Review TRENDS in Ecology and Evolution Vol.21 No.7 July 2006

A future for models and data in environmental science

James S. Clark[1] and Alan E. Gelfand[2]

[1]Nicholas School of the Environment and Department of Biology, Duke University, Durham, NC 27708, USA
[2]Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708, USA

Together, graphical models and the Bayesian paradigm provide powerful new tools that promise to change the way that environmental science is done. The capacity to merge theory with mechanistic understanding and empirical evidence, to assimilate diverse sources of information and to accommodate complexity will transform the collection and interpretation of data. As we discuss here, we specifically expect a shift from a focus on simple experiments with inflexible design and selection among models that embrace parts of processes to a synthesis of integrated process models. With this potential come new challenges, including some that are specific and technical and others that are general and will involve reexamination of the role of inference and prediction.

Introduction

Be it climate change, exotic invasions, extinction risk, or emerging diseases, the contrast between the simple models and experiments used to learn about ecosystems versus actual ecosystem behavior has never been more apparent. With a growing appreciation of ecosystem complexity has come scrutiny and criticism of traditional models, appeals for synthesis, and frustration with statistical methods used to understand and predict nature [1–3]. Ecosystems are increasingly seen as the product of huge numbers of interacting forces [4,5], food webs being among the few examples where ecologists have attempted to enumerate the complex interactions. Even here, efforts to define a seemingly obvious relationship (whether a species eats another) have been viewed as highly subjective [6]. Influences vary with setting and scale, nonlinearities abound and ecosystem properties emerge from interacting local and global influences.

For many pressing environmental challenges, ecologists do not feel qualified to anticipate ecosystem change [7] and extrapolation from small-scale experiments to relevant settings can be an act of faith. The challenges faced can be expressed in several ways: How do we combine observations that derive from many sources? Can we join what is learnt about parts of a process in isolation (e.g. controlled experiments) in ways that integrate their interactions in nature? How do we connect observations that are specific to location, time and setting with understanding that comes from a diverse body of nonspecific theory? Can we accommodate the uncountable interactions among unseen forces that collectively determine ecosystem behavior? Given that learning requires models, and modeling demands simplification, how is simplicity achieved from such overwhelming complexity?

The long-standing dichotomy of 'statistical' or 'empirical' models versus 'theoretical' or 'mechanistic' models contributes to the challenge of synthesis [8]. 'Statistical modeling' is applied to experiments for purposes of inference. Few ecologists make predictions from such models, appreciating the specific settings and scales from which they derive [1,9] and their 'non-mechanistic' nature. Rather, experimental results are extrapolated to nature in informal ways, as demonstrated by the many examples included in debates on carbon cycling and increasing atmospheric CO_2 concentrations [10].

'Theoretical' or 'mechanistic' modeling is used for understanding and prediction. Simple models with few parameters are analyzed to discover general features of model behavior. Complex models with many parameters are explored with simulation. Parameter values, rather than data, are the inputs, which might be scavenged from the literature, as they are rarely available for the specific contexts in which the models are applied. For example, ecological and conservation journals include lively debate on topics such as predicted time to extinction for endangered species, based on models and/or parameter values that might not apply to the future or to the specific setting [11,12].

The divide between statistical versus theoretical models is partly a reaction to the complexity challenge. Classical statistical practice demands simplicity in terms of strict experimental design and control. Efforts to shoehorn relevant observations into a rigid framework have contributed to debates on classical hypothesis testing [13,14], the importance of controlled experiments [15], pseudoreplication [16] and the expanding application of Bayesian inference [17]. Uncertainty comes with observations [18–20], model parameters [11,21,22], the specification of processes [23,24] and scenarios developed for the exploration of possible futures [7,25]. An abiding consensus cutting across the debates is that proper statistical practice must be the objective arbiter of evidence. But synthesis of general theory, mechanistic understanding

Corresponding author: Clark, J.S. (jimclark@duke.edu).
Available online 19 April 2006

