Ecology Letters, (2005) **8**: 2–14 doi: 10.1111/j.1461-0248.2004.00702.x

IDEAS AND PERSPECTIVES

Why environmental scientists are becoming Bayesians

James S. Clark
Nicholas School of the Environment and Department of Biology, Duke University, Durham, NC 27708, USA
Correspondence: E-mail: jimclark@duke.edu

Abstract

Advances in computational statistics provide a general framework for the high-dimensional models typically needed for ecological inference and prediction. Hierarchical Bayes (HB) represents a modelling structure with capacity to exploit diverse sources of information, to accommodate influences that are unknown (or unknowable), and to draw inference on large numbers of latent variables and parameters that describe complex relationships. Here I summarize the structure of HB and provide examples for common spatiotemporal problems. The flexible framework means that parameters, variables and latent variables can represent broader classes of model elements than are treated in traditional models. Inference and prediction depend on two types of stochasticity, including (1) *uncertainty*, which describes our knowledge of fixed quantities, it applies to all 'unobservables' (latent variables and parameters), and it declines asymptotically with sample size, and (2) *variability*, which applies to fluctuations that are not explained by deterministic processes and does not decline asymptotically with sample size. Examples demonstrate how different sources of stochasticity impact inference and prediction and how allowance for stochastic influences can guide research.

Keywords

Data modelling, Gibbs sampler, hierarchical Bayes, inference, MCMC, models, prediction.

Ecology Letters (2005) 8: 2–14

INTRODUCTION

Ecologists are increasingly challenged to anticipate ecosystem change and emerging vulnerabilities (Costanza *et al.* 1997; Daily 1997; Carpenter 2002; Peterson *et al.* 2003; Pielke & Conant 2003). Efforts to predict ecosystem states highlight long-standing obstacles that apply not only to forecasts, but also to the seemingly pedestrian practice of inference. There is growing awareness of how difficult it can be to connect predictive intervals obtained from models back to the data that went into their construction.

Models of nature, including experimental ones, routinely entail dilemmas: simplify the research problem in the interest of generality, or admit the complexity to attain some realism. The tradeoffs are well known. On the one hand, simple experiments may not 'scale' to nature – the settings where we would like to apply them lie outside the experimental conditions. They engage situation-specific and scale-dependent effects (Levin 1992; Carpenter 1996; Skelly 2002). Likewise, simple models may not accommodate the range of influences that can operate in different settings and at different scales. On the other hand, complicated experiments are rarely feasible (Caswell 1988). Complicated models have traditionally been specific, containing many deterministic relationships and associated parameters. Specifying many 'effects' in models leads to overfitting, in the sense that we can fit this data set, yet promise little predictive capacity for the next. Arguably few ecological predictions with generality claim strong empirical support.

Stochasticity is central to the complexity dilemma, because it encompasses the elements that are uncertain and those that fluctuate due to factors that cannot be fully known or quantified. Decisions concerning what will be treated deterministically in models, what is assumed stochastic, and what can be ignored are the basis for model

and experimental design (Pearl 2002; Clark & LaDeau 2004; Dawid 2004). Such decisions define the complexity of the problem and, thus, the dimensionality of models.

Complexity and scale challenges translate directly to prediction. Despite a long research tradition on demography, population dynamics, and species interactions, ecologists frequently have little to offer when confronted with pending climate-forced range shifts, fragmentation, design of reserves, and where and when biodiversity loss is likely to have feedback effects on ecosystem services (e.g. Emlen *et al.* 2002; Ellner & Fieberg 2003). When challenged for answers, there is temptation to abandon all pretence of prediction and fall back on scenarios that are loosely linked to data. Scenarios can help inform decisions (Clark *et al.* 2001; Carpenter 2002). But the data and models that fill scientific journals might have a more direct role. The capacity to more directly apply ecological understanding should be a compelling justification for research.

Advances in computational statistics of the last decade have produced new tools for inference and prediction. Whereas modelling constraints have long caused most of the complexity to be ignored, new hierarchical Bayes (HB) structures bring flexibility. Developing sampling-based algorithms allow for analysis of complex models. Ecologists will soon become consumers of new techniques, if not practitioners, and insightful interpretation will rely on grasp of terms and basic assumptions. In this paper, I describe the underlying structure of HB that can be exploited for a broad range of ecological problems. I attempt to clarify some of the motivation for Bayesian approaches and some concepts that are often vague or contradictory, even in statistics literature. Because techniques have developed in parallel across many disciplines, the literature on uncertainty is vast and diffuse. Several recent overviews include Pearl (2002), Halpern (2003) and Zidek (unpublished data). Some discussion from an ecological perspective is contained in Dixon & Ellison (1996), Hilborn & Mangel (1997), Clark (2003), and Ellison (2004). Even a cursory summary of the 'uncertainty literature' is beyond the scope of this paper. While HB is clearly not the only way to address uncertainty, it stands out as an approach that can accommodate complex systems in a fully consistent framework. That is, once there is a model, only accepted rules of probability take us from data to inference to prediction.

For context, I begin by outlining a different perspective on the relationship between classical and Bayesian inference from the one that is pervasive in most ecological discussions of the subject. I then move to a simple overview of HB structure, demonstrating capacity to efficiently exploit and combine information from multiple sources, and to generate predictive intervals that are compatible with the process and unknowns in ecological models. Space does not permit a full review or 'how-to' treatment of HB, principally because it demands some sophistication with distribution theory. Rather, I focus on basic concepts that distinguish it from the classical approach and from simple Bayes (SB), and I use examples to highlight how those differences affect inference and prediction. Finally, I mention that, as complexity increases, computational issues emerge as an important challenge.

PHILOSOPHY AND PRAGMATISM

Previous reviews of Bayesian inference in the ecological literature have emphasized philosophical differences between classical approaches and SB[1]. A comparison of results obtained for a process model analysed under the frameworks of classical vs. SB is often provided to emphasize the differences. I will only say enough about these familiar topics to motivate why I do not dwell on them and move immediately to HB.

First, I do not believe that philosophical issues are the principal motivation for the direction of modern statistical computation. The philosophical emphasis of ecological writings on this subject do not raise new issues, but rather rework long-recognized ones available in many texts on Bayesian inference. Debates in the statistical literature were especially lively in the 1960s through early 1990s, but have been around much longer. Ecological debates on this topic cover the same ground, reviewing the different concepts of probability, how these views manifest themselves in classical and Bayesian inference, with attention to differences and points of potential controversy. This emphasis can leave the impression that philosophy determines approach. Those leaning towards a subjective probability view go Bayesian; a frequency view of probability steers one in a classical direction.

This importance of philosophy seems to be reinforced by examples aimed at demonstrating, comparing, and/or contrasting classical vs. SB. Such examples show that, with similar underlying assumptions (e.g. vague priors), classical methods and SB yield near identical confidence envelopes – with a lot more work, one can expect to obtain a Bayesian credible interval (CI) that is not importantly different from a classical confidence interval. (I use this term 'confidence envelope' when I do not care to distinguish between a classical 'confidence interval' and a Bayesian CI). If one gets essentially the same answer, philosophy must be the motivation. Despite the counsel to start from one's view of probability, ecologists having no philosophical axe to grind might question the point; if Bayes requires more work to arrive at the same interpretation, why bother with Bayes?

