in the ocean equation results in an increase in SFR with increases in phytoplankton biomass, while in freshwater systems SFR initially declines and eventually flattens with increases in phytoplankton biomass across the range of observations. These equations for SFR as a function of chlorophyll make distinct predictions that conform to data (Baines et al. 1994).

This study illustrates how empirical models can be used to establish and evaluate ecosystem patterns and to make unexpected discoveries about differences among ecosystems. The insights gained are about ecosystems in general and not about a specific system. The evaluation of simple statistical models not only leads to significant generalizations but also raises important questions about the reasons for the difference in how lakes and oceans function as revealed by the Baines et al. (1994) study. While much of ecological science necessarily focuses on method, measurement, and mechanism, synthesis of existing data into simple models as in this example is an equally significant activity that often helps identify critical conceptual questions.

**Dynamic Models and Ecosystem States**

Another class of models that are more explanatory and are based on dynamic differential equations is useful for analyzing and characterizing states and state transitions in ecosystems. These models help determine if changes observed in ecosystems are consistent with a relatively small set of mechanisms represented by only a few equations with a small number of parameters. Alternate states and state transitions arise from certain combinations of parameters, which may represent relatively common or relatively rare conditions. For example, Carpenter and Pace (1997) used a simple model of lakes to address whether the conditions of dystrophy (brown water due to high organic carbon loading) and eutrophy (green water due to high nutrient loading) constituted alternate stable states. Analysis of the model revealed that less than 10% of the parameter sets
produced model solutions that were consistent with alternate stable states, suggesting that internal processes do not normally cause lakes to change rapidly between green and brown water conditions. Instead, the model indicates that these lake states are ends along a continuum that is determined by relative rates of nutrient and organic matter loading.

Better evidence for alternate states comes from studies of shallow lakes. Some shallow lakes shift between a vegetation-dominated state and a turbid state where algae and/or suspended sediments dominate the water column and inhibit the growth of vegetation (Scheffer et al. 1993). Lake Zwenlust in the Netherlands provides an example (Van Donk and Gulati 1995; Scheffer 1998). Prior to 1988, algae dominated the lake with massive sustained blooms occurring each year. The lake was manipulated to remove most of the fish, and this allowed grazers to flourish, holding the phytoplankton blooms in check. With increased transparency, the favorable light conditions allowed submersed vegetation to colonize and grow on the bottom. The lake shifted to a clear-water state with little phytoplankton and a dominance of submerged, rooted vegetation that was maintained through much of the 1990s. Subsequently, the lake has shifted back to an algae-dominated state. Changes of this type have been observed in a number of different shallow lakes and appear to be driven by several mechanisms, including changes in nutrient loading, storms, changes in water level, and shifts in food web structure (Scheffer 1998).
Several switches appear to cause the shift from an algae-dominated state to a clear-water, submersed vegetation state (Figure 4.2). One set of switches is due to the food web. Shallow, eutrophic lakes often support large populations of planktivorous fish that severely limit grazer populations. If these fish are removed, grazers can exert some control over algae and promote a shift to vegetation dominance. Another switch is related to resuspension of sediments and turbidity due to algae that collectively determine light availability for submersed plants. Once some vegetation is established, there is an inhibition of wave formation and resuspension of sediments allowing greater light penetration with further development of vegetation. These interactions can promote a switch from algae to vegetation dominance. Fish can also have differential net effects based on the relative importance of benthic- versus pelagic-feeding forms. Through their feeding activities, benthic fish help to sustain turbid-water states and inhibit shifts toward a vegetated state.

The effects of turbidity and vegetation can be modeled by two equations that take into account turbidity and the fraction of the lake covered by vegetation together with feedback effects of each variable on the other (Scheffer 1998). In absence of vegetation, turbidity depends on the concentration of algae, which is a function of nutrient loading. This maximum turbidity is modified by vegetation cover, with turbidity declining as vegetation increases. The fraction of the lake covered by vegetation is similarly influences by turbidity.

When run with standard parameters for a shallow Dutch lake, this type of model leads to two stable equilibria (Figure 4.3). The lake exists either in a turbid state or in a vegetated state; intermediate conditions are unstable. Importantly, the behavior of this type of model is consistent with numerous observations. In addition to establishing that this conceptualization of the functioning of shallow lakes is consistent with alternate stable states, the model makes a number of qualitative predictions about how these systems change. It indicates that there are discontinuous responses at thresholds, shifts among states exhibiting hysteresis, and dramatic responses to large perturbations. These predictions agree with various observations, experiments, and long-term studies of shallow lakes, and the general concept of alternate states driven by switches is now being used in lake restoration and management (Moss 1996).

