Bayesian Data–Model Integration in Plant Physiological and Ecosystem Ecology

Kiona Ogle(✉) and Jarrett J. Barber

Contents

1 Introduction: Data–Model Synthesis 282
1.1 The Data 282
1.2 The Process Model 283
1.3 Data–Model Integration 283
2 Bayesian Applications in Plant and Ecosystem Ecology 284
3 Overview of the Bayesian Approach 286
3.1 Hierarchical Probability Models 286
3.2 Bayes Theorem and Bayesian Inference 289
3.3 Hierarchical Bayesian Modeling 291
4 Example Applications 293
4.1 Example 1: Leaf-Level Gas Exchange 293
4.2 Example 2: Use of Prior Information with Process Models 304
5 Final Remarks 304
5.1 Scaling Plant Physiological and Ecosystem Ecology 304
5.2 Implementation Issues 305
5.3 Thoughts on Statistical Education and Practice 306
References 307

Abstract This paper reviews and illustrates the use of modern methods for integrating diverse data sources with process-based models for learning about plant physiological and ecosystem processes. The particular focus is on how such data sources and models can be coupled within a hierarchical Bayesian modeling framework. This framework, however, has been underutilized in physiological and ecosystem ecology, despite its great potential for data–model integration in these areas. This paper provides a summary of the use of Bayesian methods in ecological research and gives detailed examples highlighting existing and potential uses of Bayesian and hierarchical Bayesian methods in plant physiological and ecosystem ecology. This paper also provides an overview of the statistical theory underlying the development of hierarchical Bayesian methods for analyzing complex ecological problems.

Kiona Ogle
Department of Botany and Department of Statistics, Department of Botany, Box 3165, University of Wyoming 1000 E. University Ave. Laramie, WY 82071, USA
e-mail: kogle@uwyo.edu

U. Lüttge et al. (eds.), *Progress in Botany 69*.

The methods are applied to specific examples that include a detailed illustration of a hierarchical Bayesian analysis of leaf-level gas exchange data that are integrated with models of photosynthesis and stomatal conductance, and Bayesian approaches to estimating parameters in complex ecosystem simulation models. The paper concludes with some practical issues and thoughts on the direction of hierarchical Bayesian modeling in plant physiological and ecosystem ecology.

1 Introduction: Data–Model Synthesis

This paper is a review and prospectus of modern methods for integrating diverse data sources with process-based models for learning about plant physiological and ecosystem processes. In particular, we discuss how such data sources and models can be coupled within a hierarchical Bayesian modeling framework (Wikle 2003b; Clark 2005; Clark et al. 2005). This framework, however, has been underutilized in physiological and ecosystem ecology (Ellison 2004), despite its great potential for data–model integration in these areas. We summarize the use of Bayesian methods in ecological research and provide detailed examples highlighting uses of Bayesian and hierarchical Bayesian methods in plant physiological and ecosystem ecology. In the process, we provide a sort of tutorial on the use of hierarchical Bayesian methods for analyzing complex ecological problems. We conclude with some practical issues and thoughts on the direction of hierarchical Bayesian modeling in plant physiological and ecosystem ecology.

1.1 The Data

Plant physiological and ecosystem ecologists are proficient at collecting enormous amounts of data. This proficiency is partly due to advances in automated data collection devices (e.g. weather stations, soil moisture probes, eddy flux towers, continuous soil flux chambers, tunable diode lasers, sap flow sensors) that record at frequent intervals (e.g. every 30 min for multiple days, months, or years), and to sophisticated and expensive tools allowing hands-on field measurements (e.g. portable leaf-level gas exchange analyzers, stable isotopes, minirhizotrons). These diverse data represent ecological or physical processes that span different spatial scales (e.g. eddy flux and whole-ecosystem gas exchange vs soil chambers for measuring soil respiration), different temporal scales (e.g. carbon isotopes in leaf tissue and integrated water use efficiency vs leaf-level gas exchange and instantaneous water use efficiency), and different levels of biological organization (e.g. leaf temperature images from infrared cameras to infer patchy stomatal conductance vs sapflux and plant- or canopy-level transpiration). In addition, there is a plethora of information that already exists in the literature. How does one deal with such data? How can we extract information from different types of data to make inferences about complex, interrelated mechanisms governing plant and ecosystem processes?

1.2 The Process Model

Process models traditionally provide the means by which mechanisms operating at different temporal, spatial, and biological scales are formalized and explicitly linked. Development of a process model (or models) forces us to quantify our understanding of ecological systems, thus improving our heuristic and predictive understanding of plant physiological and ecosystem responses (e.g. Levin 1992; Canadell et al. 2000; Rastetter et al. 2003). Process models have been developed to describe and predict an array of physiological and ecosystem behaviors. Such models include, for example, the "Farquhar" biochemical-based model of leaf-level photosynthesis (Farquhar et al. 1980; Farquhar and Sharkey 1982; Farquhar and von Caemmerer 1982), the canopy-level photosynthesis and evaptranspiration model PnET (Aber and Federer 1992; Aber et al. 1996), and the plant–soil nutrient cycling model CENTURY (Parton et al. 1987). Components of such process models, including functional forms and parameter values, are often derived from empirical information, and these models are frequently used to learn about the real system through simulation experiments. However, a model of even moderate complexity is rarely fit to observed field or experimental data in a rigorous manner. Instead, data summaries (typically means) may be used for parameter values in a model. Or, a model may be "tuned" – parameters adjusted in a relatively ad-hoc manner – so that it fits a particular dataset or, more often, a summary thereof. Such approaches are often referred to as "forward modeling" where the primary focus is on understanding and predicting model and system behavior via simulation – running the model "forward" given a set of parameter values – with comparatively little emphasis on quantifying parameter uncertainty, or more generally, on rigorous data–model integration.

1.3 Data–Model Integration

So, we have varied and large datasets that are generated by complex, interrelated ecological processes, and we have process models designed specifically to learn about these complexities. When faced with several types of data that may inform the same process (or a set of processes), the standard approach is to analyze data in a piece-wise fashion – each dataset analyzed independently of others, largely ignoring the interconnectedness of the data. Instead of a separate analysis approach or a relatively ad-hoc tuning approach to process simulation, we argue a need for a more rigorous method that allows all data and prior knowledge to inform process parameters without requiring undue process model simplifications that could compromise the integrity of the inherent complexity of the ecological system embodied by the model. Ultimately, this data–model integration should allow exploration of ecological questions that are otherwise difficult to address via traditional approaches.

Recent developments in computing technology and Bayesian statistical methodology now allow for such integration of complex process models with large and diverse

datasets within a unified statistical inferential framework. The Bayesian approach is emerging as a useful tool in population and community ecology (Clark 2003; Ellison 2004; Clark 2005; Clark and Gelfand 2006), but it has received comparatively little attention in plant physiological and ecosystem ecology (e.g. Ellison 2004; Van Oijen et al. 2005). The perspective of this paper is similar to that of Clark and Gelfand (2006), our intention here being to raise awareness of hierarchical Bayesian methods for coupling process models and data in plant physiological and ecosystem ecology.

In Section 2, we give a brief review of existing applications of Bayesian and hierarchical Bayesian methodologies in plant physiological and ecosystem ecology and closely related fields. Section 3 presents basic probability results, including hierarchical probability model specification and a statement of Bayes theorem, that are fundamental to hierarchical Bayesian modeling of complex processes. Section 4 presents detailed examples that illustrate how these results may be applied to relatively complex data–model systems in plant physiological and ecosystem ecology. In Section 5, we conclude with some practical suggestions and offer a few remarks on the future of hierarchical Bayesian modeling in these fields.

2 Bayesian Applications in Plant and Ecosystem Ecology

Bayesian statistical methods are relatively new to ecologists (Ellison 2004; Clark 2005). Ellison (2004) surveyed articles published in major ecological journals between 1996 and 2003 and found only 62 articles that used Bayesian data analysis tools. The ecological topics of these papers centered on population and community ecology, and plant physiological and ecosystem studies lagged behind in their use of Bayesian methods. Given the recent rapid increase in Bayesian applications in ecology, we expanded the scope of Ellison (2004) to include top plant physiological and ecosystem journals for the period 1996 to 2006. The results from our survey are summarized in Table 1 and Fig. 1.

Bayesian applications in population and community ecology continue to grow (Fig. 1), with only 20–25% of these studies addressing problems of a botanical nature. It should be noted that Bayesian inference in phylogenetics (Yang and Rannala 1997; Larget and Simon 1999; Mau et al. 1999; Huelsenbeck and Ronquist 2001) has become popular due to statistical and computational demands associated with analyzing enormous molecular datasets within the context of complex evolutionary models (e.g. Huelsenbeck et al. 2001). A large number of papers – 58 for the search criteria in Table 1 and 1411 for unrestricted dates and journals – employed Bayesian methods to construct phylogenies to infer, for example, evolutionary relationships among species (e.g. Miller et al. 2004; McKown et al. 2005). Despite a comparable level of complexity in plant physiological and ecosystem ecology, Bayesian methods remain underutilized in these areas (Fig. 1, Table 1). Once the plant physiological and ecosystem community becomes more aware of these methods, we anticipate a rapid increase in the application of Bayesian methodologies.