[1]By 'simple Bayes', I mean a minimal model of likelihood and prior. An example will follow in the next section.

Of course, there is more to discuss, classical hypothesis testing and the role of priors being central. And the importance of philosophy should not be understated. The tension between the need for objectivity in science and the inevitable subjectivity of statistics (Berger & Berry 1988) poses a serious challenge for philosophers and statisticians. The philosophical issues can be deeply metaphysical and bear on the very nature of probability (Dawid 2004 provides a recent perspective). The points I emphasize here are (1) that these issues are not new, (2) that they will be alive and well long after HB pervades a staggering breadth of scientific disciplines, and (3) that philosophy has little to do with the transformation in statistical computation that has emerged over the last decade. No philosophical war has been won to support an emerging consensus. Subjective or 'personal' probability does not now occupy the high ground once held by a frequency interpretation of probability. Rather, there is a growing appreciation of the compatibilities between frequentist and subjective probability views (reviewed in Clark 2004). Those already possessing a healthy scepticism of classical hypothesis tests can continue to estimate classical confidence intervals without offending many Bayesians.

The focus of this paper stems from an alternative view that the emergence of modern Bayes has little to do with philosophy, but comes rather from pragmatism. I have previously discussed how, from a pragmatic standpoint, one could find more to contrast between HB vs. both SB *and* frequentist views (Clark 2003). In that example, I demonstrated why frequentist and Bayesian assumptions about parameters are more similar than is apparent from many ecological writings on the subject. Although Bayesians speak of parameters as 'random', and frequentists do not, SB shares with classical approaches the assumption that there is an underlying 'true' parameter value that is incrementally approached with increasing sample size, in the same way and at the same rate as obtained with a frequentists confidence interval. HB relaxes this assumption in the sense that a 'parameter' can vary. I say more about this in the next section.

A growing number of practitioners are willing to let the choice between frequentist and Bayes rest on complexity. Simple problems are most readily analysed with standard software options. Although Bayesians will not pay much attention to the *P*-values, they may not advocate unnecessarily complicating the model when something off the shelf will do. While preferring the concept of a prior-likelihood-posterior update cycle in many applications, I find instances where a frequency interpretation of probability is sensible. The power of HB comes from a capacity to accommodate complexity. Clark *et al.* (2003) used a specific example to demonstrate how HB allows for errors in variables, random effects, hidden variables, and multiple data sets at different scales. Yet, for a highly simplified version of the model, SB yielded results that did not importantly differ from a classical implementation.

In the present paper I discuss the general framework of HB that makes it powerful. Rather than focus on specific examples, I emphasize the general framework and then discuss several examples within the context of this framework.

FROM SIMPLE TO HB MODELS

Hierarchical Bayes transformed computational statistics in the 1990s providing a framework that can accommodate nearly all high-dimensional problems (Gelfand & Smith 1990; Carlin & Louis 2000). The structure is so flexible that it not only opens doors to complex problems informed by messy data. The intuitive approach facilitates a deeper understanding of the processes and the importance of treating unknown elements in appropriate ways. In this section I provide a brief overview of the framework with two examples.

A decomposition for complexity

Consider a model that involves equations with parameters θ. A traditional analysis involves embedding a deterministic process model within a stochastic shell (e.g. a 'sampling distribution') to allow for the scatter in data that the process model cannot account for. Together this deterministic process and its stochastic shell comprise the likelihood function. We write the probability for a data set y as $p(y|\theta)$, the probability of obtaining data y under the assumption that they are generated by a model containing parameters θ. The likelihood function is central to most methods of inference. However, by itself, it cannot accommodate complex relationships.

Hierarchical Bayes accommodates complexity by allowing us to dissect the problem into levels. The likelihood still plays a prominent role, providing a first level, $p(y|\theta) \equiv p(\text{data}|\text{process,parameters})$. The likelihood can be our 'data model', which is conditioned on a process model and on parameters. We can write models for them, $p(\text{process}|\text{process,parameters})$ and $p(\text{all parameters})$. The full model is a joint distribution of unknowns, which includes parameters (and latent variables),

$$p(\text{parameters}|\text{model,data}) \propto p(\text{data}|\text{process,data parameters}) \times p(\text{process}|\text{process parameters}) \times p(\text{all parameters})$$

This decomposition comes from Berliner (1996), and has been adopted in a number of recent applications (Wikle *et al.* 2001; Clark *et al.* 2003, 2004; Clark 2003; Wikle 2003a).

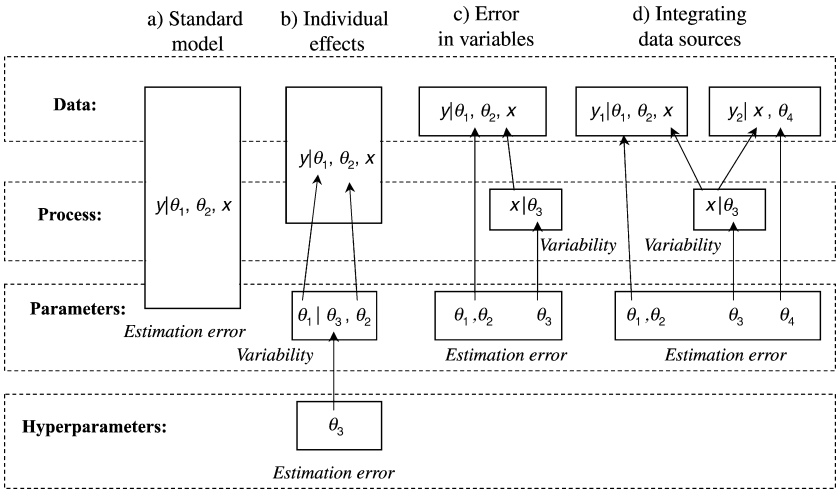

Figure 1 Four examples of how the Bayesian framework admits complexity (see text). Models can be viewed as networks of components, some of which are known and many unknown. The stages shown here, Data, Process, Parameters, and Hyperparameters, represent an overarching structure that admits complex networks. A model might include structure in space, time, or among individuals or groups (b), hidden processes (c,d), and multiple sources of information that bear on the same process (d). Acknowledging variability in a 'parameter' θ_1 (b) is accomplished by conditioning on additional parameters (θ_3). Now θ_1 occupies a middle stage and is truly variable, not just uncertain; θ_3 is asymptotic. Acknowledging variability in a predictor variable x (c) is accomplished in the same fashion.

It represents but one of many ways in which we might care to structure a model under HB. We now have several levels – if the likelihood represents a single stage in the traditional framework (Fig. 1a), HB might have elements organized over various levels (Fig. 1b–d). We can think of a network, suggesting application of graphical methods. Consider some of the ways in which the network can develop, beginning with SB, followed by straightforward extension to complex problems.