 0169-5347/$ - see front matter doi:10.1016/j.tree.2006.03.016

Box 1. Challenges for modeling

A 'simple' population dynamic model is complicated enough to highlight the challenges for inference and prediction. Consider a single population, the dynamics of which depend on recruitment, growth and mortality. Many population models involve dozens of life stages, but we can illustrate the complexity with only a few relationships. The issues we highlight here apply, regardless of whether a population is modeled as a system of differential equations, difference equations, or transition matrices. For a typical analysis, parameters would be estimated from a range of different studies and plugged into a dynamic model (Figure I). Parameter estimates have different levels of support and are usually obtained from scales and settings that are different from the dynamic model in which they will be used for simulation. The disparate settings for parameter estimates is unavoidable, because observational data are limited and indirect, and experimental manipulations come from controlled settings and scales that do not experience the range of natural influences. Furthermore, the analyses are disconnected from one another. Fecundity and mortality risk can be related to growth rate, but it is hard to integrate these processes. It is also difficult to specify a model for a realistic process, because most aspects of the process are not observed. Uncertainty only enters at the 'data' stage (the sampling distribution, or 'likelihood') and the process model is taken to be deterministic and known.

The biggest casualty in this type of modeling is the estimate of uncertainty. Point estimates from such models can sometimes be reasonable, but estimates of uncertainty are necessarily poor. Given that the confidence envelope is the most important product of model fitting [3], this cost is severe. The problem starts with the fact that there is no probabilistic interpretation of the envelope.

Figure II provides an alternative view of integrated processes, organized as a level in a larger model that could admit many types of data and unknowns at all stages. Because 'arrows' are stochastic, we learn about the processes and state variables simultaneously. We admit relationships among them, by modeling their interactions. The interactions are partly known in advance, information that enters as functional forms and prior estimates of parameters. The balance between deterministic versus stochastic contributions throughout the model reflects the degree of current understanding. Upon updating or 'fitting the model', we learn from posterior parameter distributions and criteria that reflect overall model performance. The novelty here is treatment as a single integrated process. The demographic interactions can be carried forward in stand simulators to predictive distributions of dynamics [32].

Some ecologists might ask how Figure II differs from large ecosystem models that share this box-and-arrow presentation. Ecosystem ecologists have long been adept at thinking 'globally', but beginning with the same reductionism used to construct Figure II. There are many complex simulation models in use for biogeochemical cycling [37] to numerical weather prediction [55], including those used to inform the general public. Until recently, these models have been almost exclusively deterministic, with stochastic outputs generate by stochastic inputs. In ecosystem models, simulators are spun forward based on fixed parameter values. The key innovation is making such models inferential, starting with a similar graphical structure, but using it to construct algorithms to both infer and predict.

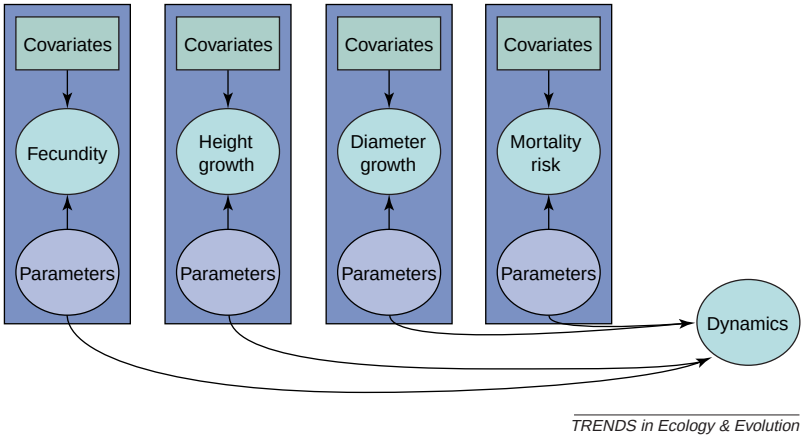

Figure I. The traditional data modeling approach, where four independent analyses are used to obtain estimates of demographic rates that could be taken as inputs to a dynamic model. The demographic process is effectively 'disaggregated' into many separate analyses. The dynamic model is typically applied to different scales or settings from that from where the parameter values were obtained.

and heterogeneous data can be among the casualties of efforts to satisfy rigid statistical assumptions [8].