There are, however, a few notable examples of Bayesian applications in plant physiological and ecosystem ecology and closely related fields. For example, global

Table 1 Survey of ecological studies using Bayesian methods published between 1996 and 2006 in the following major ecological journals: *American Journal of Botany*, *American Naturalist*, *Conservation Biology*, *Ecological Applications*, *Ecological Monographs*, *Ecology*, *Ecology Letters*, *Ecosystems*, *Functional Ecology*, *Global Change Biology*, *Journal of Applied Ecology*, *Journal of Animal Ecology*, *Journal of Ecology*, *New Phytologist*, *Oecologia*, *Oikos*, *Plant Cell and Environment*, and *Tree Physiology*. See Ellison (2004) for a similar survey, based on a subset of these journals, for papers published between 1996 and 2003. *Note*: we used the ISI Web of Science search engine to find papers in the focal journals with "Bayes" or "Bayesian" in the title, abstract, or subject. Thus, this is not a comprehensive survey of all ecology-related papers using Bayesian methods, as there are surely other papers in journals that we did not consider and there are likely papers using Bayesian methods that do not have the term in the title, abstract, or subject. Papers related to population ecology were split into two categories, those focusing on population-related and life-history aspects of single species and those focusing on interacting species (e.g. predator–prey dynamics). Our rules for classifying articles may have differed from Ellison (2004), and thus our numbers are not in perfect agreement with Ellison's. We also tallied papers that presented a problem with a clear animal or plant focus. The numbers of plant and animal papers do not always add up to the total number of papers because some papers were classified as both (e.g. plant–animal interactions), and some were not classified as either (e.g. they may have focused on, for example, soil biogeochemistry). A full bibliography of these studies can be obtained from the authors

Topic	Number of papers (1996–2006)	Primary focus		Selected examples
		Plant	Animal	
Population ecology (single species)	73	10	63	O'Hara et al. (2002), Clark et al. (2003a, 2005), Wikle (2003a)
Population ecology (interacting species)	12	3	11	Dixon et al. (2005), Ovaskainen and Laine (2006)
Community ecology	24	12	14	Crome et al. (1996), Cottingham and Schindler (2000), Etienne and Olff (2005)
Physiological ecology	1	1	0	Ogle et al. (2004)
Ecosystem ecology	9	6	1	Carpenter et al. (1996), Rains et al. (2004), Braswell et al. (2005)
Review/comment	9			Ellison (1996, 2004), Clark (2005)
Total	128	32	89	

change scientists have embraced Bayesian inversion methods that couple various CO_2 flux and concentration data with atmospheric transport and/or terrestrial biosphere models to predict regional, continental, and global sources and sinks of atmospheric CO_2 (Enting et al. 1995; Dargaville et al. 2002; Law et al. 2004). Bayesian applications are surfacing in smaller-scale studies related to plant physiological ecology (e.g. Ogle et al. 2004) and ecosystem carbon, water, and nutrient dynamics (e.g. Braswell et al. 2005; Hong et al. 2005), and the ecosystem-related studies in Section 4.2 serve as good examples of how to incorporate informative prior information. Examples of Bayesian applications in ecosystem-level studies include: partitioning of ecosystem carbon fluxes into photosynthesis and respiration (Ogee et al. 2004), inverting a carbon cycle model to infer soil carbon loss and transfer between different litter, microbial, and soil organic matter pools (Xu et al. 2006), and parameterizing process-based models of tree growth and forest productivity

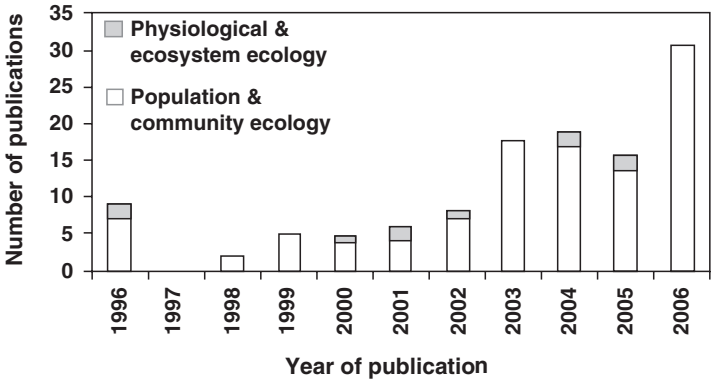


Fig. 1 Number of publications related to physiological and ecosystem ecology (*gray bars*) vs population and community ecology (*open bars*) that used Bayesian methods during the period 1996–2006. Details of the literature survey are given in Table 1 and discussed in Section 2.

(Gertner et al. 1999; Radtke et al. 2002; Van Oijen et al. 2005). These studies and those listed in Table 1 are prime examples of how Bayesian statistical methods can be used to integrate complex ecological models and data related to plant physiological and ecosystem processes.

3 Overview of the Bayesian Approach

In this section, we give a necessarily broad overview of hierarchical Bayesian modeling. Our intent is not to extol the virtues or to criticize the weaknesses of statistical paradigms, Bayesian or otherwise – we would have little more to offer to the philosophical debates among statistical schools of thought. Instead, we attempt to focus the discussion toward current developments of hierarchical statistical modeling of complex ecological systems. We start by presenting basic probability rules that are useful for a subsequent discussion of hierarchical probability modeling and for a statement of Bayes theorem. We conclude this section with what is loosely becoming known as the "process sandwich", a largely conceptual way to view the incorporation of process models into a hierarchical probability framework.

3.1 Hierarchical Probability Models

We begin with some basic probability rules. Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ denote a vector of n random variables, with $p(\mathbf{y})$ denoting a mathematical formula describing their probabilistic behavior. That is, $\mathbf{y}$ is a random (vector) variable with (joint probability) distribution $p(\mathbf{y})$. For example, $\mathbf{y}$ may be a collection of n observed leaf transpiration

rate values, and $p(\mathbf{y})$ may be a multivariate Gaussian (or "normal") probability density function. We use $\mathbf{y} \sim p(\mathbf{y})$ to denote the distribution of a random variable – i.e., $\mathbf{y}$ is "distributed as" $p(\mathbf{y})$.

Often, we wish to learn about the joint behavior of two random variables, for example, $\mathbf{x}$ and $\mathbf{y}$, together, and we write $(\mathbf{x}, \mathbf{y})' \sim p(\mathbf{x},\mathbf{y})$ to denote the joint distribution of $\mathbf{x}$ and $\mathbf{y}$. In this context, we now refer to $p(\mathbf{x})$ and $p(\mathbf{y})$ as marginal distributions where $p(\mathbf{y}) = \int p(\boldsymbol{x}, \boldsymbol{y}) d\mathbf{x}$ or $p(\mathbf{y}) = \sum_{\mathbf{x}} p(\mathbf{x}, \mathbf{y})$ if $\mathbf{x}$ is discrete-valued.

In a modeling situation, the marginal distributions $p(\mathbf{x})$ or $p(\mathbf{y})$ may be relatively easy to specify, but the joint nature of the relationship between $\mathbf{x}$ and $\mathbf{y}$ may make direct specification of $p(\mathbf{x}, \mathbf{y})$ more difficult. Also, if given a value of, say, $\mathbf{x}$, then the problem of specifying a distribution for $\mathbf{y}$ may become easier. That is, the data analyst or modeler often finds the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ easier to specify than $p(\mathbf{y})$. We use $p(\mathbf{y}|\mathbf{x})$ to denote the dependence of the distribution of $\mathbf{y}$ on the value of $\mathbf{x}$, and we read "$\mathbf{y}|\mathbf{x}$" as $\mathbf{y}$ "given" or "conditional upon" $\mathbf{x}$. As we discuss below, the idea of conditioning is fundamental to building complex joint probability models from relatively simple "pieces."

Perhaps the most familiar example of conditional model specification occurs with regression analysis or with analysis of variance (ANOVA) whereby the distribution – typically Gaussian – of the response $\mathbf{y}$ is specified conditionally on the value of the covariate $\mathbf{x}$ usually through the mean of $p(\mathbf{y}|\mathbf{x})$, i.e., through the "regression function." In many circumstances, depending on the assumptions for $\mathbf{x}$ (e.g. $\mathbf{x}$ may be assumed "fixed" or measured without error), it suffices to specify only $p(\mathbf{y}|\mathbf{x})$ in order to "best" learn about the parameters (e.g. the intercept and slope coefficient) in the regression function. Other cases lead to specification of a distribution for $\mathbf{x}$, $p(\mathbf{x})$, which, incidentally, is done when modeling "measurement error" or "error in variables" (e.g. Stefanski 2000). If we consider both $\mathbf{y}$ and $\mathbf{x}$ as random quantities, then, by basic probability rules, $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$ specify a joint probability distribution for $\mathbf{x}$ and $\mathbf{y}$ together,

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \,|\, \mathbf{x}) \cdot p(\mathbf{x}). \tag{1}$$

As is frequently the case, if we assume independence across observations, then we may work component-wise and use the multiplication rule for multiplying distributions of independent quantities to get the joint distribution,

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}) &= \prod_i p(x_i, y_i) \\ &= \prod_i p(y_i \,|\, x_i) \cdot p(x_i) \end{aligned} \tag{2}$$

To illustrate this simple hierarchical probability model (Eqs. 1, 2), consider the problem of measuring leaf-level transpiration rates (T) and stomatal conductance to water vapor (g), and let T_i and g_i represent the i^{th} observation of each. Let us also assume that the mean transpiration rate, denoted as $E(T)$, is given by Fick's law of diffusion, $E(T) = g \cdot \Delta W$, where ΔW is the gradient in water vapor from inside the

leaf to the atmosphere (Pearcy et al. 1989; Nobel 1999). For simplicity, we treat ΔW as a fixed covariate and assume that the mean conductance value is $E(g) = \mu_g$, with $p(T_i \mid g_i)$ and $p(g_i)$ given by Gaussian probability densities with variances σ_T^2 and σ_g^2. That is:

$$T_i | g_i \sim No(g_i \cdot \Delta W, \sigma_T^2) \tag{3}$$

and:

$$g_i \sim No(\mu_{g_i}, \sigma_g^2). \tag{4}$$

We use Eq. 1 to combine the conditional and marginal distributions in Eqs. 3 and 4 to obtain a bivariate normal density for the joint distribution of T_i and g_i:

$$\begin{pmatrix} T_i \\ g_i \end{pmatrix} \sim MNo\left(\begin{pmatrix} \mu_{g_i} \cdot \Delta W \\ \mu_{g_i} \end{pmatrix}, \begin{pmatrix} (\Delta W)^2 \cdot \sigma_g^2 + \sigma_T^2 & (\Delta W)^2 \cdot \sigma_g^2 \\ (\Delta W)^2 \cdot \sigma_g^2 & \sigma_g^2 \end{pmatrix} \right) \tag{5}$$

where the unconditional mean of T_i is $\mu_{gi} \cdot \Delta W$, the unconditional variance of T_i is $(\Delta W)^2 \cdot \sigma_g^2 + \sigma_T^2$, the covariance between T_i and g_i is $(\Delta W)^2 \cdot \sigma_g^2$, and μ_{gi} and σ_g^2 are as given above. In this example, the conditional and marginal distributions for $T_i|g_i$ and g_i are relatively straightforward to specify, but it is less obvious how to specify the joint distribution in Eq. 5 directly, without the relationships in Eqs. 3 and 4. Assuming independence across $i = 1,\ldots, n$ observations gives:

$$p(\mathbf{T}, \mathbf{g}) = \prod_{i=1}^{n} p(T_i, g_i) \tag{6}$$

where $p(T_i, g_i)$ is given by Eq. 5.