Simple Bayes

Beginning with SB, a model with two parameters (θ_1, θ_2) looks like this

$$p(\theta_1, \theta_2|x, y) \propto p(y|\theta_1, \theta_2, x) \quad \{\text{likelihood (data model)}$$
$$\times\, p(\theta_1)p(\theta_2) \quad \{\text{prior (parameter model)}$$

where $p(\,)$ represents a distribution or density. On the left-hand side is the posterior, with a vertical bar that separates unknowns or 'unobservables' to the left and 'knowns' to the right. This is the posterior distribution of unknowns. Unknowns are assigned distributions – unknowns are stochastic. On the right-hand side, there is no distribution for x, because this predictor variable is assumed to be known (the standard assumption). The response variable y is treated as known on the left-hand side, because it is already observed. On the right it is stochastic. There are distributions for the parameters. The priors allow for uncertainty; they are stochastic only in the subjective probability sense (see below).

Parameters θ_1 and θ_2 need not be independent *a priori*. For example, *a priori* assumptions about the joint density $p(\theta_1,\theta_2)$ can increase efficiency or allow for known dependencies. However, we may assume independence, in which case the data will determine their posterior relationship. If we have no external insight, we can make them rather flat. Such 'non-informative' priors are used to allow that data dominate the answer, rather than a combination of data and external information. Priors can admit information that does not enter through the likelihood. Asymptotics apply at the lowest stage – as sample size increases, uncertainty about the true values of θ_1 and θ_2 declines to zero. Hereafter, when I speak of 'asymptotics' I mean it in this sample-size sense: the larger the sample size, the narrower the confidence envelope.

The foregoing is the SB framework, and the credible interval that results may not substantially differ from a traditional confidence interval. For example, if priors are relatively flat, we would obtain parameter estimates close to those coming from a classical analysis. There can be important exceptions to this, but they apply primarily to complex models.

Considerable confusion remains over the interpretation of parameters in the Bayesian context. The distinction between classical and SB usually emphasized is the 'random' nature of a Bayesian parameter. As I discussed in the context of demographic change, the *estimate* is random, not the parameter itself (Clark 2003). The posterior density

describes *uncertainty*, not *variability* or *fluctuation* in the sense that ecologists use these terms. This is stochasticity in the subjective probability sense. The stochastic treatment of the *estimate* allows 'learning' as, say, data accumulate. Yet SB shares with the frequentist approach the assumption that the parameter itself is a fixed constant. This shared assumption that the true 'parameter' (as opposed to the estimate of it) is fixed makes classical and SB more compatible than is commonly believed. For simple problems (including non-informative priors), we can expect them to provide us with essentially the same confidence envelopes – they are subject to the same asymptotic relationships with sample size n. Then why bother with Bayes?

At this stage, we might make a philosophical case for a classical or Bayesian approach. Or we could argue pragmatically that available software makes a classical treatment more expedient. If a confidence envelope is the goal, we could claim that Bayes provides no practical advantage for such simple problems. The limitations of classical approaches become formidable as we move beyond simple problems, because they place the full burden on the likelihood. SB is likewise limited. We have added priors, which allow us to learn about parameters, but we are stuffing everything else into the likelihood.

The advantage of HB comes as complexity increases. By relying on straightforward probability rules for obtaining the posterior, Bayes has flexibility: we can construct models from simple interactions, factor complex relationships into simple pieces, and reorganize them to facilitate computation. The term 'HB' refers to the fact that models can be constructed and solved in terms of stages.

Hierarchical Bayes models

Consider now the case of multiple sources of stochasticity, i.e. more unknowns. Suppose that our 'parameter' θ_1 is a demographic rate that might vary among individuals (Clark *et al.* 2003; Clark 2003) or locations (Wikle 2003a) due to processes that cannot be measured. We cannot ignore this variability, because the resulting confidence or confidence envelope will be inaccurate. Without changing basic model structure, we simply add a level to the foregoing model,

$$\begin{aligned} p(\theta_1,\theta_2,\theta_3|x,y) \propto\ & p(y|\theta_1,\theta_2,x) && \text{likelihood}\\ & \times p(\theta_1|\theta_3)p(\theta_2) && \text{prior}\\ & \times p(\theta_3) && \text{hyperprior} \end{aligned}$$

(Fig. 1b). Because the parameter θ_1 occupies a middle stage, it is no longer subject to the 'asymptotic collapse' with sample size. It is more like a variable, in the sense that our CI describes variability in the population. Like a parameter, it also possesses uncertainty. The parameters θ_2 and θ_3 are not conditioned on lower stages and, thus, are subject to 'asymptotic collapse'.

Alternatively, suppose that we cannot observe x (it is a latent variable) or it is sampled with error. This stochasticity in x entails an additional parameter θ_3,

$$\begin{aligned} p(\theta_1,\theta_2,\theta_3,x|y) \propto\ & p(y|\theta_1,\theta_2,x) && \text{likelihood}\\ & \times p(x|\theta_3) && \text{process}\\ & \times p(\theta_1)p(\theta_2)p(\theta_3) && \text{prior} \end{aligned}$$

(Fig. 1c). Here again, there is stochasticity at several stages. The likelihood captures the variability associated with sampling. The process that generates the latent variable x, occupies a middle stage in Fig. 1c (right-hand side). We must estimate x (it possesses uncertainty), but it is not subject to asymptotic collapse (it is a variable). Moreover 'collecting more data' means adding more things to be estimated (the xs). The parameters θ_1, θ_2, and θ_3 are subject to asymptotic collapse.

No new approaches are needed to assimilate multiple data sources for the process x, call them data sets y_1 and y_2. We could think of the same model with two likelihoods:

$$\begin{aligned} p(\theta_1,\theta_2,\theta_3,x|y_1,y_2) \propto\ & p(y_1|\theta_1,\theta_2,x)p(y_2|\theta_4,x) && \text{likelihoods}\\ & \times p(x|\theta_3) && \text{process}\\ & \times p(\theta_1)p(\theta_2)p(\theta_3)p(\theta_4) && \text{prior} \end{aligned}$$

At first glance, this might set off an alarm. We are trained to assume independent samples, an assumption that allows us to write the likelihood for a data set as the product of likelihoods for each datum. Here data sets y_1 and y_2 cannot possibly be independent – they derive from the same x (Fig. 1d). How can we simply multiply them together? Again, conditioning is the answer. The two data sets are *conditionally independent*, each with a data model and conditioned on quantities that are also modelled. In this example, the first data model involves parameters θ_1 and θ_2. The second involves parameter θ_4. Parameters might describe observation error. If one 'data' type is model output (e.g. Wikle *et al.* 2001; Fuentes & Raftery 2003), parameters might involve model error and bias. The process that generates y_1 and y_2 then is taken up at the process stage (Fig. 1d). Again, we are building a network, focusing on local connections among elements.

In summary, the hierarchical structure allows for stochasticity at multiple levels. By working with low-dimensional pieces of the problem, we are always 'conditioning'. Rather than ask 'How does the whole process work?', we ask 'How does this component work, conditioned on those elements that directly affect it?' Complex relationships in space, time, and among individuals or groups emerge when we marginalize across the components, a mindless operation

suited for computers (but requiring some algorithmic sophistication).