In a discussion of modeling shallow lakes, Scheffer (1998) suggests that simple models are, "best used to study the properties of isolated mechanisms... not to obtain quantitative insights into the relative importance of different mechanisms." Yet, ecosystems are complex and subject to multiple mechanisms that interact and operate simultaneously or in sequence. So are simple dynamic models simply too simplistic to be worthwhile? The answer depends on the scientific question. My sense is that many of the more interesting questions in ecosystem science are about key processes that determine the similarities and differences within and among systems. For this purpose, simple models can be illuminating, helping to isolate features of ecological systems and assessing whether specific mechanisms are consistent with dynamics. This perspective suggests that simple models help unfold the complexity of ecological
Figure 4.3. Results of model for vegetation and turbidity given parameters typical for a shallow Dutch lake. Solid and dotted lines indicate stable and unstable equilibrium solutions respectively. Figure is redrawn from Scheffer 1998 and used with permission.

systems and provide a pathway for further investigation of critical features (e.g., see Scheffer et al. 2001).

Prediction

Prediction is an important goal for ecology and for modeling in particular (Clark et al. 2001; Pace 2000). Predictions are formally defined as derivations from the logical structure of theory that serve as a means of testing and evaluating theory (Pickett et al. 1994). The means for deriving theoretical predictions is frequently via mathematical models such as the turbidity-vegetation model of shallow lakes discussed above. Predictions are also needed for the application of ecological knowledge to specific situations and to inform the public and policy makers of the likely consequences of current actions and social trends (but see Chapters 2 and 7 for pitfalls and cautions in this context). Uncertainty accompanies most ecological predictions, and the explicit description of uncertainty is another reason that ecological research benefits from a predictive orientation (Clarke et al. 2001; Pace 2001).
The term "prediction" is used informally to describe a number of activities that are associated closely with modeling. These include forecasting and scenario building. At the Cary Conference, there was considerable discussion of what types of predictions could be expected from ecosystem modeling especially with regard to global environmental change. In this context, predictions will be required in a number of different forms, including the consequences of projecting current trends into the future (e.g. Sala et al. 2000; Tilman et al. 2001). Predictions of this nature, however, may tell us little about the real future and may not reduce uncertainty for decision making (see Chapter 7). When entering this arena, ecologists must be clear about the shortcomings and limitations of their models and explicit that changes in human activity will alter the trajectories of many forecasts, such as those related to climate change effects, loss of biodiversity, alterations of habitat, cutting of forests, accumulations of toxins, and degradation of water resources. Forecasting the future precisely is not going to be possible, but evaluating consequences of current trajectories and casting alternate possibilities via scenarios remains a need.

There are many examples where simple models provide useful predictions and clearly demonstrate the uncertainty associated with them; one such example is a study of Lake Mendota, Wisconsin. Because of heavy loading of nutrients from urban and agricultural areas in the watershed, Lake Mendota has long experienced nuisance blooms of cyanobacteria. To assess what level of nutrient loading resulted in blooms, Stow et al. (1997) analyzed a 19-year data set on cyanobacterial dynamics and phosphorus. While the data are noisy, a reasonably good fit was obtained between spring total phosphorus concentrations and summer cyanobacterial biomass with a logistic regression model that assumes error in both variables. Having an established model, its implications can be recast in terms of probabilities of a bloom at specific spring total phosphorus (TP) concentration. Defining a nuisance bloom as exceeding a specific cyanobacterial biomass, the model indicates that there is less than a 5% chance of a bloom at TPs below 0.08 mg L\(^{-1}\), but at TPs of 0.12 mg L\(^{-1}\), the chance of a bloom is greater than 60%. Management to limit blooms might effectively target a spring TP concentration of 0.08 mg L\(^{-1}\) or less.

This example illustrates how a management-relevant prediction can be derived from a simple empirical model. While the model is not mechanistic, it is strongly based on underlying principles about nutrient limitation of phytoplankton in lakes and extensive experience in lake management. The uncertainties of the model are clear and their basis transparent. Moreover, this simple model had an important influence on efforts to study the sources of nutrients deriving from the watershed of Lake Mendota and on management plans to reduce these loadings (S. Carpenter, pers. comm.).

A second problem related to models and prediction concerns how well simple models compare to models that are more complex in their ability to forecast. There are relatively few examples of forecasting in the ecological literature, let alone comparisons of forecasts made by different models (Chapter 2). There are examples from other disciplines, such as a recent model comparison for fore-
casting El Niño–Southern Oscillation (ENSO) events. ENSO events are large-scale ocean-atmosphere interactions that result in the signature condition of a warming of the eastern tropical Pacific Ocean.

Because of the widespread impact of ENSO on global climate, numerous simulation and statistical models forecast this phenomenon. A major event occurred during 1997–1998 for which there were prior model forecasts. Landsea and Knaff (2000) compared these forecasts to assess the “skill” of the models relative to a base model that they viewed as having no skill since it was simply a set of multiple regressions derived from past data. El Niños vary in the rapidity with which they become established, their magnitude and duration, and the rate of decline during the decay phase. The goal of the model comparison for the 1997–1998 event was to determine how well the model forecasts fit the entire dynamic including the onset, duration, and decay of the El Niño.