Equation 1 is a particular case of a completely general result for obtaining joint distributions from the product of conditional and marginal distributions. For a generic random vector $\mathbf{z} = (z_1, z_2,\ldots, z_n)'$, basic rules of probability allow us to write the joint distribution of $\mathbf{z}$, $p(\mathbf{z}) = p(z_1) \times p(z_2|z_1) \times p(z_3|z_2,z_1) \times \ldots \times p(z_n| z_{n-1},\ldots,z_1)$, where the components z_i may themselves be sub-vectors of $\mathbf{z}$, with the particular sub-vectors and sequencing of conditioning being arbitrary. For example, with $\mathbf{z} = (\mathbf{z}_1', \mathbf{z}_2')'$, we might choose to model $p(\mathbf{z}) = p(\mathbf{z}_1|\mathbf{z}_2) \times p(\mathbf{z}_2)$ or, instead, $p(\mathbf{z}) = p(\mathbf{z}_2|\mathbf{z}_1) \times p(\mathbf{z}_1)$. It should be noted that when building a joint $p(\mathbf{z})$ from marginals and conditionals, the resulting $p(\mathbf{z})$ may differ, depending on the specified marginal and conditional distributions. Whichever way we choose to condition, we get a joint distribution $p(\mathbf{z})$ in the end. For example, we may have specified a conditional distribution for $\mathbf{g}|\mathbf{T}$ and marginal for $\mathbf{T}$, but the resulting joint distribution would generally be different from that in Eq. 5. Again, the usefulness of this simple result is that each conditional distribution $p(\mathbf{z}_j|\mathbf{z}_{j-1},\ldots,\mathbf{z}_1)$ is likely easier to specify, whereas the task of specifying a model for a joint distribution $p(\mathbf{z})$ directly is often beyond our ability – it is simply easier to think about the smaller dimensions of $\mathbf{z}_1$, $\mathbf{z}_2$, through $\mathbf{z}_n$ than to think at the dimensionality of $\mathbf{z}$. The conditional approach allows us to work with relatively simple pieces to arrive at a complex joint model that would otherwise be

difficult or impossible to specify directly. This approach is sometimes loosely referred to as "modeling locally."

Initially, the flexibility of the conditional modeling approach may seem bewildering. Which conditional specification should we choose? But, the particular specification used in practice is often motivated and aided by consideration of the problem at hand, including the available data and ecological or biophysical theory. In other words, complex models of data, quantitative theory, and conceptual understanding can often be accommodated via the flexibility of conditional specification. It is the specification of a joint model via a sequence of conditional models that is generally referred to as hierarchical modeling. Some of the examples that we consider later illustrate how the conditional specification is key to modeling complex ecological processes. Other tools such as graphical models can facilitate the derivation of the corresponding probabilistic model by helping to identify relationships among components, which ultimately informs the conditional specification (see Brooks 1998, and references therein), and which may be particularly useful to ecologists with a conceptual model in mind (e.g. Clark and Gelfand 2006). For example, ecologists often use "systems diagrams" (see Kitching 1983), a form of graphical modeling, to help conceptualize the ecological system, and thus they should find graphical models familiar and helpful in building probability models.

3.2 Bayes Theorem and Bayesian Inference

Of course, we want to put the above discussion on hierarchical probability modeling into a Bayesian statistical framework, and we start with a statement of Bayes theorem, at the foundation of Bayesian statistical inference. The theorem has existed in various forms since the eighteenth century when introduced by Thomas Bayes (Barnard 1958; Malakoff 1999). The result of the theorem is a straightforward consequence of fundamental probability rules. Bayes theorem (or Bayes rule) is without controversy, existing separate from notions of statistical inference:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x})}{p(\mathbf{y})}. \tag{7}$$

To place this into a Bayesian statistical context, consider a probability distribution function representing the distribution of data, **y**, and depending on other quantities, θ (generally a vector), called the parameter(s) of the distribution. Such dependence is often made explicit via notation such as $p(\mathbf{y}|\theta)$ or $p(\mathbf{y}|\mathbf{x}, \theta)$ if considering covariates **x**. Continuing the transpiration–conductance example (Eqs. 3–6), the data are $\mathbf{y} = (\mathbf{T},\mathbf{g})$, the parameters are $\theta = (\mu_g, \sigma_T^2, \sigma_g^2)$, and the covariates are $\mathbf{x} = (\Delta\mathbf{W})$. The distribution $p(\mathbf{y}|\theta)$ is often called the data distribution, the sampling distribution, or, when viewed as a function of θ with **y** fixed at observed data values, it is called the likelihood (of the data). Of course, θ is almost always unknown to some degree, and the objective is to infer about θ, to predict unobserved **y** values, or, more generally, to infer about some function of these quantities.

To enable inference about the parameter θ, we introduce a prior probability distribution $p(\theta)$ – an explicit summary, in the form of a probability distribution, of what we know about θ prior to observing the data. So, with $p(\mathbf{y}|\theta)$ as our sampling distribution and $p(\theta)$ as our prior, then, by the basic probability rule in Eq. 1, we have a "full" probability model for data and parameters,

$$p(\mathbf{y},\theta) = p(\mathbf{y}|\theta) \cdot p(\theta). \tag{8}$$

Bayes theorem is used to actuate the common (but not universally followed) statistical principle of conditioning on what is known (**y**, data) in order to learn about what is unknown (θ, parameters). In other words, once the probability model is specified, Bayes theorem provides a complete prescription for statistical inference. It is often said that Bayes theorem "updates" prior knowledge of θ as quantified by $p(\theta)$ to reflect the information in the data, arriving at the posterior (conditional) distribution of θ (given the data **y**),

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta) \cdot p(\theta)}{p(\mathbf{y})} \tag{9}$$

We provide a detailed example in Section 4.1, similar to the transpiration–conductance problem, of how to arrive at the posterior distribution of unknown quantities.

Usually, only in the simplest of cases are we able to use analytic techniques to obtain a closed form solution to $p(\theta|\mathbf{y})$, and often we must resort to numerical or stochastic techniques to approximate the distribution $p(\theta|\mathbf{y})$ or to obtain a sample of θ values from $p(\theta|\mathbf{y})$. Widespread use of Bayesian methodology occurred with the development of computing power (Malakoff 1999) and concomitant developments of algorithms for sampling from distributions (e.g. Gelfand et al. 1990; Gilks et al. 1996; Gamerman and Lopes 2006). These sampling methods are generally referred to as Markov chain Monte Carlo (MCMC) methods (Gilks et al. 1996; Brooks 1998; Cappe and Robert 2000; Robert and Casella 2004; Gamerman and Lopes 2006), including Gibbs (Gelfand et al. 1990; Casella and George 1992), Metropolis (Metropolis et al. 1953; Robert and Casella 2004), and Metropolis–Hastings algorithms (Hastings 1970; Chib and Greenberg 1995). Although these methods are not inherently Bayesian, their usefulness for sampling from the posterior distribution has made them standard fare in Bayesian methodology. Together, MCMC methodology and Bayesian inference are becoming a very useful combination for complex models (e.g. Clark and Gelfand 2006). We do not cover MCMC for Bayesian inference but refer the reader to the above references, especially the texts by Gamerman and Lopes (2006), Gilks et al. (1996), and Robert and Casella (2004).

Upon obtaining a sample of θ values from $p(\theta|\mathbf{y})$, inference about θ then proceeds via ordinary computation of means, medians, quantiles, credible regions, etc. (e.g. Smith and Gelfand 1992), exactly analogous to summarizing a sample of data. Furthermore, the (marginal) posterior distribution and summaries for any function $f(\theta)$, or more generally, for any function $f(\theta, \mathbf{y}, \mathbf{x})$, may then be obtained

in a straightforward manner as well. For example, for each value of θ, $f(\theta)$ is computed, yielding a sample of $f(\theta)$ values, which can be summarized in the same way as θ.

All of the most popular computational methods for obtaining samples from posterior distributions depend only on knowing $p(\theta|\mathbf{y})$ up to some multiplicative constant, a factor that does not depend on θ. When the data are given, $p(\mathbf{y})$ is constant with respect to θ, and we often see an alternative form of Bayes theorem:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta) \cdot p(\theta), \tag{10}$$

and we often hear the chant "posterior is proportional to likelihood times prior."

3.3 Hierarchical Bayesian Modeling

Recalling the previous discussion (Section 3.1) of modeling joint distributions via conditional specifications, we see that the right-hand side of Bayes theorem as expressed in Eq. 10 is the joint distribution $p(\mathbf{y},\theta)$. Thus, the enterprise of Bayesian statistical modeling can be seen as the development of a full probability model $p(\mathbf{y},\theta)$ for the data and unknown parameters, inference being fully prescribed, in principle, via Bayes theorem once the model is specified. Once again it is important to note that we may avail ourselves of the flexibility of the hierarchical specification of joint distributions, which is especially important for complex modeling problems – modeling locally and, now, inferring globally so-to-speak.

Before proceeding to discuss hierarchical Bayesian modeling in more detail, we remind the reader that hierarchical statistical modeling is, of course, not exclusively Bayesian. In fact, we suspect that the vast majority of literature on hierarchical modeling remains nonBayesian despite recently increasing activity in hierarchical Bayesian modeling. The literature on hierarchical statistical modeling is extensive and diverse. We provide only a short, incomplete and somewhat idiosyncratic set of references that is, nonetheless, sufficient to allow the interested reader to explore this literature more fully.

Hierarchical statistical modeling, especially its nonBayesian versions, tends to fall under the guises of repeated measurement models (Lindsey 1993), longitudinal data models (e.g. Diggle et al. 2002), multilevel models (Goldstein 2002) or, more generally, mixed effects models. Laird and Ware (1982) are frequently cited for their work on linear mixed effects (LME) models, and Lindley and Smith (1972) introduce the hierarchical Bayesian linear model. Lindstrom and Bates (1990) discuss the nonlinear mixed effects model (NLME), with Gelfand et al. (1990) providing a Bayesian perspective. The extension of hierarchical models to nonGaussian data is often attributed to Breslow and Clayton (1993) for their treatment of generalized linear mixed models (GLMM). The text by Searle et al. (1992) is a comprehensive review of LME and variance component models. Davidian and Giltinan (1995) present a nice discussion of linear and nonlinear hierarchal models, and Gelman et al.

(2004) provides a general introduction to Bayesian methodology with chapters on hierarchical modeling. Popular software for fitting nonBayesian hierarchical models includes the R (R Development Core Team 2006) package by Pinheiro et al. (2006) for LME and NLME models (see also Pinheiro and Bates 2000), the SAS MIXED procedure for LME models, and SAS %NLINMIX macro for NLME models (SAS Institute 2001; see also Little et al. 1996). The SAS GLIMMIX procedure is designed for fitting GLMM models (SAS Institute 2006). WinBUGS (Lunn et al. 2000) is a general purpose software package for implementing Bayesian models, including hierarchical Bayesian models.