Analysis is accomplished by a sampling-based approach, a key innovation being application of Gibbs sampling, a Markov chain Monte Carlo (MCMC) technique. The joint posterior is proportional to the product of many distributions. We cannot integrate such distributions directly. But we can factor a high-dimensional posterior to obtain a collection of low dimensional (typically univariate) conditional distributions and sample alternately. For a deterministic analogy, suppose you receive directions to take street A, then B, then C. You need not memorize the full map, because each decision is conditionally independent of past ones. Your decision at C does not require that you remember A. You simply condition on your current location (B).

Gibbs sampling is a stochastic version of this process. Conditioned on all else, you can sample from a conditional (simple) distribution, obtained by factoring the joint posterior. Now conditioned on this value, sample the next. In this way, the algorithm marginalizes over the full model. CIs, predictive intervals, and even decision analyses (associating some utility or cost to particular outcomes) are readily assembled or derived from MCMC results. The advantage comes from the fact that complex problems are handled like simple ones. Two examples in the next section summarize applications for the most difficult (and most common) class of ecological problems, spatiotemporal ones.

Prediction follows directly. Suppose we did like to predict y' at some new location or in the future. The predictive distribution comes from integrating over the posterior distribution for things that have been estimated, indicated here by θ,

$$p(y'|y) = \int p(y'|\theta)p(\theta|y)\mathrm{d}\theta$$

The first component of the integrand will usually be available from the likelihood. The second component is the joint posterior distribution taken over all θ. This integral is typically not available, but it can be readily assembled from MCMC output. We thus have a direct link from data to inference to prediction.

Application to a 'simple' demographic process

Efforts to understand and predict forest diversity focus on early life history stages, particularly tree recruitment (Rees *et al.* 2001). The most fundamental aspect of recruitment involves seed production. Ecologists assume that fecundity follows an allometric relationship with tree diameter (e.g. Harper 1977), implemented as a regression on log variables. But fecundity schedules are more complex than this, and we cannot control variables in a way that would meet assumptions of classical models. For example, a simple allometric relationship does not describe the nonlinearities associated with maturation and senescence. We do not expect that all trees of the same size produce the same number of seeds. This variation cannot be linked to measurable variables, suggesting random effects; the process model is stochastic. To accommodate autocorrelation, we view each tree as a time series (as opposed to treating each year as an independent observation). If we simply model average seed production, we cannot infer masting. Although this is already a hard problem, we could still make some headway with classical models (e.g. Lindsey 1999). What causes us to abandon a traditional approach is the most daunting aspect of the problem: the process is hidden. We cannot count the seeds on trees in a closed stand, not even approximately.

We lack the data we want (seeds produced by a tree), but we have information we can use. There are two sources of 'data', both indirect (i.e. neither involves counting seeds on trees). First are seeds that settle in seed traps. Classical methods are used to estimate fecundity from this type of data (Ribbens *et al.* 1994; Clark *et al.* 1999). These models embed within a sampling distribution a deterministic model of seed production and dispersal to seed traps. With HB we can accommodate a second data type, observations of whether or not a tree has any seed, providing evidence of its 'maturation status'.

A fully spatiotemporal analysis of this problem requires estimates of an individual effect for each tree and fecundity for each tree in each year. If we have a decade of observations relating to thousands of trees, we might need to estimate thousands of unobservables (Clark *et al.* 2004). Why so many estimates? Because anything unobserved must be estimated. The classical options do not allow for the complexity of the problem, and they cannot exploit the multiple data types. HB allows us to model seed production as a latent process and to admit information that comes from several sources (Fig. 2). Although the full model is complex, the principal elements are tree size, reproductive status, and dispersal. Most of the structure accommodates the unknowns.

What do we gain from a full accounting of stochasticity? First, we do not require a rigid design, abstracted from the natural setting, to avoid violating model assumptions. Rather, the model is constructed to accommodate the uncertainty, and it is conditioned on what was actually observed. Second, it provides detailed insight into fecundity schedules. Previous approaches only identified a single fecundity parameter and, thus, an unrealistic fecundity schedule. Order of magnitude bias results when the dominant sources of stochasticity (interannual and individual effects) are omitted (Fig. 3) (Clark *et al.* 2004). With HB we obtain a full accounting of size effects, variability among individuals and years (variance and autocorrelation),

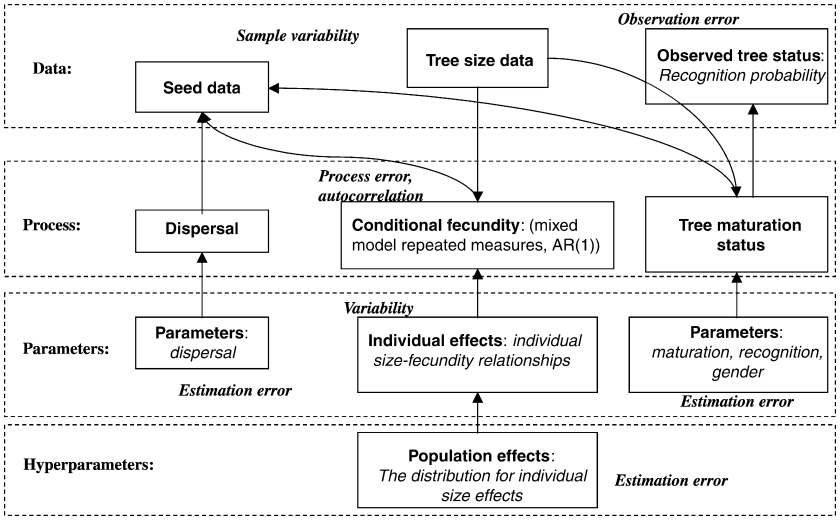

Figure 2 The hierarchical Bayes model used to estimate tree fecundity schedules. There are three data sets (upper level). The 'response variable' (conditional fecundity) is actually a latent variable and is never seen. Like all other unobservables in the model, this latent variable must be estimated – it possesses both uncertainty and variability. Inference involves a joint distribution of all parameters and latent variables, including recognition errors of tree status (lower right) and sampling variability (upper left) (Clark *et al.* 2004).

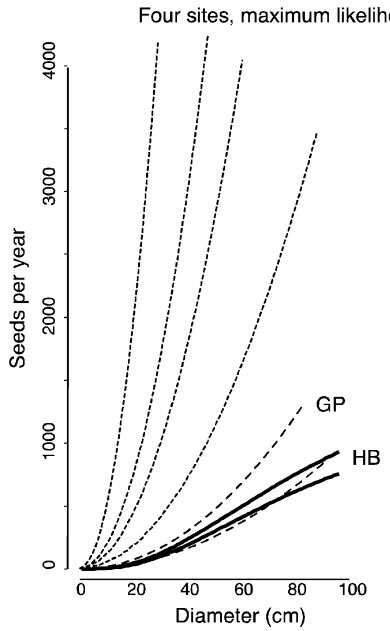

Figure 3 Estimates of oak fecundity using a classical maximum likelihood (four sites from Clark *et al.* 1999) and hierarchical Bayes (HB) compared with independent estimates from the same region, where crowns of conspecific individuals did not overlap (GP, Greenberg and Parrasol 2002). Hierarchical Bayes estimates agree with independent evidence (Clark *et al.* 2004), whereas ML estimates from the traditional model are an order of magnitude too high. (Reproduced with permission of the Ecological Society of America).

dispersal, and sex ratio, together with estimates of observation errors.