Now, in the early 21st century, a profound change is afoot in the perception of what can be learned from data and models. The developing modeling tools and growing appreciation of complexity are complementary trends that can take us from the current disaggregated study of process components toward a more integrated understanding of process behavior. We believe that multilevel modeling using graphical approaches provides the general platform. The familiar 'box-and-arrow' diagram, previously used for forward modeling (taking parameter values, rather than data, as inputs), will take on an expanding role, organizing model building and computation for inference and prediction. This transition is already underway in settings as diverse as weather and climate, environmental health, gene expression and financial markets, settings where the scientific community has already recognized the need for a more global perspective of interacting processes.

Here, we summarize elements of the move beyond the focus on rigid experiments for abstracted ecosystems and the machinery that makes it possible. We begin with some of the recognized model limitations that have prepared ecologists for change and then discuss why the most powerful current approach to synthesis requires the combined advantages of graphs and Bayesian inference. Finally, we summarize some of the emerging challenges, both technical and conceptual.

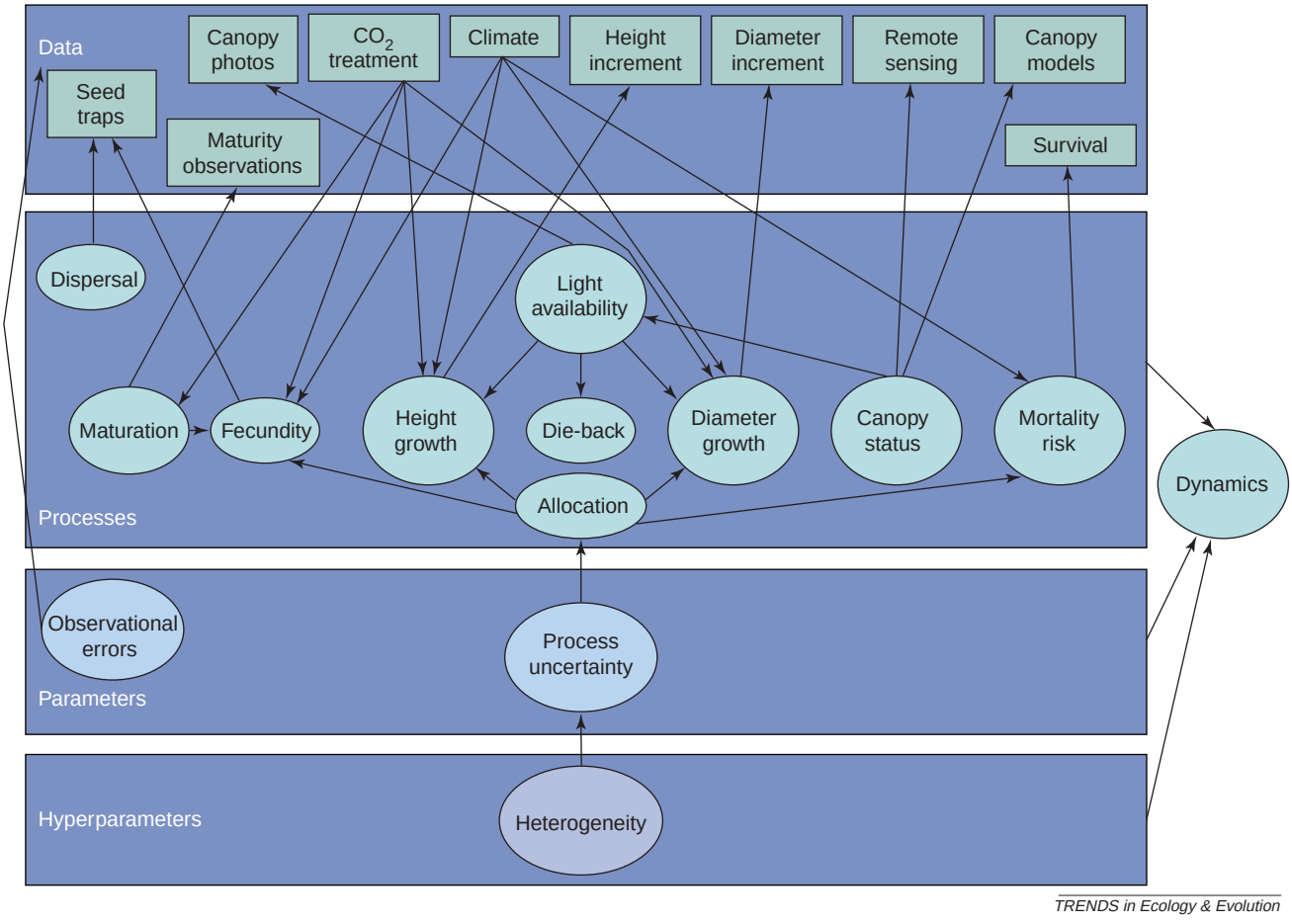

Figure II. An integrated process model that could embrace the same elements as Figure I. Rather than four separate analyses, followed by prediction, inference on the integrated process is done in a single step, along with dynamic prediction. In effect, data, rather than parameter estimates, are the inputs. Estimates and predictions are the outputs.

Where we have come from: limitations of tradition

The tradition in ecology of rigid design owes much to modeling limitations. Although ecologists have long used and valued data from many sources (e.g. [26–28]), a successful advocacy for simple experiments during the 1980s soon made them the dominant research tool. This emphasis reflects an appreciation for the importance of data that meet model assumptions. The understandable preoccupation with design involves considerations such as independent observations, full knowledge of predictor variables and the need for controlled manipulation. Departures from restrictive assumptions have regularly brought dispute over 'misuse of statistics', definitions of treatment and controls, and the 'correct' test or the 'best' interval estimate. Loose interpretation of assumptions for nonexperimental data was inevitable as models with relevant assumptions were not available. Such inflexible designs address simple, typically local, relationships between variables, but provide only a highly abstracted view of most environmental process [29,30].