So, hierarchical statistical modeling is neither new nor exclusively Bayesian. Why focus on hierarchical Bayesian models? Hierarchical Bayesian modeling is proving to be the primary methodology to study complex relationships embodied in process models, data, and prior information, and it is the only method that allows for explicit incorporation of prior information. Berliner (1996) is often credited with introducing what has become loosely referred to as the "process sandwich" as an idealized way of thinking about how to incorporate the information contained in a process model into a Bayesian statistical framework along with (i.e., "between") the data and prior information. Similar to Eq. 10, the process sandwich is:

$$p(\text{process}, \text{parameters} \mid \text{data}) \propto$$

$$p(\text{data} \mid \text{process}, \text{parameters}) \cdot p(\text{process} \mid \text{parameters}) \cdot p(\text{parameters}), \quad (11)$$

and Bayes theorem and (usually) MCMC methodology are applied to learn about the (unknown or latent) process and parameters via the posterior p(process, parameters|data). In this context, the process is now modeled as a stochastic quantity with probability distribution p(process|parameters) with a mean that is usually specified by a deterministic process model – what we have been calling the "process model" up to this point. Now, we may refer to the process (probability) distribution model or to its mean, the process model. The process model may describe a theoretical physical process in the form of, say, partial differential equations, but in its most general form may have some combination of mechanistic, semi-mechanistic, phenomenological, empirical, or stochastic components that represent, for example, real underlying physiological, ecological, or biogeochemical dynamics.

The process sandwich approach is espoused and illustrated by Wikle and others in, for example, Wikle et al. (1998), Wikle et al. (2001), and Wikle (2003b); with much of this work having a spatial or spatial-temporal statistical bent within a geophysical application. Wikle (2003a) applies similar methodology in modeling the spread of the house finch (*Carpodacus mexicanus*) using a reaction–diffusion partial differential equation, and Clark and colleagues have applied the approach to a variety of population and community ecology problems (e.g. Clark 2003; Clark et al. 2003b, 2005). We return to hierarchical Bayesian modeling and the process sandwich in our examples below.

4 Example Applications

In this section, we provide some specific examples of Bayesian and hierarchical Bayesian applications in plant physiological and ecosystem studies. None of the physiological- and ecosystem-related papers in the literature survey (Table 1, Fig. 1) employed a complete hierarchical Bayesian framework as defined by Eq. 11. That is, although many incorporated complex ecological process models, none explicitly modeled process error – the process is treated as deterministic rather than stochastic. Thus, in Section 4.1, we begin by providing our own detailed example of a hierarchical Bayesian modeling problem using process models of photosynthesis and stomatal conductance with leaf-level gas exchange data. In Section 4.2, we then highlight three published examples that use (nonhierarchical) Bayesian methods with informative priors to parameterize complex ecosystem simulation (process) models.

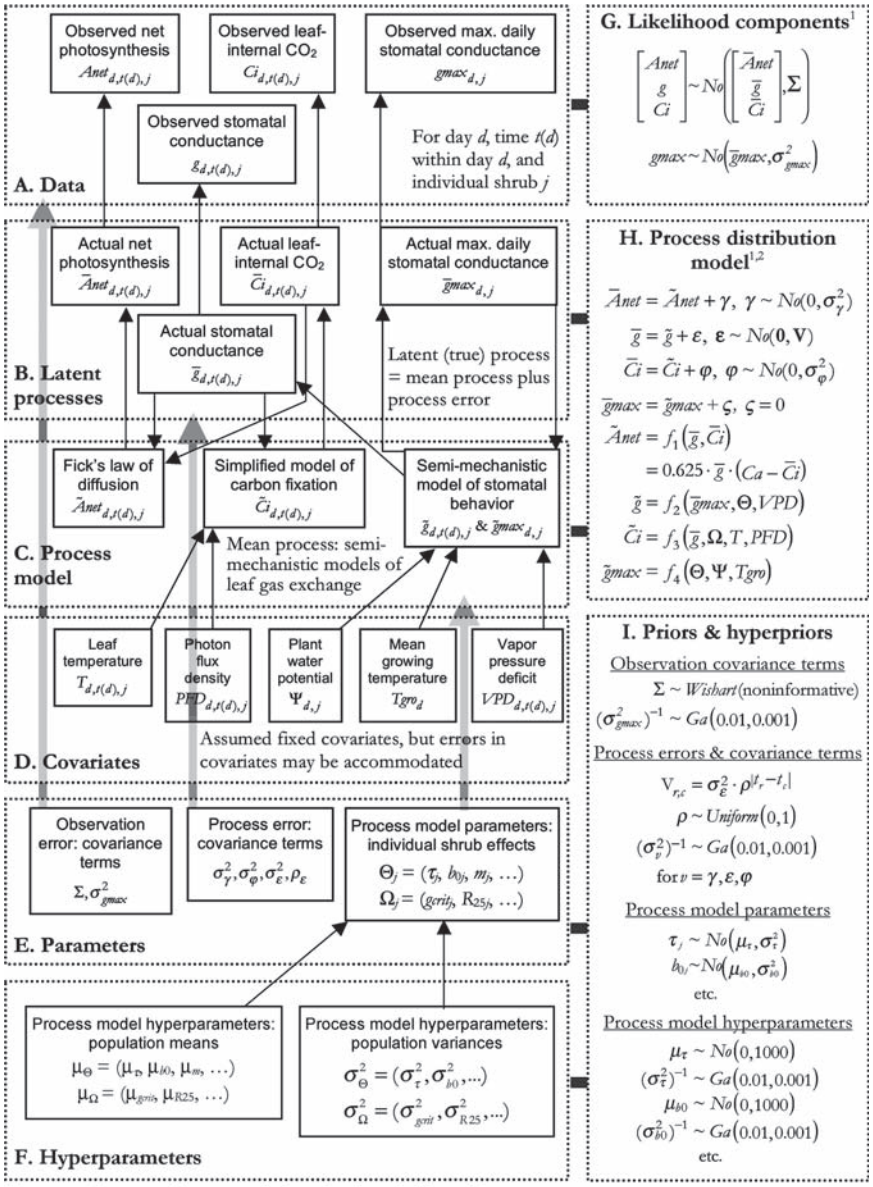
4.1 Example 1: Leaf-Level Gas Exchange

4.1.1 The Problem

This example builds on a study by Ogle and Reynolds (2002) that focused on *Larrea tridentata* (creosote bush), a common evergreen shrub of the hot deserts of North America (Barbour et al. 1977). *Larrea* experiences large year-to-year, seasonal, and diurnal fluctuations in temperature and water availability, yet it remains metabolically active year-round (e.g. Oechel et al. 1972) due to its ability to tolerate severe temperature and water stress. Thus, *Larrea* makes an interesting candidate for studying plant physiological strategies for coping with environmental stress in field settings. A primary goal of this study was to learn how temperature acclimation and stomatal control of photosynthesis in *Larrea* contribute to its ability to tolerate severe water and temperature stress. Ogle and Reynolds (2002) combined extensive field data with physiological process models to tease apart potential mechanisms underlying *Larrea*'s ability to tolerate harsh arid conditions.

4.1.2 The Data

Ogle and Reynolds (2002) used an open-system infrared gas exchange analyzer to measure leaf-level, instantaneous net photosynthesis (*Anet*), stomatal conductance to water vapor (g), and leaf-internal CO_2 concentration (Ci). Diurnal gas-exchange measurements were made on eight shrubs on each of 16 days throughout the spring and summer seasons in the Chihuahuan desert of New Mexico. Additional stomatal conductance measurements were made with a steady-state porometer on 16 shrubs (including the eight above) and 25 days



[1]For ease of presentation, we omit the subscripts d, $t(d)$, and j.

[2]Process errors represent diurnal (time within day) effects, and exploratory analyses suggest that the $\bar{g}$ process errors are temporally autocorrelated, but the $\bar{A}net$ and $\bar{C}i$ process errors are independent with respect to time of day. The $\bar{g}$ process errors were nonidentifiable, thus we set them equal to zero. See Ogle & Reynolds (2002) for the explicit equations used to specify f_2, f_3, and f_4; note that f_1 is Fick's law.

Fig. 2 Graphical model (*Boxes A–F*) and corresponding distribution and process models (*Boxes G–H*) associated with the leaf-level gas exchange example in Section 4.1

(including the 16 days above) to estimate daily maximum conductance (*gmax*). Referring to Fig. 2 (Box A), *d* indexes the day of observation, *t*(*d*) refers to the time of measurement within day *d*, and *j* indexes the shrub. Using the notation introduced in Section 3.1, the dataset includes 934 observations of the triplet $(Anet_{d,t(d),j}, g_{d,t(d),j}, Ci_{d,t(d),j})'$, 218 observations of $gmax_{d,j}$, and various biotic and abiotic covariates that were measured at the same time (see Fig. 2, Box D). See Ogle and Reynolds (2002) for a detailed account of the field experiment and sampling design and methods.

4.1.3 The Process Model

Ogle and Reynolds (2002) modified the Hybrid model (Katul et al. 2000) to describe *Larrea*'s *Ci* dynamics. This model is derived from simplified versions of an A-Ci curve (e.g. see pp 17–19 in Lambers et al. 1998) and a biochemical model of CO_2 assimilation (Farquhar et al. 1980; Farquhar and Sharkey 1982; Farquhar and von Caemmerer 1982). Stomatal conductance is an important factor controlling *Ci*. Instantaneous *g* is modeled as a threshold-type function of *gmax* and vapor pressure deficit (*VPD*) in such a way that agrees with a biophysical model of how stomata regulate leaf water potential (see Oren et al. 1999; Ogle and Reynolds 2002); *gmax* is assumed to vary with plant water stress and growing temperature. Once the *Ci*, *g*, and *gmax* models are specified, than net assimilation (*Anet*) is simply given by Fick's law of diffusion. The coupling of *Ci*, *g*, *gmax*, and *Anet* is illustrated in Fig. 2 (Boxes B, C, H), and we cover more details in the next section.

4.1.4 Data–Model Integration

We develop a hierarchical Bayesian model to couple the data and process models discussed above. We provide a detailed discussion of how the conditional modeling framework established in Section 3 can be applied to the *Larrea* leaf-level gas exchange problem. Although this is a specific example, it is representative of the types of data and process models that are commonly encountered by plant ecophysiologists, and the steps that we discuss here are applicable to a variety of problems. Of course, in practice, a primary goal of implementing such a framework is to obtain the posterior distribution of all ecologically meaningful process parameters, latent variables, and covariance terms that describe errors in observation (i.e. measurement) and process (i.e. model), thereby updating our understanding of the underlying physiological system. However, our goal here is to emphasize the means by which to arrive at the posterior distribution.

Using notation similar to that in Eq. 11, we present the stages of the hierarchical model. We elaborate on Eq. 11 by (1) partitioning the "parameters" into "process parameters", "process hyperparameters", and "covariance parameters", (2) renaming "process" as "latent process" and introducing explicit dependence on the deterministic

"process model", and (3) indicating explicit dependence on "covariates". Thus, the posterior distribution for all unknown parameters and latent processes:

$$p(\text{parameters, latent process} \mid \text{process model, data, covariates}) \cdot \quad (12)$$

is proportional to:

$$p(\text{data} \mid \text{latent process, covariance parameters}) \cdot \quad (12.1)$$

$$p(\text{latent process} \mid \text{process model, process parameters, covariates}) \cdot \quad (12.2)$$

$$p(\text{process parameters} \mid \text{process hyperparameters}) \cdot p(\text{covariance parameters}) \cdot \quad (12.3a)$$

$$p(\text{process hyperparameters}) \cdot \quad (12.3b)$$

The first stage (Eq. 12.1) is the likelihood of the data, which corresponds directly to p(data|process, parameters) in Eq. 11. The second stage (Eq. 12.2) is the stochastic (or probability) model for the "unobservable" latent processes, i.e. p(process|parameters) in Eq. 11. The third stage (Eqs. 12.3a, b) is p(parameters) in Eq. 11 and is modeled, using an assumption of independence of process-related parameters and covariance parameters and applying the hierarchical specification in Eq. 1 to p(process parameters, process hyperparameters). Notice that the joint prior is specified hierarchically, and we may want to distinguish different levels of the joint prior as the "prior" (Eq. 12.3a) and the "hyperprior" (Eq. 12.3b).

In the following sections, we discuss each of the above stages (Eqs. 12.1–12.3) in the context of the leaf gas exchange problem. Following Clark (2005) and Clark and Gelfand (2006), we provide a graphical model that illustrates these stages (Fig. 2, Boxes A–F), and we also provide a summary of the probability distributions associated with each stage (Fig. 2, Boxes G–I).

Stage 1

The first stage is represented by Boxes A and G (Fig. 2), which denote the likelihood components of individual observations of $(Anet_{d,t(d),j}, g_{d,t(d),j}, Ci_{d,t(d),j})'$ and $gmax_{d,j}$. The gas-exchange instrument software computes "observed" values of *Ci* from "observed" photosynthesis rates (*Anet*), transpiration rates, and stomatal conductance (*g*; LI-COR 2004). Similarly, *Anet* and *g* are computed from other measured quantities (e.g. mole fractions of water vapor and CO_2 and flow rates). Thus, it seems reasonable to consider the "observation" errors in *Ci*, *Anet*, and *g* to be correlated, and we assume a multivariate normal distribution for $(Anet_{d,t(d),j}, g_{d,t(d),j}, Ci_{d,t(d),j})'$, as given in Box G of Fig. 2 [subscripts d, $t(d)$, and j are omitted for ease of presentation]. (In fact, the posterior distribution of Σ supports this assumption.) Given that a different instrument was used to measure *gmax*, we assume independence of *gmax* observations and use a univariate normal distribution to describe the likelihood of $gmax_{d,j}$ (Fig. 2, Box G). The likelihood of all data (i.e. **Anet**, **g**, **Ci**, **gmax**)

is obtained by multiplying the likelihood components (Fig. 2, Box G) across all days, shrubs, and times, giving:

$$p(\text{data} = \{\mathbf{Anet, g, Ci, gmax}\} | \text{latent process, covariance parameters}) =$$

$$\prod_{d}\prod_{j} No\left(gmax_{d,j} \left| \bar{g}max_{d,j}, \sigma^2_{gmax}\right.\right) \left(\prod_{t(d)} No\left(\begin{bmatrix} Anet_{d,t(d),j} \\ g_{d,t(d),j} \\ Ci_{d,t(d),j} \end{bmatrix} \middle| \begin{bmatrix} \bar{A}net_{d,t(d),j} \\ \bar{g}_{d,t(d),j} \\ \bar{C}i_{d,t(d),j} \end{bmatrix}, \Sigma \right) \right) \quad (13)$$

(see examples discussed with Eqs. 2, 6). Note that when the data are conditioned on the true or latent processes (i.e. $\bar{\boldsymbol{A}}\boldsymbol{net}$, $\bar{\boldsymbol{g}}$, $\bar{\boldsymbol{C}}\boldsymbol{i}$, $\bar{\boldsymbol{g}}\boldsymbol{max}$), as indicated in Eq. 13, then any remaining unexplained variability should be attributable to measurement error, and we assume that these errors are uncorrelated across shrubs and time. Although this is a common assumption, it may not always be appropriate depending on the measurement instrument used.

Stage 2

The second stage (Eq. 12.2) is depicted by Boxes B, C, D, and H in Fig. 2. Stage 2 embodies the specification of the stochastic and deterministic process models (Fig. 2, Boxes B, C, H), and we incorporate process error into the stochastic process model (Box H). The process error accounts for the fact that we are approximating the true photosynthetic and stomatal dynamics with relatively simple equations that cannot reproduce the truth exactly. Thus, we model the latent process as the mean process plus some random process error (see Fig. 2, Box H).

One approach to modeling process errors is to assume that they are independent. However, if observation errors are also assumed independent, then this may lead to identifiability problems – i.e. process and observation variance components may not be disentangled – unless a tight prior is specified for the variance components of the observation error. Instrument manuals often provide estimates of instrument precision, which can be used to construct tight (or informative) priors for observation error variances. In the absence of informative (or informative) priors, identifiability of process error and observation error variance terms may be facilitated by incorporation of different error structures for process and observation errors. We cannot possibly measure and account for all factors affecting the true process, so we cannot model the true process perfectly via the process mean alone and are led to consider the structure of the process error. These unobserved factors are almost invariably temporal, spatial, or biological in nature and usually have little impact on measurement error. Thus, we may want to incorporate temporal, spatial, or biological structure into the process errors.

In the leaf-gas exchange problem, we did not have sufficient information to construct tight priors for the observation covariance parameters. Thus, for illustration, we opted to incorporated temporal structure into the *Anet*, *g*, and *Ci* process errors

(Fig. 2, γ, ε, and φ in Box H). Based on previous experience (Ogle and Reynolds 2002), we assumed diurnal effects such that each time period within a day gets a particular random effect (error) that applies to all days. For example, the diurnal effect at, say, noon is the same for all days – effects do not vary by day. Exploratory analysis with different temporal error structures suggested the *Anet* and *Ci* diurnal effects were uncorrelated with respect to time. However, the *g* diurnal effects appeared autocorrelated, and we used an exponential covariance function – defined by the variance parameter σ_ε^2 and the correlation coefficient ρ – to describe this correlation (see Fig. 2, Boxes H, I). The shrub itself is where the physiological processes are taking place, but shrubs differ in their genetic composition, age, and growth history. Thus, rather than incorporating additive shrub effects into the process error, we allow the process parameters themselves (i.e. Θ_j and Ω_j in Box E) to vary by shrub, with distributions specified in Stage 3. Specification of the deterministic process model is highly dependent on existing ecological theory, the understanding of the specific system of interest, and the particular objectives of the study. Process models have a long history in ecology, and we do not cover details of process model development. However, the process model is key because it contains important information about the ecological system, and we illustrate how it is incorporated into the hierarchical Bayesian modeling framework. In this example, each shrub *j* gets a unique process mean (i.e. Fig. 2, $\tilde{A}net$, $\tilde{g}$, $\tilde{C}i$, $\tilde{g}max$ in Boxes C, H) because the deterministic process models (Box C) depend on the shrub-specific process parameters (Θ_j and Ω_j in Box E). The specific deterministic process models that we employ in this example are discussed in Section 4.1.3 and their equations are outlined in Box H (Fig. 2). The shrub-specific process model parameters (Boxes E, I) include, for example, maximum potential conductance (τ_j), sensitivity of stomata to changes in *VPD* (represented by b_{0j} and m_j), leaf dark respiration rate at 25 °C (R_{25j}), and so forth (for a list of parameters, see Ogle and Reynolds 2002). The process models also incorporate the effects of important plant and environmental covariates (see Box D for specific covariates, Box H for model dependencies on covariates). In this example, we assume that the covariates are measured without error, which is probably a reasonable assumption because there is likely relatively little error associated with measuring these quantities compared to "measuring" *Anet*, *g*, *Ci*, and *gmax*. However, it is fairly straightforward to incorporate errors in covariates into a hierarchical modeling framework (Clark et al. 2003b).

Stage 3

Finally, Stage 3 (Eq. 12.3) involves defining the priors for all parameters (process parameters, covariance parameters, process hyperparameters; Fig. 2, Boxes E, F, I). As mentioned above, we lack information on the instrument measurement error and thus use fairly noninformative (relatively diffuse) priors for the observation covariance parameters (see *Wishart* for Σ and gamma for precision term σ^2_{gmax} in Box I). Similarly, we specified noninformative priors for the process error variance parameters [see gammas for $(\sigma_\gamma^2)^{-1}$, $(\sigma_\varepsilon^2)^{-1}$, and $(\sigma_\varphi^2)^{-1}$ in Box I] and a uniform (flat) prior

for the correlation parameter ρ (Box I). All distribution parameterizations follow Gelman et al. (2004).

As noted in Stage 2, we assume that the process parameters (Θ_j and Ω_j, Box E) vary by shrub. The shrub effects are viewed as coming from population distributions with means μ_Θ and μ_Ω and variances σ_Θ^2 and σ_Ω^2 (see Box E). For example, Θ_j represents the shrub-specific process parameters related to the stomatal conductance models (see f_2 and f_4 in Box H), and let us consider one element of Θ_j, maximum potential conductance (τ_j). We first note that τ_j is a positive-valued quantity, and thus we model $\log(\tau_j)$ – which can be positive- or negative-valued – as coming from a population distribution, with mean μ_τ and variance σ_τ^2, that describes shrub-to-shrub variability in τ_j,

$$\log(\tau_j) \sim No(\mu_\tau, \sigma_\tau^2). \tag{14}$$

We assume a priori that the elements of Θ_j and Ω_j are independent given μ_Θ, μ_Ω, σ_Θ^2, and σ_Ω^2, and thus priors similar to Eq. 14 are specified for each element of Θ_j and Ω_j (Box I). We assume independence of the covariance parameters and the shrub-specific parameters so that the complete prior in Eq. 12.3a is given by the multiplication of the marginal priors specified in first three sections of Box I (Fig. 2), analogous to Stage 1, Eq. 13. Note that the simplifying modeling assumption of a priori independence does not necessarily imply a posteriori independence.

Finally, we specify the hyperprior (Eq. 12.3b) for the hyperparameters (i.e. μ_Θ, μ_Ω, σ_Θ^2, σ_Ω^2) describing the population distributions of shrub effects. Continuing to illustrate using τ_j, we specify independent diffuse priors for μ_τ and σ_τ^2 (process model hyperparameters, Box I), with similar a priori independent priors for the remaining hyperparameters, and the full hyperprior (Eq. 12.3b) following again by multiplication of distributions. Note that in most cases, we are actually most interested in learning about such hyperparameters (population-level parameters); for example, here, μ_Θ and μ_Ω tell us about the physiological behavior of *Larrea* in general (we may not be interested in the behavior of a particular shrub).

Now that we have described the various components making up the different stages of the hierarchical model, some in more detail than others, they may be combined to yield the full conditional specification in Eq. 12. The resulting model is rather cumbersome, thus we omit the technical details. Moreover, we implemented the model in WinBUGS, and the WinBUGS environment does not require one to specify the fully probability model in Eq. 12, but only requires one to specify each of the conditional components, just as we have done here.

4.1.5 Comments

We do not discuss specific results of the analysis because this is not the objective of this example, but we would like to address the question: now that we have samples from the posterior, what to we do with them? We can summarize the posterior results by computing posterior statistics such as means, variances, and percentiles (a measure

of "uncertainty") for each parameter of interest. We may also want to estimate correlations betweens parameters or to learn about the distributional form of the marginal posterior distributions. For example, many marginals are often normal-like, but some can exhibit interesting behavior such as a high degree of skewness or multiple modes. Multiple modes may indicate indentifiability problems that may be addressed via tighter priors, reparameterization, or restructuring of related model components. In principle, we can obtain the posterior distributions for most any ecologically relevant functions of the parameters, data, or covariates. For example, we could compute daily-integrated photosynthesis from predictions of diurnal photosynthesis (e.g. $\bar{A}day_{d,j} = \Sigma_{t(d)} \bar{A}net_{d,t(d),j}$), yielding a posterior distribution for $\bar{A}day_{dj}$. We could also use the joint posterior distribution of the parameters within a forward modeling context to learn how uncertainty in the photosynthesis and conductance parameters are propagated to model output related to leaf-level gas exchange under, for example, different scenarios of temperature and water stress. We discuss the use of the posterior in forward modeling simulations in Section 4.2. How does the hierarchical Bayesian modeling approach compare with other, more traditional approaches? We note that Ogle and Reynolds (2002) fitted the aforementioned process models to the gas exchange using a piece-wise approach that proceeded as follows: (1) fit *gmax* observations to the *gmax* process model using nonlinear least squares, (2) fit the *g* observations to the *g* process model using a nonlinear mixed effects (NLME) approach, with the *gmax* process parameters fixed at the point estimates obtained in (1), and (3) fit *Anet* observations to the *Anet* process model using a NLME approach, with *gmax* and *g* parameters fixed at the point estimates obtained in (1) and (2). The NLME approach was considered a fairly rigorous and state-of-the-art approach at the time (e.g. Ogle and Reynolds 2002; Peek et al. 2002).

In hindsight, this piece-wise approach is unsatisfactory in several ways. First, standard NLME packages cannot accommodate a multivariate likelihood, thus compelling an assumption of independence of the *Anet*, *g*, and *Ci* observations, but the posterior distribution of Σ (Fig. 2, Box G) suggested strong dependence. Second, uncertainty in parameter estimates could not be propagated from one model fit to the next. And, third, there were convergence issues with the NLME routines such that some of the parameters could not be estimated and thus were fixed at "realistic" values to ensure convergence of the estimates of the other parameters. The hierarchical Bayesian analysis was not subject to these problems, and more importantly, it simultaneously integrates all data sources and the interrelated process models.

4.2 Example 2: Use of Prior Information with Process Models

4.2.1 The Problem

In this example, we highlight three studies that discuss and illustrate the utility of Bayesian methods for parameterizing relatively detailed ecosystem process models. These models include the coupled hydrologic-nitrogen cycle model SINIC

(Hong et al. 2005), the ecosystem-level photosynthesis and evapotranspiration model SIPNET (Braswell et al. 2005), and the forest carbon and nitrogen simulation model BASFOR (Van Oijen et al. 2005). The primary goal of each study was to demonstrate how rigorous parameter estimation could be accomplished within a simple Bayesian framework that incorporates different types of data and prior information. Similar across all studies is the specification of priors based on knowledge embodied in the scientific literature. Each study also addresses the questions: how "well constrained" are the parameters; how well does the data update prior knowledge and uncertainty about the parameters; and, how does parameter uncertainty propagate to model output uncertainty?

4.2.2 The Data

The three studies vary in the types and amount of data that they use to construct the likelihood. Hong et al. (2005) used long-term (1964–1994) measurements of annual streamflow and annual streamflow nitrate flux from a watershed located in the Hubbard Brook Experimental Forest in New Hampshire. Braswell et al. (2005) used net ecosystem CO_2 exchange (NEE) data for daily daytime and nighttime intervals for the period 1992–2002. NEE values were calculated from eddy flux measurements made at the Harvard Forest Long-Term Ecological Research site in Massachusetts. Van Oijen et al. (2005) utilized forest inventories from Skogaby, Sweden that provided measurements of, for example, tree heights, carbon and nitrogen contents of different plant and soil pools, leaf area index, and net primary productivity (data from Schulze 2000). In all studies, the data are directly related to outputs generated by the process models.

4.2.3 The Process Model

Although the three models were constructed for dissimilar purposes and represent different ecological systems, they are similar in their level of complexity. Each model contains a series of interrelated equations describing the daily dynamics of various ecosystem processes of interest, and the process models require a moderately large number of parameters. For example, SINIC simulates a variety of hydrologic processes (e.g. evapotranspiration, vertical and horizontal soil water flux, runoff, groundwater flow) and nitrogen cycling processes (e.g. atmospheric deposition, mineralization, plant uptake, nitrification, denitrification, horizontal and vertical fluxes in the soil, discharge to streams; Hong et al. 2005). Twenty parameters are needed for the hydrologic submodel and 17 for the nitrogen submodel. SIPNET is derived from the canopy photosynthesis and evapotranspiration model PnET-Day (Aber et al. 1996), and it models NEE as the difference between respiration and gross primary productivity. These flux processes depend on 25 parameters related to the initial pool values, photosynthesis, autotrophic and heterotrophic respiration, and plant–soil water relations (Braswell et al. 2005). BASFOR is a forest simulator

that contains 11 state variables representing pools of carbon and nitrogen in different tree, soil, and litter compartments (Van Oijen et al. 2005). The model produces a variety of outputs including, but not limited to, time series of the state variables, net primary productivity, and leaf area index. BASFOR has 39 parameters related to, for example, initial pool values, carbon allocation, tree morphology, tissue turnover rates, decomposition, and environmental response functions. All models require similar climate-related input data such as precipitation, light, and temperature.

4.2.4 Data–Model Integration

Unlike Example 1 (Section 4.1), none of the three studies couch the Bayesian data–model integration within a hierarchical framework. The process modeling stage is based on the deterministic output of the simulation models, and there are no random effects associated with, for example, different trees, time periods, or locations. An important element of all three studies, however, is the use of prior information obtained from the scientific literature. Braswell et al. (2005) and Van Oijen et al. (2005) implemented relatively wide uniform priors for all stochastic parameters. The upper and lower bounds were primarily derived from the literature, but some were based on conventional knowledge or ranges of values used in other similar ecosystem models (e.g. Braswell et al. 2005). Hong et al. (2005) implemented a set of rules for choosing the type of distribution and the level of prior uncertainty. Most priors were bounded either with a uniform distribution (for parameters with high prior uncertainty) or a beta distribution (for parameters with intermediate prior uncertainty). Normal distributions were used for those parameters where direct data were available, and the mean and variance were computed from available data. Hong et al. (2005) also incorporated logical constraints that eliminated physically or biologically impossible parameter values. For example, “day of leaf out” must be less than “day of leaf fall” and “soil water content at field capacity” must be less than “saturated soil water content”.

4.2.5 Comments

We applaud Braswell et al. (2005), Hong et al. (2005), and Van Oijen et al. (2005) for introducing and demonstrating the utility of Bayesian data–model integration methods to their respective communities. The Bayesian parameter estimation approach that they employed could be easily adapted to a hierarchical modeling framework, and Braswell et al. (2005) acknowledge that the approach can be extended to include “structural failures in the model” – or process error as defined herein. For example, the SIPNET model had difficulty capturing interannaul patterns of NEE (Braswell et al. 2005), indicating that the incorporation of temporal process error may be appropriate here.

These studies provide valuable insight into the identification of parameters that were constrained by the observed data. For example, Braswell et al. (2005) found that approximately half of the parameters were poorly constrained (little updating of the prior), or their marginal posterior distributions were concentrated at one of the edges of the uniform prior, suggesting discordance between prior and data. The other half were well constrained by the NEE data such that their marginal posteriors were concentrated inside the uniform prior. Hong et al. (2005) combined a sensitivity analysis of the SINIC model with a rigorous evaluation of the posterior distribution to identify key process parameters controling streamflow and steamflow nitrate flux. Perhaps not surprisingly, they found that, in general, those parameters having the most influence on the output of the SINIC model were also those that were most well constrained by the data. They also note that, because the SINIC model is relatively insensitive to some of the other parameters, improved (less uncertain) estimates of such parameters may not necessarily improve predictions of steam nitrate flux.

How can the posterior distributions of poorly constrained parameters be improved? Aside from certain types of parameters like, for example, the individual shrub random effects Θ_j and Ω_j of the *Larrea* example of Section 4.1, more data is often the answer. But, what kind of data and how much? To answer this question, one could conduct a statistical experiment similar to that carried out by Van Oijen et al. (2005). They systematically left out different data sources, reran the Bayesian analysis, and evaluated the resulting posteriors. This allowed them to identify the most useful data – for example, tree height growth data were most informative for the BASFOR model – providing a guide for future data collection efforts. Although Braswell et al. (2005) constructed the likelihood function based on only one data source (i.e. computed NEE values), they noted that the incorporation of other data sources – such as water vapor fluxes, soil respiration measurements, and carbon stock assessments – would likely improve estimates of poorly constrained parameters. Because the model parameters are associated with a variety of different, yet interrelated processes, the incorporation of diverse datasets reflecting these different processes is expected to refine our understanding of the process parameters and underlying ecological mechanisms. A primary goal of Bayesian parameter estimation is to produce "good" estimates of the process model parameters that are informed by information-rich data sources. Given these parameter estimates, one may wish to use the process model in a series of simulation experiments to explore, for example, potential effects of environmental perturbations (e.g. altered precipitation, elevated CO_2, increased nitrogen deposition) on model behavior, with implications for real system responses. As mentioned earlier, such simulations are often referred to as "forward modeling," but the incorporation of the Bayesian parameter estimates adds a new and improved (i.e. data- and prior-informed) dimension to this predictive modeling approach. For example, we could carry out Monte Carlo simulations whereby process parameters are drawn from the joint posterior distribution, input to the process model, a process model simulation is initiated, and the resulting output is stored. Repeated simulations are conducted and distributions of simulation outputs – induced by the posterior distribution of the process parameters – can be evaluated. Drawing parameter values from the joint posterior distribution

preserves important correlations between parameters, and Monte Carlo simulations that draw parameters independently (e.g. from the priors) are likely to draw "biologically unrealistic" combinations and greatly over-estimate model output uncertainty (see Van Oijen et al. 2005).

5 Final Remarks

We conclude with three sections on various issues that we consider important to our treatment of hierarchical Bayesian statistical modeling of complex ecological systems. Section 5.1 discusses the relevance of this modeling framework to the issue of scaling in plant physiological and ecosystem ecology. Section 5.2 provides some practical computational suggestions and a brief discussion of convergence diagnostics and model checking. The final section (5.3) offers some direction on statistical education and practice.

5.1 Scaling Plant Physiological and Ecosystem Ecology

One particularly exciting aspect of the hierarchical Bayesian modeling framework is the potential to link plant and ecosystem processes that span a range of temporal, spatial, and biological scales. For example, we may want to learn about the underlying mechanisms controlling carbon fluxes and stocks at different scales – spanning leaves, canopies, soils, and whole-ecosystems – and how mechanisms operating at one scale may feed back to another scale. Process models are very useful tools for linking mechanisms across scales (Rastetter et al. 2003), but a rigorous coupling of such models with data that also span a range of scales is generally lacking. Importantly, the process models contain ecologically and biophysically meaningful parameters related to scale-dependent mechanisms, and some parameters may even be important to scaling from one level to another. A hierarchical Bayesian integration of process models and datasets spanning different scales provides a method for obtaining "good" estimates of these process parameters.

For example, consider linking leaf-level gas exchange (fast, small scales), ecosystem carbon fluxes (fast, larger scales), and ecosystem carbon stocks (slow, larger scales). We may have data related to each of these processes, such as measurements of leaf-level photosynthesis, canopy-level gas exchange, soil respiration, net ecosystem exchange, and soil carbon stocks. We could use the conditional specification introduced earlier to write, for example, the likelihood of the carbon stock data given latent carbon stocks (predicted by an ecosystem carbon cycle model; part of Stage 1, Eq. 12.1), the distribution of latent carbon stocks given latent soil respiration, latent net ecosystem exchange, and latent initial carbon stocks (part of Stage 2, Eq. 12.2), the likelihood of the net ecosystem exchange data given latent net ecosystem exchange (part of Stage 1), and so forth.

Note that data related to a specific process are conditioned on the associated latent process, as described in Section 4.1 (Fig. 2, Box G). Here, we may assume that the different types of observations (data) are conditionally independent given the latent processes such that Stage 1 can be written as:

$$\begin{gathered} p(\text{data} \mid \text{latent process}, \ldots) = \\ p(\text{carbon stocks data} \mid \text{latent catbon stocks}, \ldots) \times \\ p(\text{net ecosystem exchange data} \mid \text{latent net ecosystem exchange}, \ldots) \times \\ \ldots \times \\ p(\text{photosynthesis data} \mid \text{latent photosynthesis}, \ldots), \end{gathered} \tag{15}$$

where "…" refers to "covariance parameters." Likewise, the explicit linkage of processes operating at different scales is achieved in Stage 2 (Eq. 12.2), and we can also use the conditional specification ("modeling locally") to decompose the latent process model in Eq. 12.2 into a hierarchy of conditional latent process models, for example:

$$\begin{gathered} p(\text{latent process} \mid \ldots) = \\ p(\text{laent carbon stocks} \mid \text{latent soil respiration, latent net ecosystems} \\ \text{exchange, latent initial stocks}, \ldots) \times \\ p(\text{latent net ecosystem exchange} \mid \text{latent soil respiration, latent} \\ \text{photosynthesis}, \ldots) \times \\ \ldots \times \\ p(\text{latent photosynthesis} \mid \ldots), \end{gathered} \tag{16}$$

where, here, "…" is short-hand for "process model(s), process parameters, covariates." Note, the latent processes depend on ecological process models and process model parameters, in addition to other covariance parameters associated with process error. The final product is a full probability model from which we get the posterior distribution of all unknown quantities, including latent processes and process parameters, allowing inference about mechanisms operating across different scales and governing whole-system behavior.

5.2 Implementation Issues

Many hierarchical models may be "fit" using specialized software packages as mentioned in Section 3.3. For Bayesian computations, we recommend using WinBUGS when possible because the language is relatively transparent and development time is usually short enough to justify the longer run-times typical of a high-level, specialized language. One of the advantages of WinBUGS is that it does not require specification of the details of sampling from the posterior. However, the complexity of the problem may not be amenable to implementation in WinBUGS (e.g. the simulation models discussed in Section 4.2 cannot be accommodated by WinBUGS) or run-time may be prohibitive and other options may be more suitable. For example, R and MATLAB (The MathWorks) are less specialized, but still convenient enough to keep production time reasonable. If run-time is still prohibitive, then a language like C, C++, or Fortran

may be worth the extra development time usually required of these lower-level, but generally faster languages. Often a model may be coded in a higher-level language first, followed by a faster running lower-level version. This dual approach can provide results with a relatively small initial development investment and can serve as a means to check code and results between the two implementations.

There are other important issues that we have avoided thus far but deserve some discussion. In particular, we should mention something about convergence and model checking. Convergence refers to iterative algorithms used to sample from the posterior distribution. Many of these algorithms do not immediately obtain parameter values from the posterior, requiring some number – often many thousands – of iterations before samples can be reasonably assumed to come from the posterior distribution. Visual and other standard diagnostics are available to assess convergence (e.g. Cowles and Carlin 1996; Brooks and Gelman 1998). Sometimes, apparent convergence issues may be due to nonidentifiability problems that may be resolved by re-parameterization or by the introduction of more informative priors. Also, the efficiency with which the posterior parameter space is explored does vary. One solution to a poorly efficient sampling algorithm is simply to let the sampling algorithm continue for a large number of iterations – which may take days, weeks, or months! When run-time is of the essence, substantial energy and experience may be required to devise more efficient sampling strategies (see Gamerman and Lopes 2006).

In principle, once the probability model is specified, inference is fully prescribed via Bayes theorem. This is the strength of Bayesian inference. But this is completely model dependent. So, how can we be sure that we have a "good" model? Of course, "model" refers to any component of a hierarchical model including the likelihood, prior, and any intervening hierarchical structures, including, for example, the process model. Addressing this question requires "model checking" tools, including, for example, evaluating the sensitivity of the posterior to the prior specification, constructing plots of observed vs predicted (e.g. $\bar{A}net_{d,\,t(d),j}$ vs $\bar{A}net_{d,t(d),j}$ in Section 4.1) to evaluate model goodness-of-fit (e.g. Ogle et al. 2006), comparing posterior predictive abilities of competing models (e.g. using DIC; Spiegelhalter et al. 2002), and evaluating parameter identifiability by comparing prior vs posterior distributions (Braswell et al. 2005, Hong et al. 2005). Many standard model-checking methods are described by Gelman et al. (2004). Finally, we mention the use of synthetic data generated from models with known parameter values, which can be compared with the posterior distribution obtained by subjecting the synthetic data to the hierarchical Bayesian modeling framework (see e.g. Ogle et al. 2004, Braswell et al. 2005).

5.3 Thoughts on Statistical Education and Practice

In his commentary on the Bayes/frequentist "roadmap" for the future of statistics as a discipline, Little (2006) concludes with three primary suggestions on future directions for statistics education and practice. Although aimed at statisticians, we believe Little's suggestions are more broadly relevant to include statistics as taught

to and practiced by ecologists, including the hierarchical framework emphasized in this paper.

1. Given the current prominence of Bayesian statistics in science, Bayesian statistics must be taught, beginning at the undergraduate level. Our experience is that the vast majority of statisticians have little to no formal training or professional interest in Bayesian statistics and, consequently, have been slow to incorporate Bayesian methods into the statistics curriculum. Ecologists can help to change this by encouraging their statistician colleagues to teach Bayesian statistical methods.
2. More emphasis should be given to statistical modeling over the methods. We suggest that this include instruction on development of graphical models and their associated probability models, something we only briefly mention in Section 3.1 (for more discussion on graphical and hierarchical modeling, see Clark and Gelfand 2006). We need to provide students and researchers with the tools to allow complex problems to be solved rather instructing them on how to design a study to fit a relatively restrictive analysis framework.
3. More attention should be devoted to assessing model fit. Model-checking typically involves comparisons of predictions to observed data, i.e. to empirical evidence, and in the case of a Bayesian model, includes comparison of posterior quantities to our current knowledge about the parameters. In as much as the models that we discuss here serve as scientific hypotheses, model-checking is then vital to their use in the inherently empirical enterprise of the scientific method. We believe this to be especially important for the complex models discussed in this paper.

Acknowledgements This material is based on work supported by the National Science Foundation (under a Biological Informatics Postdoctoral Fellowship awarded to K.O. in 2003 and Biological Informatics Starter Grant 0630119 awarded to K.O. in 2006) and the Department of Energy National Institute for Climatic Change Research (grant awarded to K.O. in 2006). The authors would like to thank Kimberly Garvie for conducting the literature searches associated with Table 1 and Fig. 1.

References

Aber JD, Federer CA (1992) A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. Oecologia 92:463–474

Aber JD, Reich PB, Goulden ML (1996) Extrapolating leaf CO_2 exchange to the canopy: A generalized model of forest photosynthesis compared with measurements by eddy correlation. Oecologia 106:257–265

Barbour MG, MacMahon JA, Bamberg SA, Ludwig JA (1977) The structure and distribution of *Larrea* communities. In: Mabry TJ, Hunziker JH, Difeo DR (eds) Creosote bush: biology and chemistry of *Larrea* in New World deserts. (US/IBP synthesis series 6) Halsted, New York, pp 227–251

Barnard GA (1958) Studies in the history of probability and statistics. IX. Thomas Bayes's essay towards solving a problem in the doctrine of changes. Biometrika 45:293–295

Berliner LM (1996) Hierarchical Bayesian time series models. In: Hanson KM, Silver RN (eds) Maximum entropy and Bayesian methods. Kluwer, Dordrecht, pp 15–22

Braswell BH, Sacks WJ, Linder E, Schimel DS (2005) Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Global Change Biol 11:335–355

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88:9–25

Brooks SP (1998) Markov chain Monte Carlo method and its application. J R Stat Soc D 47:69–100

Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7:434–455

Canadell JG, Mooney HA, Baldocchi DD, Berry JA, Ehleringer JR, Field CB, Gower ST, Hollinger DY, Hunt JE, Jackson RB, Running SW, Shaver GR, Steffen W, Trumbore SE, Valentini R, Bond BY (2000) Carbon metabolism of the terrestrial biosphere: a multitechnique approach for improved understanding. Ecosystems 3:115–130

Cappe O, Robert CP (2000) Markov chain Monte Carlo: 10 years and still running! J Am Stat Assoc 95:1282–1286

Carpenter SR, Kitchell JF, Cottingham KL, Schindler DE, Christensen DL, Post DM, Voichick N (1996) Chlorophyll variability, nutrient input, and grazing: evidence from whole-lake experiments. Ecology 77:725–735

Casella G, George EI (1992) Explaining the Gibbs sampler. Am Stat 46:167–174

Chib S, Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. Am Stat 49:327–335

Clark JS (2003) Uncertainty and variability in demography and population growth: a hierarchical approach. Ecology 84:1370–1381

Clark JS (2005) Why environmental scientists are becoming Bayesians. Ecol Lett 8:2–14

Clark JS, Gelfand AE (2006) A future for models and data in environmental science. Trends Evol Ecol 21:375–380

Clark JS, Lewis M, McLachlan JS, Hiller IS, Lambers J (2003a) Estimating population spread: what can we forecast and how well? Ecology 84:1979–1988

Clark JS, Mohan J, Dietze M, Ibanez I (2003b) Coexistence: how to identify trophic trade-offs. Ecology 84:17–31

Clark JS, Ferraz G, Oguge N, Hays H, DiCostanzo J (2005) Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. Ecology 86:2232–2244

Cottingham KL, Schindler DE (2000) Effects of grazer community structure on phytoplankton response to nutrient pulses. Ecology 81:183–200

Cowles MK, Carlin BP (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. J Am Stat Assoc 91:883–904

Crome FHJ, Thomas MR, Moore LA (1996) A novel Bayesian approach to assessing impacts of rain forest logging. Ecol Appl 6:1104–1123

Dargaville R, McGuire AD, Rayner P (2002) Estimates of large-scale fluxes in high latitudes from terrestrial biosphere models and an inversion of atmospheric CO_2 measurements. Clim Change 55:273–285

Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data. Chapman and Hall/CRC, Boca Raton

Diggle PJ, Heagerty P, Liange KY, Zeger SL (2002) Analysis of longitudinal data. Oxford University Press, New York

Dixon PM, Ellison AM, Gotelli NJ (2005) Improving the precision of estimates of the frequency of rare events. Ecology 86:1114–1123

Ellison AM (1996) An introduction to Bayesian inference for ecological research and environmental decision-making. Ecol Appl 6:1036–1046

Ellison AM (2004) Bayesian inference in ecology. Ecol Lett 7:509–520

Enting IG, Trudinger CM, Francey RJ (1995) A synthesis inversion of the concentration and $\delta^{13}C$ of atmospheric CO_2. Tellus B Chem Phys Meteorol 47:35–52

Etienne RS, Olff H (2005) Confronting different models of community structure to species-abundance data: a Bayesian model comparison. Ecol Lett 8:493–504

Farquhar GD, Caemmerer S von (1982) Modelling of photosynthetic response to environmental conditions. In: Lange OL, Nobel PS, Osmond CB, Ziegler H (eds) Water relations and carbon assimilation. (Encyclopedia of plant physiology. Physiological plant ecology II) Springer, Berlin Heidelberg New York, pp 549–587

Farquhar GD, Sharkey TD (1982) Stomatal conductance and photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 33:317–345

Farquhar GD, Caemmerer S von, Berry JA (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta 149:78–90

Gamerman D, Lopes HF (2006) Markov chain Monte Carlo. Chapman and Hall/CRC, Boca Raton

Gelfand AE, Hills SE, Racinepoon A, Smith AFM (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. J Am Stat Assoc 85:972–985

Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis. Chapman and Hall/CRC, Boca Raton

Gertner GZ, Fang SF, Skovsgaard JP (1999) A Bayesian approach for estimating the parameters of a forest process model based on long-term growth data. Ecol Model 119:249–265

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. Chapman and Hall/CRC, Boca Raton

Goldstein H (2002) Mulitlevel statistical models. Hodder Arnold, London

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Hong B, Stawrman RI, Swaney DP, Weinstein DA (2005) Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed six. Water Resour Res 41. DOI 10.1029/2004WR003551

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314

Katul GG, Ellsworth DS, Lai CT (2000) Modelling assimilation and intercellular CO_2 from measured conductance: a synthesis of approaches. Plant Cell Environ 23:1313–1328

Kitching RL (1983) Systems ecology: an introduction to ecological modelling. University of Queensland Press, St. Lucia

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963–974

Lambers H, Chapin FS, Pons TL (1998) Plant physiological ecology. Springer, Berlin Heidelberg New York

Larget B, Simon DL (1999) Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol Biol Evol 16:750–759

Law RM, Rayner PJ, Wang YP (2004) Inversion of diurnally varying synthetic CO_2: network optimization for an Australian test case. Global Biogeochem Cycles 18:GB1044. DOI 1010.1029/2003GB002136

Levin SA (1992) The problem of pattern and scale in ecology. Ecology 73:1943–1967

LI-COR (2004) Using the LI–6400: Portable Photosynthesis System, ver 5. LI-COR Biosciences, Lincoln

Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. J R Stat Soc B Stat Method 34:1–41

Lindsey JK (1993) Models for repeated measurements. Oxford University Press, New York

Lindstrom MJ, Bates DM (1990) Nonlinear mixed effects models for repeated measures data. Biometrics 46:673–687

Little RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS system for mixed models. SAS Institute, Cary

Little RJ (2006) Calibrated Bayes: a Bayes/frequentist roadmap. Am Stat 60:213–223

Lunn DJ, Thomas A, Best N, Spiegelhalter DJ (2000) WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10:325–337

Malakoff D (1999) Bayes offers a 'new' way to make sense of numbers. Science 286:1460–1464

Mau B, Newton MA, Larget B (1999) Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55:1–12

McKown AD, Moncalvo JM, Dengler NG (2005) Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. Am J Bot 92:1911–1928

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Miller RE, McDonald JA, Manos PS (2004) Systematics of *Ipomoea* subgenus *quamoclit* (Convolvulaceae) based on its sequence data and a Bayesian phylogenetic analysis. Am J Bot 91:1208–1218

Nobel PS (1999) Physicochemical and environmental plant physiology. Academic, San Diego

O'Hara RB, Arjas E, Toivonen H, Hanski I (2002) Bayesian analysis of metapopulation data. Ecology 83:2408–2415

Oechel WC, Odening WR, Strain BR (1972) Photosynthetic rates of a desert shrub, *Larrea divaricata* Cav., under field conditions. Photosynthetica 6:183–188

Ogee J, Peylin P, Cuntz M, Bariac T, Brunet Y, Berbigier P, Richard P, Ciais P (2004) Partitioning net ecosystem carbon exchange into net assimilation and respiration with canopy-scale isotopic measurements: an error propagation analysis with $^{13}CO_2$ and $CO^{18}O$ dGB2019ata. Global Biogeochem Cycles 18:GB2019. DOI 10.1029/2003GB002166

Ogle K, Reynolds JF (2002) Desert dogma revisited: coupling of stomatal conductance and photosynthesis in the desert shrub, *Larrea tridentata*. Plant Cell Environ 25:909–921

Ogle K, Wolpert RL, Reynolds JF (2004) Reconstructing plant root area and water uptake profiles. Ecology 85:1967–1978

Ogle K, Uriarte M, Thompson J, Johnstone J, Jones A, Lin Y, McIntire EJB, Zimmerman JK (2006) Implications of vulnerability to hurricane damage for long-term survival of tropical tree species: a Bayesian hierarchical analysis. In: Clark JS, Gelfand AE (eds) Hierarchical modelling for the environmental sciences: statistical methods and applications. Oxford University Press, New York, pp 98–117

Oren R, Sperry JS, Katul GG, Pataki DE, Ewers BE, Phillips N, Schafer KVR (1999) Survey and synthesis of intra- and interspecific variation in stomatal sensitivity to vapour pressure deficit. Plant Cell Environ 22:1515–1526

Ovaskainen O, Laine AL (2006) Inferring evolutionary signals from ecological data in a plant–pathogen metapopulation. Ecology 87:880–891

Parton WJ, Schimel DS, Cole CV, Ojima DS (1987) Analysis of factors controlling soil organic matter levels in Great Plains grasslands. Soil Sci Soc Am J 51:1173–1179

Pearcy RW, Schulze ED, Zimmermann R (1989) Measurement of transpiration and leaf conductance. In: Pearcy RW, Ehleringer J, Mooney HA, Rundel PW (eds) Plant physiological ecology. Chapman and Hall, London, pp 137–160

Peek MS, Russek-Cohen E, Wait DA, Forseth IN (2002) Physiological response curve analysis using nonlinear mixed models. Oecologia 132:175–180

Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, Berlin Heidelberg New York

Pinheiro JC, Bates DM, DebRoy S, Sarkar D (2006) Linear and non-linear mixed effects models: R package version 3.1–73. R Foundation for Statistical Computing, Vienna

R Development Core Team (2006) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Radtke PJ, Burk TE, Bolstad PV (2002) Bayesian melding of a forest ecosystem model with correlated inputs. For Sci 48:701–711

Rains MC, Mount JE, Larsen EW (2004) Simulated changes in shallow groundwater and vegetation distributions under different reservoir operations scenarios. Ecol Appl 14:192–207

Rastetter EB, Aber JD, Peters DPC, Ojima DS, Burke IC (2003) Using mechanistic models to scale ecological processes across space and time. Bioscience 53:68–76

Robert CP, Casella G (2004) Monte Carlo statistical methods. Springer, Berlin Heidelberg New York

SAS Institute (2001) The MIXED procedure and %NLINMIX macro. SAS Institute, Cary. Available at http://www.sas.com/statistics/doc.html)

SAS Institute (2006) The GLIMMIX procedure, June 2006. SAS Institute, Cary. Available at http://www.sas.com/statistics/doc.html)

Schulze ED (2000) Carbon and nitrogen cycling in european forest ecosystems. Springer, Berlin Heidelberg New York

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York

Smith AFM, Gelfand AE (1992) Bayesian statistics without tears: a sampling-resampling perspective. Am Stat 46:84–88

Spiegelhalter DJ, Best NG, Carlin BR, Linde A van der (2002) Bayesian measures of model complexity and fit. J R Stat Soc B Stat Method 64:583–616

Stefanski LA (2000) Measurement error models. J Am Stat Assoc 95:1353–1358

Van Oijen M, Rougier J, Smith R (2005) Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiol 25:915–927

Wikle CK (2003a) Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84:1382–1394

Wikle CK (2003b) Hierarchical models in environmental science. Int Stat Rev 71:181–199

Wikle CK, Berliner LM, Cressie N (1998) Hierarchical Bayesian space-time models. Environ Ecol Stat 5:117–154

Wikle CK, Milliff RF, Nychka D, Berliner LM (2001) Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds. J Am Stat Assoc 96:382–397

Xu T, White L, Hui DF, Luo YQ (2006) Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. Global Biogeochem Cycles 20:GB2007. DOI 10.1029/2005GB002468

Yang ZH, Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol Biol Evol 14:717–724