Estimates of seed production by all trees in all years showed that masting involved subsets of populations – not all individuals participate in quasiperiodic, synchronous seed production. Year-by-tree fecundity estimates are 'latent variables' and possess both variability and uncertainty (Fig. 4). For a given tree, the year-to-year variability is large, and correlations among trees indicate 'levels' of masting. Comprehensive treatment of stochasticity provides reliable estimates of uncertainty – if confidence envelopes for asymptotic parameters are unacceptably wide, additional insight requires more seed traps. However, confidence envelopes for fecundity estimates (upper left in Fig. 4) cannot be substantially reduced with more data.

The dominant stochasticity represented by interannual and individual effects influence how we view colonization capacity. Species coexistence is hypothesized to depend on tradeoffs among species in terms of their abilities to colonize vs. compete, their abilities to exploit abundant resources vs. survive when resources are scarce, or both (reviewed by Rees *et al.* 2001). When inference is based on models that treat process (interannual and individual) variability as though it was 'error' we miss the fluctuations that can affect the outcome of competition.

Application to population spread

A second example of migration exploits the same framework used for fecundity. The eastern house finch population spread west following release on Long Island, New York in 1940. The North American Breeding Bird Survey engages volunteers who attempt to identify birds by sight or sound. Sequential maps reveal westward expansion (Fig. 5). That spread is often modelled as a reaction–diffusion process, whereby individuals move some average mean squared distance D during a generation, and they reproduce at per capita rate r.

Wikle's (2003a) approach allowed for complexity and unknowns. First, the assumption that D and r are

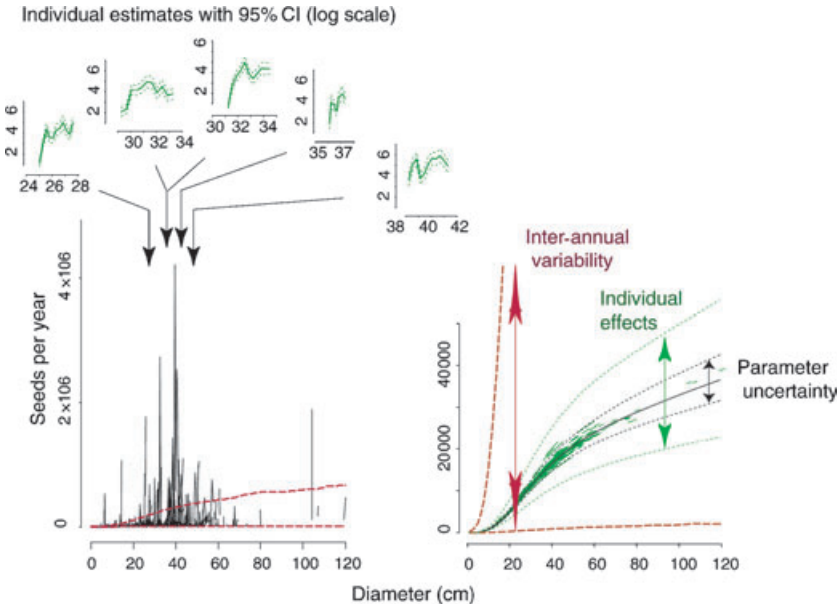

Figure 4 Sources of stochasticity in the fecundity schedule of *Liriodendron tulipifera*. At left are year-by-year fecundity estimates for all trees plotted against diameters. The low values are difficult to see on the linear scale at lower left, but are represented at right by the population mean response (black with 95% CI for parameter uncertainty), random effects of individuals (green, showing contributions of each tree and the 95% CI), and interannual variability (red 95% CI). At upper left are shown the variability (dominated by interannual variability) and uncertainty (dashed lines are 95% CI) for several individuals selected at random plotted on a log scale (Clark *et al.* 2004). (Reproduced with permission of the Ecological Society of America).

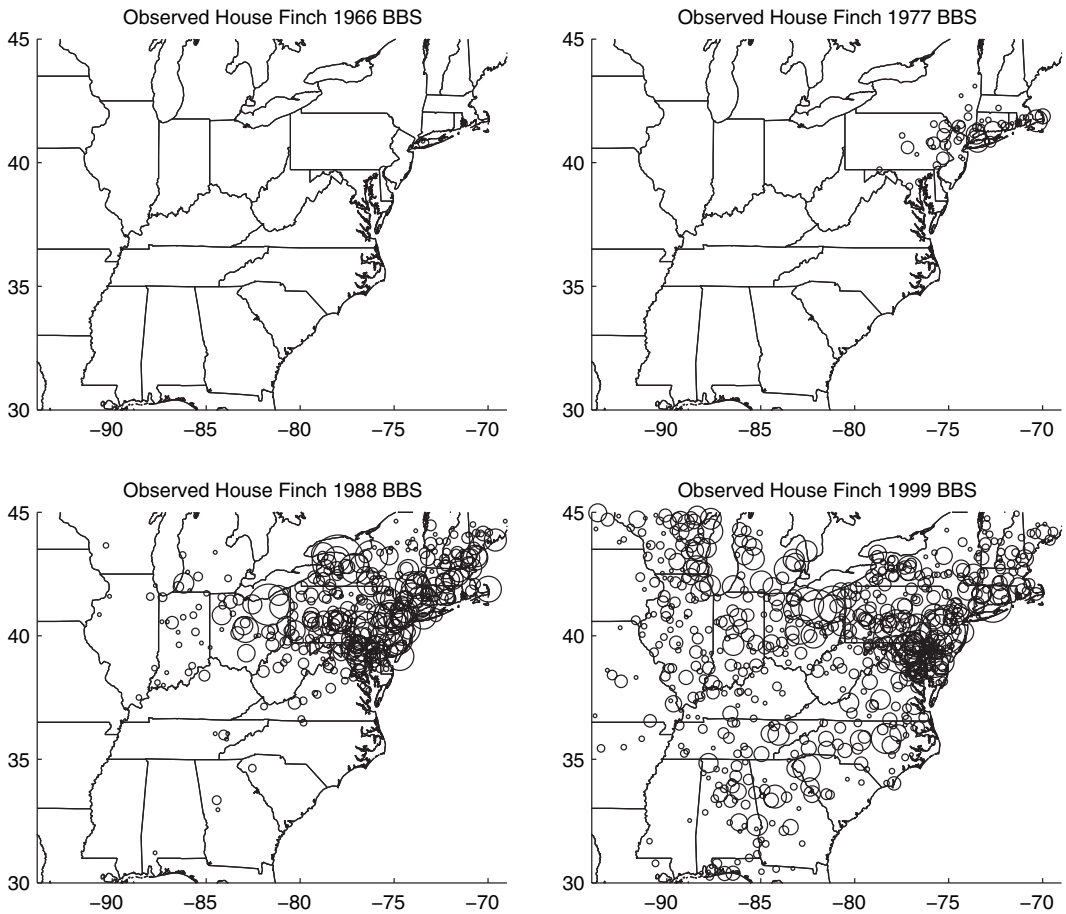

Figure 5 Breeding bird survey data for the house finch. Circle radius is proportional to BBS counts. (Reproduced with permission of the Ecological Society of America).

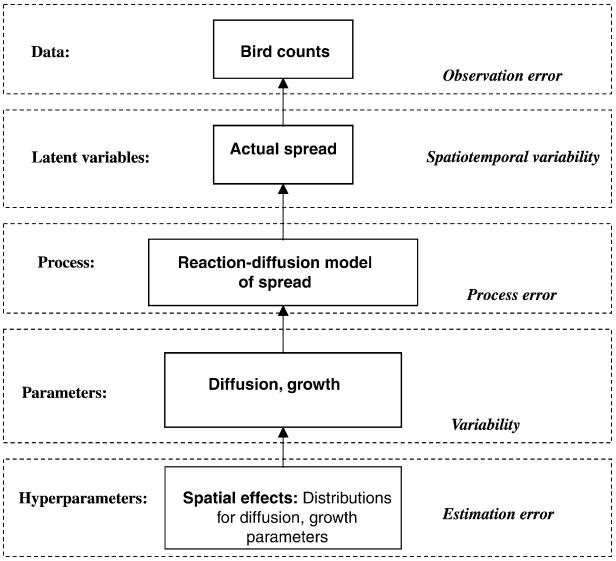

Figure 6 HB structure of the diffusion model to infer spread of the house finch.

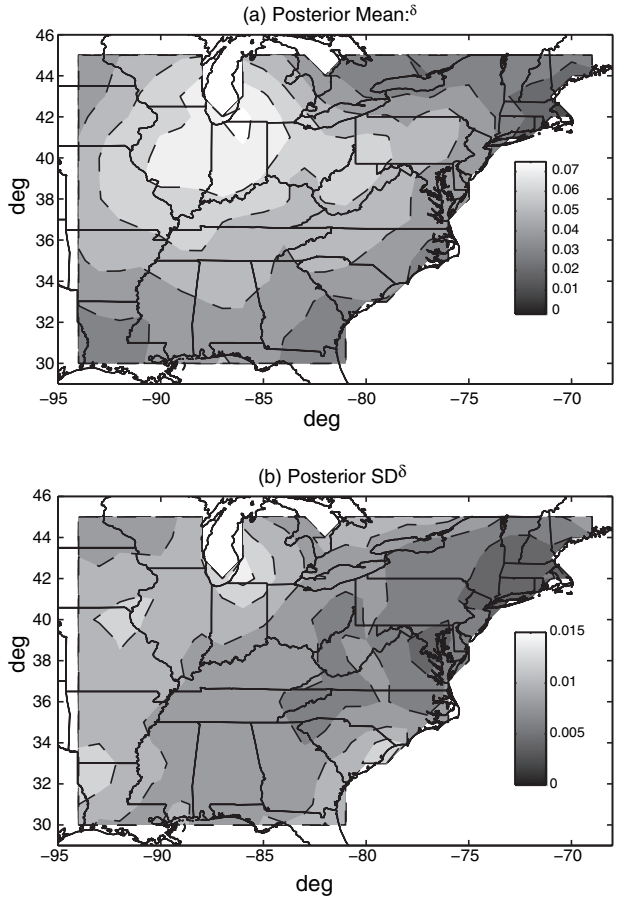

Figure 7 Posterior mean and standard deviation for westward 'diffusion' of the house finch (Wikle 2003a). Diffusion increases as populations expand into the Midwest, and variability increases. (Reproduced with permission of the Ecological Society of America).

everywhere the same is unrealistic. We do not expect birds to survive, reproduce, and disperse in New Jersey as they might in, say, Illinois. A parameter model added at a lower stage (Fig. 6) accounts for this variability. Asymptotics apply to the 'hyperparameters', in the sense that CIs will decline with sample size, while the local parameter estimates can vary.

Second, the process itself is not exactly reaction–diffusion. Diffusion may provide a rough caricature of population spread, but birds are not gas molecules. Population movement involves many factors that are not included in the diffusion model and could not be identified from BBS data. This 'process error' (model misspecification) is an additional stage (Fig. 6). Stochasticity steps in to accommodate the unknowable factors that distinguish actual population spread from strict reaction diffusion.

Finally, we do not observe population spread, but rather a crude approximation in the form of BBS data. If we ignore this 'observation error', then we are, in effect, modelling a diffusion process on observations (e.g. Fig. 5), rather than on the bird populations. The distinction is critical, because the future population should not depend on whether or not there is an observer in this county, an observer can distinguish one species from another, and so forth.

What do we gain from the broad treatment of complexity? First, it looks hard, but its easier than would be any traditional attempt to accomplish the same thing. The estimate of density at one location depends on observations everywhere else at all other times (forward and backward). But with HB, we need only specify the model for a location and its neighbours; construct the model in parts and let the computer marginalize.

Second, the parameter estimates and uncertainties are more realistic, because the model represents how most ecologists believe that the process operates. The population 'parameters' we wish to estimate are actually latent variables – r and D vary from place to place, just as we expect for a population. Unlike standard errors for parameters subject to asymptotics, these standard deviations describe variability in population growth and diffusion. For example, Wikle found that diffusion not only increases as the population spreads west (Fig. 7a), it also becomes more variable (Fig. 7b). Estimates of mean and standard deviation are relatively unbiased by the large stochasticity associated with observations and model misspecification. By contrast, a classical approach would provide point estimates for D and r and standard errors that provide no insight on variability.

Model size and information in data

In the early 1990s the number of unknowns estimated with HB gave pause to classically trained statisticians. Classical inference is limited by a relationship between model dimension and information in data. In the classical framework, model dimension and information bear simple relationships to number of parameters and sample size, respectively. The notion of estimating more parameters than there are data points violates ingrained conventions that are indeed sacred in the traditional setting.

With HB, simple conventions do not hold. We can place bounds on model dimension, but a precise definition of model size is only possible in the context of specific assumptions that can be debated (e.g. Spiegelhalter *et al.* 2002). Likewise, the information in data is only loosely connected to sample size, depending on where data types enter the model network, the nature of conditional relationships among variables (arrows in Figs 1 and 2), and so forth. For a complex model, it should be apparent that a small set of strategic observations can have impact that overwhelms a large set of relatively non-informative ones. For example, consider a simple model of exponential population growth with observations taken every 10 years. Our model of exponential growth is one of the assumptions that can be evaluated (e.g. through goodness of fit, model selection, and so on), along with the estimates of parameters and of population densities, which are typically sampled with error. If the model fits well, we have information on population size not only for sample years, but also for years in which we did not sample. Otherwise, how could we claim any confidence in the model? The uncertainty in those estimates will depend on how well the model 'fits' the data and on the time since or until the next sample. We might use the model to draw inference on population sizes at many more dates than those for which we have observations (Clark & Bjornstad 2004).

With complex networks, it becomes impossible to precisely define model dimension or to anticipate how large n should be to achieve a confidence envelope of a particular width. In the fecundity example, we had two different data types with different sample sizes from which we estimated seed production from far more tree-years than there were observations for either data set. In the house finch example, the amount of information in data depends on, among other things, spatiotemporal correlations. Samples obtained within the correlation length scales contain substantial redundancy. 'Effective n' is much lower than the number of observations.

Although model size and information cannot be simply defined, we can say how many unknowns we are estimating – each is assigned an explicit place in the network and has a prior and a posterior distribution.

THE CHALLENGES IN COMPLEX MODELS AND DATA

For simple problems, Bayes is more difficult than traditional methods, because it requires integration, and there are fewer software options. As problems become more complex, the situation is reversed. The prior-likelihood structure can be broadly extended, without changing the basic approach. MCMC methods for computation allow us to 'model locally', leaving integration to the computer. Conceptually, models for 'difficult' space–time problems are not too different from simple ones. As complexity increases the challenges become computational. MCMC algorithms must 'converge' to a proper posterior density. In principal we might construct models with a large number of levels, hoping for some degree of learning that comes from a 'borrowing of strength' across groups. In practice, it may be difficult to specify 'non-informative' priors for models with deep hierarchies and even to identify when there are problems (e.g. Berger *et al.* 2001). Although there are number of indices to help gauge convergence (e.g. Gelman & Rubin 1992), for complex models one may never be completely sure.

The fact that practitioners of HB devote substantial effort to computational challenges has fostered the impression among some ecologists that it must represent one of the 'cons' on the Bayes side of the ledger to be weighed when choosing between Bayes vs. a traditional approach. The complexity of problems addressed by HB requires sophistication. For sure, we bump up against computational issues with HB, but its because we are dealing with far more complex problems than we would attempt with a classical approach.

THE BROADER ROLE OF HB IN THE PREDICTION PROCESS

Development of new tools makes timely a reappraisal of the promise and limitations of ecological forecasting. Arguably both are underestimated. Ecological problems are high-dimensional. With the advent of computers, ecological models expanded to embrace many deterministic relationships. Some disillusion came with growing appreciation of the overfitting problem and inevitable ad hockery needed to 'parameterize' the many unmeasurables. Overfitted models have enough dimensions to fit scatter, but lack predictive capacity.

Simple models were a logical reaction to frustration with intractable, poorly parameterized simulation models. If complex models fail, measure what we can hope that the unmeasurables will have limited impact. The numbers of birds recorded by a volunteer one day a year is not the actual number of birds, but perhaps its close enough. The number

of seeds on the ground bears a complex relationship to fecundity, but we might overlook that. Although simple models can be more 'predictive' than complex ones (because they are not overfitted), several decades of focus on such models have not fostered confidence in a predictive science. When pressed for guidance, ecologists and managers tend to bypass models fitted to data and move directly to qualitative approaches. The recent emergence of scenarios is a healthy reaction to the need for thoughtful treatment of uncertainty and limited information (Clark *et al.* 2001; Scheffer *et al.* 2001). It complements, but does not substitute for, quantitative assessments that could be informed by rapidly expanding data sets.

The promise of HB includes the potential to treat high dimensional problems with full exploitation of information and accommodation of the unknowns. By contrast with early ecological models, much of the dimensionality comes from stochastic components. The distinction is important. Inability to measure an effect need not be an excuse for ignoring it. The fecundity example exploited two different data sets, linked them together with three processes, and fully exploited the spatial and temporal information they contained. As with the diffusion example, most ecologists could agree on the structure of the unknowns, and that can be enough. Unlike a complex deterministic model, we can move ahead despite being unable to measure all influences. Unlike a simple deterministic model that ignores these contributions, an appropriate structure allows inference for the simple process. Stochasticity stands in for the unknowns and unmeasurables.

Although HB seems 'hard', apparent difficulty comes from the complex relationships it can address. Ecologists have long been consumers of a few basic alternatives broadly available in software packages. Limited options in standardized software insulate the user from technical detail. With the flexibility of HB comes the need for environmental scientists to be more actively engaged. Ecologists may gain facility with distributional theory and computation or collaborate more closely with statisticians (Wikle 2003b). There are some software tools (e.g. Winbugs, http://www.mrc-bsu.cam.ac.uk/bugs/winbugs), and these will continue to improve. The challenges of learning basics of HB are weighed against benefits of its generality. In contrast to separate texts and courses for each model class, recent references on HB (Gelman *et al.* 1995; Carlin & Louis 2000; Congdon 2001) find basic models, mixed models, and spatiotemporal models *in many of the same chapters*. The mind-boggling proliferation of test-statistics we have come to accept with traditional approaches is avoided.

Hierarchical Bayes is gaining wide acceptance for its potential to accommodate high-dimensional problems and obscure evidence. The growing examples include genomics (Hartemink *et al.* 2002), clinical trials (Spiegelhalter *et al.* 2003), atmospheric sciences (Berliner *et al.* 1999), fisheries (Meyer & Millar 1999), population dynamics (Ver Hoef 1996; Calder *et al.* 2003; Clark 2003; Clark *et al.* 2003, 2004; Clark & Bjornstad 2004), biodiversity (Gelfand *et al.* 2004), the internet (Mitchell 1997), and finance (Jacquier *et al.* 1994). Structural equation models, of recent interest to ecologists (e.g. Shipley 2000), can be treated more flexibly within a Bayesian framework (distributional assumptions can be relaxed) than is possibly with classical approaches. This expansion of Bayes is moving ahead without convergence of views on philosophy (e.g. frequency vs. subjective notions of probability). The prior-likelihood structure opens possibilities for extension to a broad range of challenges that face many disciplines.

Finally, expectations should be realistic. Informative predictions are not to be expected for all applications. For example, expanding application of Bayesian models in finance does not mean that investors successfully anticipate stock market fluctuations. Realistic goals will emphasize what is predictable, and they will rely on models that link predictions to data in appropriate ways. The comprehensive accounting of variability and uncertainty is central. The distinction will provide guidance as to what is 'predictable', what is inherently unpredictable, and where additional data can provide the most benefit (Sarewitz & Pielke 2000; Clark *et al.* 2001).

ACKNOWLEDGEMENTS

For comments, I thank M. Dietze, A. Gelfand, F. He, I. Ibanes, S. LaDeau, A. McBride, J. McLachlan, Nathan Welch, Chris Wikle, and three anonymous referees. Research support was provided by NSF grants DEB-9632854 and DEB-9981392 and DOE grant 98999.

REFERENCES

Berger, J. & Berry, D. (1988). Analyzing data: is objectivity possible? *Am. Sci.*, 76, 159–165.

Berger, J., DeOliveira, V. & Sanso, B. (2001). Objective Bayesian analysis of spatially correlated data. *J. Am. Statist. Assoc.*, 96, 1361–1374.

Berliner, M. (1996). Hierarchical Bayesian time series models. In: *Maximum Entropy and Bayesian Methods* (eds Hanson, K. & Silver, R.). Kluwer, Norwell, MA, pp. 15–22.

Berliner, M., Royle, A.J. Wikle, C.K. & Milliff, R.F. (1999). Bayesian methods in the atmospheric sciences. In: *Bayesian Statistics*, Vol. 6 (Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F.M.). Oxford University Press, Oxford, UK, pp. 83–100.

Calder, K., Lavine, M., Mueller, P. & Clark, J.S. (2003). Incorporating multiple sources of stochasticity in population dynamic models. *Ecology*, 84, 1395–1402.

Carlin, B.P. & Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, Boca Raton, FL, USA.

Carpenter, S.R. (1996). Microcosm experiments have limited relevance for community and ecosystem ecology. *Ecology*, 77, 677–680.

Carpenter, S.R. (2002). Ecological futures: building an ecology of the long now. *Ecology*, 83, 2069–2083.

Caswell, H. (1988). Theory and models in ecology: a different perspective. *Ecol. Model.*, 43, 33–44.

Clark, J.S. (2003). Uncertainty in population growth rates calculated from demography: the hierarchical approach. *Ecology*, 84, 1370–1381.

Clark, J.S. (2004). *Models for Ecological Data*. Princeton University Press, Princeton, NJ, under contract.

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

Clark, J.S. & LaDeau, S. (2004). Synthesizing ecological experiments and observational data with hierarchical Bayes. In: *Applications of Hierarchical Bayes in the Environmental Sciences* (eds Clark, J.S. & Gelfand, A.E.), Oxford University Press, in press.

Clark, J.S., Silman, M., Kern, R., Macklin, E. & Hille Ris Lambers, J. (1999). Seed dispersal near and far: generalized patterns across temperate and tropical forests. *Ecology*, 80, 1475–1494.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J., *et al.* (2001). Ecological forecasts: an emerging imperative. *Science*, 293, 657–660.

Clark, J.S., Mohan, J., Dietze, M. & Ibanez, I. (2003). Coexistence: how to identify trophic tradeoffs. *Ecology*, 84, 17–31.

Clark, J.S., LaDeau, S. & Ibanez, I. (2004). Fecundity of trees and the colonization-competition hypothesis. *Ecol. Monogr.*, 74, 415–442.

Congdon, P. (2001). *Bayesian Statistical Modelling*. Wiley, New York.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., *et al.* (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387, 253–260.

Daily, G.C. (1997). *Nature's Services: Societal Dependence on Natural Ecosystems*. Island Press, Washington, DC.

Dawid, A.P. (2004). Probability, causality and the empirical world: a Bayes–de Finetti–Popper–Borel synthesis. *Statist. Sci.*, 19, 44–57.

Dixon, P. & Ellison, A.M. (1996). Introduction: ecological applications of Bayesian inference. *Ecol. Appl.*, 6, 1034–1035.

Ellison, A.M. (2004). Bayesian inference in ecology. *Ecol. Lett.*, 7, 509–520.

Ellner, S.P. & Fieberg, J. (2003). Using PVA for management in light of uncertainty: effects of habitat, hatcheries, and harvest on salmon viability. *Ecology*, 84, 1359–1369.

Emlen, J., Kapustka, L., Barnthouse, L., Beyer, N., Biddinger, G., Kedwards, T. *et al.* (2002). Ecological resource management: a call to action. *Bull. Ecol. Soc. Am.*, 83, 269–271.

Fuentes, M. & Raftery, A.E. (2004). Model validation and spatial interpolation by combining observations with outputs from numerical models via Bayesian melding. *J. Am. Stat. Assoc. Biometrics*, in press.

Gelfand, A.E. & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, 85, 398–409.

Gelfand, A.E., Schmidt, A.M., Wu, S., Silander, J.A.J., Latimer, A. & Rebelo, A.G. (2004). Explaining species diversity through species level hierarchical modeling. *J. R. Statist. Soc. B, Applied statistics*, in press.

Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.*, 7, 457–511.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London, UK.

Greenberg, C.H. & Parresol, B.R. (2002). Dynamics of acorn production by five species of oaks. In: *Oak Forest Ecosystems* (eds McShea, W.J. & Healy, W.M.). John Hopkins Univ. Press, Balmitore, pp. 149–172.

Halpern, J.Y. (2003). *Reasoning about Uncertainty*. MIT Press, Cambridge, MA.

Harper, J.L. (1977). *Population Biology of Plants*. Academic Press, New York.

Hartemink, A., Gifford, D., Jaakkola, T. & Young, R. (2002). Bayesian methods for elucidating genetic regulatory networks. *IEEE Intell. Syst. Biol.*, 17, 37–43.

Hilborn, R. & Mangel, M. (1997). *The Ecological Detective*. Princeton University Press, Princeton, NJ.

Jacquier, E., Polson, N. & Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *J. Business Econ. Statist.*, 12.

Levin, S.A. (1992). The problem of pattern and scale in ecology. *Ecology*, 73, 1943–1967.

Lindsey, J.K. (1999). *Models for Repeated Measurements*. Oxford University Press, Oxford, UK.

Meyer, R. & Millar, R.B. (1999). Bugs in Bayesian stock assessments. *Can. J. Fish. Aquat. Sci.*, 56, 1078–1086.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill, New York.

Pearl, J. (2002). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Peterson, G.D., Carpenter, S.R. & Brock, W.A. (2003). Uncertainty and the management of multistate ecosystems: an apparently rational route to collapse. *Ecology*, 84, 1403–1411.

Pielke, R.A. & Conant, R.T. (2003). Best practices in prediction for decision-making: lessons from the atmospheric and earth sciences. *Ecology*, 84, 1351–1358.

Rees, M., Condit, R., Crawley, M., Pacala, S. & Tilman, D. (2001). Long-term studies of vegetation dynamics. *Science*, 293, 650–655.

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

Sarewitz, D. & Pielke, R.A., Jr (2000). *Prediction: Science, Decision Making and the Future of Nature*. Island Press, Washington, DC.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C. & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, 413, 591–596.

Shipley, B. (2000). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations, and Causal Inference*. Cambridge University Press, Cambridge, UK.

Skelly, D. (2002). Experimental venue and estimation of interaction strength. *Ecology*, 83, 2097–2101.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with Discussion). *J. R. Statist. Soc. B*, 64, 583–639.

Spiegelhalter, D.J., Abrams, K.R. & Myles, J.P. (2003). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley Press, Chichester, England.

Ver Hoef, J.M. (1996). Parametric empirical Bayes methods for ecological applications. *Ecol. Appl.*, 6, 1047–1055.
Wikle, C.K., (2003a). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84, 1382–1394.
Wikle, C.K. (2003b). Hierarchical models in environmental science. *Int. Statist. Rev.*, 71, 181–199.
Wikle, C.K., Milliff, R.F., Nychka, D. & Berliner, L.M. (2001). Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds. *J. Am. Statist. Assoc.*, 96, 382–397.

Editor, Fangliang He
Manuscript received 21 June 2004
First decision made 25 July 2004
Second decision made 30 September 2004
Manuscript accepted 27 October 2004
Exceeded normal length