As a concrete example, take the common challenge of understanding density dependence and the effects of natural enemies, climate variation and management of populations. Elements of an analysis might look similar to Figure I in Box 1, where demographic rates are estimated using statistical models and plugged into theoretical models to calculate dynamics. Error in parameter estimates might be propagated to uncertainty in growth. A piecemeal approach to parameterization is standard practice, regardless of specific assumptions about interactions, time and space. Several limitations to such an approach are clear:

- No synthesis of data. In classical data modeling, there is a single data set. Deterministic treatment at the process stage precludes the assimilation of multiple data sets, because correlations among observations require stochasticity in the underlying process from which they emanate. Although 'data assimilation' and various 'inverse models' increasingly involve multiple data sets, a classical framework does not treat them coherently as it does not begin with a joint distribution of the unknowns [31].
- No synthesis of processes. We cannot 'learn' about processes that are assumed to be deterministic. Instead parameters are estimated under the assumption that the process is known. Because each process is modeled separately (e.g. Figure I, Box 1), there is no evaluation

of interactions among them. Inference devolves into an inefficient head-to-head competition among specific models, as opposed to more direct learning about the integrated model. Rather than asking about a process in which both A and B occur (and interact), standard practice is to 'select' between two models, one for A, the other for B.

- No synthesis of theoretical understanding. Theoretical knowledge is diverse, general and cannot be fully described by simple deterministic equations. Structures are needed that enable the introduction of what is already known to all parts of the model. Because theory is general and limited, knowledge and uncertainty are two sides of the same coin.
- No place for complexity. Complex interactions cannot be specified deterministically. Models require a place for the complications that arise at all levels, including within the data and mechanistic processes. Here again, stochasticity is needed to summarize complex relationships.

Clearly, the answer to these challenges entails merging the traditionally disparate treatment of 'statistics' and 'theory', bringing inference into the same models that are used to promote understanding [21,32]. Learning is the goal of both model types, and we cannot hope to separate that which comes from observations from the insight available through other sources. Here, we summarize promising new directions for answering these challenges.

Combining the advantages of graphs and Bayes

Emerging machinery for synthesis combines graphical modeling in hierarchies with Bayesian inference (Box 1). The growing appreciation of ecosystem complexity is fully compatible with this framework. Neither by itself is sufficient, as evident from past, independent, applications of graphs and Bayesian inference. Ecologists have long used graphs to represent everything from food webs [33,34], to life cycles [35,36], to ecosystem function [37], to landscapes [38]. The graph includes boxes, or 'nodes', connected by arrows that represent causal relationships or 'effects' (Box 1). Directed graphs use nodes to denote process variables (observed, unobserved or conceptual) along with arrows to indicate causality (shown by the direction of the arrows) among the variables. This reductionist approach, beginning with local building blocks, organizes models for analysis and simulation. Graphical models were traditionally deterministic: given the inputs to a node, the response is determined. However, classical interference cannot readily exploit this structure. Graphical models could be made stochastic, but without Bayesian inference, there typically was no strong motivation to do so.

Until recently, Bayesian analysis also had a long history of limited application. 'Simple Bayes' begins with a prior distribution for one or more parameters to be updated by data. The posterior distribution of parameters requires integration of the data and prior, which is simple only for simple models. Many researchers embraced the logic of proceeding from what is known, assimilating new information and then moving to an updated understanding [39]. However, without tools for practical application, the Bayesian paradigm was never a serious threat to classical methods, which, by the late 20th century, were entrenched in all widely available software.

From the convergence of graphs and Bayes comes powerful new machinery. A key breakthrough came with adaptation of Gibbs sampling to the Bayesian paradigm [40]. Not until this simulation tool was brought to bear on the posterior distribution was it possible to analyze most high-dimensional problems with a common approach. The building blocks, or conditional distributions used for Gibbs sampling, are best represented by graphs. Computer scientists interested in causality also discovered this powerful combination as an approach for 'machine learning' [41]. In statistics, the term 'hierarchical Bayes' is commonly used, and 'Bayesian network' evokes the graphical model upon which the algorithms are based.

Application to inference involves the introduction of stochasticity at each node of the graph, as discussed by Berliner [42] and, in an ecological context, by Clark [3]. The overarching hierarchical structure conceives three stages: data given process and parameters, process given parameters, and parameters. Regardless of where the complexity lies, the joint distribution of everything unknown is organized within the graph (Figure II, Box 1) and factored as the basis for algorithm development. The response of a node is uncertain and, thus, probabilistic. Introduction of stochasticity is liberating as we need not be 'correct' at each node. Rather, we anticipate variable responses to inputs, or a limited understanding of the mechanism, or we might desire a simplified specification. Because the stochastic world includes the deterministic world as a special case (zero uncertainty), the approaches are compatible. The better we understand local behavior the 'more causal' is the local submodel. Graphs and Bayes go hand-in-hand, yielding a probability model for the full (joint) density of all unknowns. Because of its fundamental reliance on probability, Bayes provided the natural framework for analysis.

Graphs and Bayes enables the introduction of all sources of information, theoretical, mechanistic and empirical (which, in fact, might have come from designed experiments), and is also flexible. Local specifications can be modified by the addition, removal, or relocation of arrows, summarizing known (deterministic) and unknown (stochastic) elements. The focus changes from that of which inferential procedure might apply to a specific design to modeling for integration, and there is an increasing number of examples from the environmental literature (e.g. [5,20,43–49]). However, the new flexibility brings with it some challenges.

Current and near-term challenges

Graphical modeling enables us to contemplate an enormous number of potential models (boxes and arrows in the graph), introducing at least two challenges: (i) how to fit complex models; and (ii) how to explore the model space with regard to adequacy (i.e. consonance with the data) and model selection (comparison of models). Owing to both model and data complexity, wholly satisfying answers are not to be expected, but they remain valuable targets.

For the first challenge, simulation is emerging as the current answer, although not as it has been applied in the

past. The traditional forward-simulation approach in ecology begins with a box-and-arrow diagram similar to that shown in Figure II, Box 1, and prescribes inputs, and produces output. Model selection is necessarily ad hoc; in not directly modeling data, inference is limited to the summarization of outputs.

The alternative approach models the data together with the rest of the graph. All variables, observed or not, are stochastic. Bayes' theorem provides a conditional (posterior) distribution for all unknowns given all observations. Although easily stated and compactly expressed probabilistically, learning about the posterior distribution of unknowns (i.e. the fit of the model) is challenging. When analytical investigation is unproductive, we can still sample from the posterior. This is simply the first principle of inference: to learn about a population, we sample from it, and the more we sample the more we learn. Conditional simulation techniques of the past 15 years [39,50,51] include Markov chain Monte Carlo approaches, as well as direct simulators using (sequential) importance sampling and particle filters [52,53] in addition to Gibbs samplers. These tools are powerful, but they cannot be efficiently applied to arbitrarily complex problems. For example, the intermediate products of an analysis might not fit in the main memory of the computer. Or the data stream might arrive faster than the model converges, a situation that is likely to arise with the expanding use of sensor network data*. Efficient computation (algorithms and data structures) thus remains an active area of research.

Owing to the complexity of the models and the novelty of developing tools, issues related to model determination require reexamination. Model adequacy is an absolute question about a given model, whereas model comparison is relative, requiring at least two models. Model adequacy is an inherently informal exercise as it is essentially impossible to 'calibrate' how well a complex process model 'fits' the data. Indeed, we might believe more in the model specification than in the data that we obtain. Especially in ecological models, data are notoriously noisy, signals might be weak and confidence in the specification might far exceed the information in the data.

Model comparison is equally problematic. Model selection depends on the intended use for the model, but requires subjective choice of a utility function [48,49,54]. More worrisome is the distillation of a complex model down to a single number used for comparison. Model performance must be considered broadly, for example, beyond the data at hand. As noted here, even if we could achieve agreement on a model selection criterion, the space of possible models will be sufficiently huge to preclude exhaustive search. How do we find the 'best' (or at least 'good') models in such a setting? Model determination is thus also an area of active research.

A future of (overly) complex models?

The emerging, synthetic view of processes, information and uncertainty represents a crucial transition, one that promises to change not only the analysis of data, but also the way in which information is perceived, collected and interpreted. The computational tools that have made complex simulation models familiar in ecology (often with limited connection to data) have now developed into a synthetic framework. However, experimental approaches will not diminish in value; on the contrary, we expect them to become all the more powerful, as flexible modeling enables us to accommodate unexpected changes in the design and availability of information [29]. The challenges noted here should not deter the use of developing tools. Rather, they reflect emerging possibilities that demand a broad overhaul of statistical inference in complex settings.

The fact that complexity often demands sophisticated models (and that we can now work with them) should not be interpreted as advocacy for ever-larger models. Where simple models work, they should be used. The prospect that ecological models will increasingly look more like Figure II than like any one of the components in Figure I in Box 1 was discussed by Clark [3] and can benefit from some elaboration here. Whereas Figure II (Box 1) appears bewildering complicated, it is important to emphasize that it substitutes for (i.e. integrates) all four boxes in Figure I (Box 1). Concern that this synthetic view can make it hard to identify the many parameters also requires context. True, we are estimating more unknowns in Figure II (Box 1), but the stochastic representation allows us to assimilate much more information. There are far more 'anchors' for the fit in Figure II (Box 1), represented by the many types of information that we can admit, including both data and mechanistic understanding.

Finally, it can be easy to mistake the introduction of stochasticity as a form of runaway model complexity. True, each stochastic element can bring parameters. But the stochastic terms simplify models by enabling much to go unspecified. Palmer *et al.* [55] provide an explicit example in numerical weather prediction. Stochasticity stands in for complexity. By making models stochastic and inferential, complexity can be represented in ways that vastly simplify modeling. Taken together, the challenge of complexity and the melding of graphs and Bayesian inference promise to change ecological science profoundly.

* Flikkema, P.G. *et al.* Model-driven dynamic control of embedded wireless sensor networks. Proceedings of the 6th International Conference on Computational Science, Workshop on Dynamic Data Driven Application Systems, Reading, UK (in press).

References

1 Carpenter, S.R. (1996) Microcosm experiments have limited relevance for community and ecosystem ecology. *Ecology* 77, 677–680
2 Hilborn, R. and Mangel, M. (1997) *The Ecological Detective*, Princeton University Press
3 Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecol. Lett.* 8, 2–14
4 Levin, S.A. (1998) Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* 1, 431–436
5 Clark, J.S. *et al.* (2004) Fecundity of trees and the colonization-competition hypothesis. *Ecol. Monogr.* 74, 415–442
6 Paine, R.T. (1988) Road maps of interactions or grist for theoretical development? *Ecology* 69, 1648–1654
7 Clark, J.S. *et al.* (2001) Ecological forecasts: an emerging imperative. *Science* 293, 657–660
8 Clark, J.S. (2006) *Models for Ecological Data*, Princeton University Press
9 Srivastava, D.S. *et al.* (2004) Are natural microcosms useful model systems for ecology? *Trends Ecol. Evol.* 19, 379–385
10 Schlesinger, W.H. (2006) Global change ecology. *Trends Ecol. Evol.* 21, doi:10.1016/j.tree.2006.03.015

11 Ludwig, D. (1999) Is it meaningful to estimate a probability of extinction? *Ecology* 80, 298–310
12 Morris, W.F. and Doak, D.F. (2002) *Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis*, Sinauer Associates
13 Roughgarden, J. (1983) Competition and theory in community ecology. *Am. Nat.* 122, 3–21
14 Strong, D.R. (1983) Natural variability and the manifold mechanisms of ecological communities. *Am. Nat.* 122, 56–80
15 Connell, J.H. (1980) Diversity and the coevolution of competitors, or the ghost of competition past. *Oikos* 35, 131–138
16 Hurlbert, S.H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54, 187–211
17 Dixon, P. and Ellison, A.M. (1996) Introduction: ecological applications of Bayesian inference. *Ecol. Appl.* 6, 1034–1035
18 Carpenter, S.R. *et al*. (1994) Fitting predator–prey models to time series with observation errors. *Ecology* 75, 1254–1264
19 de Valpine, P. and Hastings, A. (2002) Fitting population models incorporating process noise and observation error. *Ecol. Monogr.* 72, 57–76
20 Clark, J.S. *et al*. (2003) Coexistence: how to identify trophic tradeoffs. *Ecology* 84, 17–31
21 Lande, R. (1988) Demographic models of the Northern Spotted Owl (*Strix occidentalis caurina*). *Oecologia* 75, 601–607
22 Pacala, S.W. *et al*. (1996) Forest models defined by field measurements: estimation, error analysis, and dynamics. *Ecol. Monogr.* 66, 1–44
23 Pascual, M.A. and Kareiva, P. (1996) Predicting the outcome of competition using experimental data: maximum likelihood and Bayesian approaches. *Ecology* 77, 337–349
24 Peterson, G.D. *et al*. (2003) Uncertainty and the management of multistate ecosystems: an apparently rational route to collapse. *Ecology* 84, 1403–1411
25 Carpenter, S.R. (2006) Ecology for transformation. *Trends Ecol. Evol.* 21, doi:10.1016/j.tree.2006.02.007
26 Diamond, J.M. (1975) Assembly of species communities. In *Ecology and Evolution of Communities* (Cody, M.L. and Diamond, J.M., eds), pp. 342–444, Harvard University Press
27 Gauch, H.G., Jr. (1982) *Multivariate Analysis in Community Structure*, Cambridge University Press
28 Manly, B.F.J. (1997) *Randomization, bootstrap and Monte Carlo Methods in Biology* (2nd edn), Chapman & Hall
29 Caswell, H. (1988) Theory and models in ecology: a different perspective. *Ecol. Model.* 43, 33–44
30 Clark, J.S. and LaDeau, S.L. Synthesizing ecological experiments and observational data with Hierarchical Bayes. In *Applications of Computational Statistics in the Environmental Sciences: Hierarchical Bayes and MCMC Methods* (Clark, J.S. and A. Gelfand, A., eds), Oxford University Press (in press)
31 Wolpert, R.L. (1995) Comment on Inference 'from a deterministic population dynamics model for bowhead whales' by A.E. Raftery, G.H. Givens and J.E. Zeh. *J. Am. Stat. Assoc.* 90, 426–427
32 Govindarajan, S. *et al*. A scalable model of forest dynamics. Proceedings of the ACM Symposium on Computational Geometry (in press)
33 Paine, R.T. (1980) Food webs, linkage, interaction strength and community infrastructure. *J. Anim. Ecol.* 49, 667–685
34 Briand, F. and Cohen, J.E. (1984) Community food webs have scale-invariant structure. *Nature* 307, 264–266
35 Werner, P.A. and Caswell, H. (1977) Population growth rates and age versus stage-distribution models for teasel (*Dipsacus sylvestris* Huds.). *Ecology* 58, 1103–1111
36 Shugart, H.H. (1984) *A Theory of Forest Dynamics*, Springer-Verlag
37 Parton, W.J. *et al*. (1987) Analysis of factors controlling soil organic levels of grasslands in the Great Plains. *Soil Sci. Soc. Am. J.* 51, 1173–1179
38 Bunn, A.G. *et al*. (2000) Landscape connectivity: a focal-species analysis using graph theory. *J. Env. Manage.* 59, 265–278
39 Lindley, D.V. (1990) The present position in Bayesian statistics. *Stat. Sci.* 5, 44–89
40 Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85, 398–409
41 Pearl, J. (2002) *Causality: Models, Reasoning, and Inference*, Cambridge University Press
42 Berliner, L.M. (1996) Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods* (Hanson, K.M. and Silver, R.N., eds), pp. 15–22, Kluwer Academic
43 Agarwal, D. *et al*. Relating tropical deforestation and population pressure through spatially explicit misaligned bivariate regression models. *Ecol. Model.* (in press)
44 Ver Hoef, J.M. and Frost, K.J. (2003) A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska. *Env. Ecol. Stat.* 10, 201–219
45 Sauer, J.R. and Link, W.A. (2002) Hierarchical modeling of population stability and species group attributes from survey data. *Ecology* 83, 1743–1751
46 Wikle, C.K. (2003) Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* 84, 1382–1394
47 Gelfand, A.E. *et al*. (2005) Explaining species distribution patterns through hierarchical modeling. *Bayesian Anal.* 1, 41–92
48 Okuyama, T. and Bolker, B.M. (2005) Combining genetic and ecological data to estimate sea turtle origins. *Ecol. Appl.* 15, 315–325
49 Ogle, K. *et al*. (2004) Reconstructing plant root area and water uptake profiles. *Ecology* 85, 1967–1978
50 Spiegelhalter, D.J. *et al*. (2002) Bayesian measures of model complexity and fit (with Discussion). *J.R. Stat. Soc. B* 64, 583–639
51 Gelman, A. *et al*. (1995) *Bayesian Data Analysis*, Chapman & Hall
52 Robert, C.P. and Casella, G. (1999) *Monte Carlo Statistical Methods*, Springer-Verlag
53 Liu, J.S. (2001) *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag
54 Gelfand, A.E. and Ghosh, S.K. (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika* 85, 1–11
55 Palmer, T.N. *et al*. (2005) Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* 33, 163–193