Landsea and Knaff (2000) compared the models for a number of different forecast lead times. One of their results compared predictions versus observations using a linear correlation coefficient. Models were compared for forecasts that were two seasons ahead (i.e., 6 months). The more complex models were never as good as the base statistical model in this comparison, and some of the forecasts were quite poor (Figure 4.4). Landsea and Knaff made a number of comparisons with various metrics for model comparison (e.g., residual mean square errors). Their primary findings were that all models underestimated warming by at least 50% in 6- and 12-month forecasts, that most models assessed the timing of onset and/or decay of El Niño poorly, and that the complex dynamic simulation models as well as a variety of sophisticated statistical models were no better than a base statistical model derived from historical data. Landsea and Knaff concluded that the best performing model for forecasting the entirety of the very strong 1997–98 El Niño was the base statistical model. Thus, the use of more complex, physically realistic dynamical models does not automatically provide more reliable forecasts.

There is a significant lesson here for ecosystem modeling. The dynamic ocean-atmosphere models used to make El Niño forecasts are well developed and based in many aspects on established and well-understood physical interactions. They are analogous to the best ecosystem models in that they include multiple mechanisms and rich detail. Yet, their predictive ability, at least by the criteria used, was not great. It seems unlikely that in the near term most ecosystem studies will have the data, the resources, or the depth of knowledge about the modeled system that exists for ENSO. Thus, the ability of complex ecosystem models to make predictions is called into question by this example. When prediction is a critical goal, more condensed models of the type I have discussed might often be particularly useful. There are, however, many sources of uncertainty and limits to ecological prediction even for simplest empirical models (see Chapter 8). Nevertheless, there is a developing capacity and clear need for prediction and ecological forecasting, and this will require continued efforts to improve ecosystem models and foster better interaction and communication with decision makers (Clark et al. 2001; Chapter 7).
Figure 4.4. Linear correlation coefficients of predictions versus observations for several model forecasts of ENSO compared to a base model. Models are the singular spectrum analysis/maximum entropy statistical model (SSA/MEM), the Bureau of Meteorology Research Centre simplified dynamic model (BMRC), the Lamont-Doherty simplified model (LDEO), the Center for Ocean-Land-Atmosphere Studies dynamic model (COLA), and a "Base" empirical model. Forecasts are for two seasons ahead. Data are from Table 2 of Landsea and Knaff 2000 for the "Niño 3" region. The reader should consult the original source for details and citations for the models and comparisons abstracted in this figure.

Domain of Simple Models

Simple models are not a tool for all purposes but are useful for the many purposes illustrated in Figure 4.5. The two-way arrows indicate an interaction between models and uses. Perhaps the most important use of simple models is to ask general questions and depict ideas. In this context, models serve as a means of evaluation or testing questions or of exploring and illustrating new or different ideas. One of the great advantages of simple models is that they are easily discarded and replaced by something better (Chapter 25). In many ways, the science of ecology is the choice among models, and alternate simple models can be readily compared. The discussion and examples in the earlier part of this chapter have emphasized the utility of simple models in the areas of pattern and mechanism evaluation as well as prediction. Simple models can also play a role in study design by assessing, for example, an appropriate temporal scale of sampling or the number of ecosystems needed to assess a pattern given an expected trend and error.
Finally, simple models play a critical role in communication. Ecosystem science has many audiences, including scientists in other disciplines, managers, and the public. Simple models can facilitate communication of scientific results and implications to all these audiences. In this regard, simple models of ecosystems and ecosystem processes are critical in communicating the problems of ecosystem research to other scientists, including ecologists who do not focus on ecosystems. The patterns that simple models attempt either to describe (i.e., statistical models) or explain often represent important problem areas. Have ecosystem scientists been effective in broadcasting both their understanding and the remaining uncertainties for such general problems? Consider basic ecology texts, where the familiar general patterns of latitudinal gradients in species diversity, allometric relationships, and other hallmarks of population and community ecology are well represented, but some of the greatest successes of ecosystem studies, such as the generalized responses of aquatic systems to nutrient loading (Smith 1997), are often left out. Ecosystem scientists need to be more proactive in demonstrating these patterns and associated models. Simple models play a critical role in presenting ecosystem science to a wide audience.

Acknowledgments. I thank the Cary Conference organizers—Charles Canham, Jonathan Cole, and William Lauenroth—for encouragement. Helpful comments on a draft version of the paper were provided by an anonymous reviewer.
References


Models in Ecosystem Science

EDITORS

Charles D. Canham

Jonathan J. Cole

William K. Lauenroth

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD