2009

# Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling

Noel Cressie
*Ohio State University*, ncressie@uow.edu.au

Catherine Calder
*Ohio State University*

James Clark
*Duke University*

Jay Ver Hoef
*Alaska Fisheries Science Center*

Christopher Wikle
*University Of Missouri-Columbia*

# Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling


## Abstract

Analyses of ecological data should account for the uncertainty in the process(es) that generated the data. However, accounting for these uncertainties is a difficult task, since ecology is known for its complexity. Measurement and/or process errors are often the only sources of uncertainty modeled when addressing complex ecological problems, yet analyses should also account for uncertainty in sampling design, in model specification, in parameters governing the specified model, and in initial and boundary conditions. Only then can we be confident in the scientific inferences and forecasts made from an analysis. Probability and statistics provide a framework that accounts for multiple sources of uncertainty. Given the complexities of ecological studies, the hierarchical statistical model is an invaluable tool. This approach is not new in ecology, and there are many examples (both Bayesian and non-Bayesian) in the literature illustrating the benefits of this approach. In this article, we provide a baseline for concepts, notation, and methods, from which discussion on hierarchical statistical modeling in ecology can proceed. We have also planted some seeds for discussion and tried to show where the practical difficulties lie. Our thesis is that hierarchical statistical modeling is a powerful way of approaching ecological analysis in the presence of inevitable but quantifiable uncertainties, even if practical issues sometimes require pragmatic compromises.



## Keywords

statistical, strengths, modeling, limitations, hierarchical, accounting, uncertainty, ecological, analysis

## Disciplines

Physical Sciences and Mathematics

## Publication Details

Cressie, N., Calder, C., Clark, J., Ver Hoef, J. & Wikle, C. (2009). Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. Ecological Applications, 19 (3), 553-570. Copyright by the Ecological Society of America.

*Ecological Applications*, 19(3), 2009, pp. 553–570


# Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling

Noel Cressie,[1,5] Catherine A. Calder,[1] James S. Clark,[2] Jay M. Ver Hoef,[3] and Christopher K. Wikle[4]

[1]*Program in Spatial Statistics and Environmental Statistics, Department of Statistics, Ohio State University, Columbus, Ohio 43210 USA*
[2]*Nicholas School of the Environment, Duke University, Durham, North Carolina 27708 USA*
[3]*NOAA National Marine Mammal Laboratory, NMFS Alaska Fisheries Science Center, Fairbanks, Alaska 99775-7345 USA*
[4]*Department of Statistics, University of Missouri, Columbia, Missouri 65201 USA*

*Abstract.* Analyses of ecological data should account for the uncertainty in the process(es) that generated the data. However, accounting for these uncertainties is a difficult task, since ecology is known for its complexity. Measurement and/or process errors are often the only sources of uncertainty modeled when addressing complex ecological problems, yet analyses should also account for uncertainty in sampling design, in model specification, in parameters governing the specified model, and in initial and boundary conditions. Only then can we be confident in the scientific inferences and forecasts made from an analysis. Probability and statistics provide a framework that accounts for multiple sources of uncertainty. Given the complexities of ecological studies, the hierarchical statistical model is an invaluable tool. This approach is not new in ecology, and there are many examples (both Bayesian and non-Bayesian) in the literature illustrating the benefits of this approach. In this article, we provide a baseline for concepts, notation, and methods, from which discussion on hierarchical statistical modeling in ecology can proceed. We have also planted some seeds for discussion and tried to show where the practical difficulties lie. Our thesis is that hierarchical statistical modeling is a powerful way of approaching ecological analysis in the presence of inevitable but quantifiable uncertainties, even if practical issues sometimes require pragmatic compromises.

*Key words: Bayesian modeling; data model; design; empirical Bayes; harbor seals; MCMC; prior; process model; spatial process; spatiotemporal process.*

## Introduction

The field of ecology is becoming increasingly aware of the importance of accurately accounting for multiple sources of uncertainty when modeling ecological phenomena and making inferences. This development is motivated in part by the desire to provide an accurate picture of the state of knowledge of ecosystems and to be able to assess the quality of predictions of local and global change (Hilborn and Mangel 1997, Daszak et al. 2000, Clark et al. 2001, Beckage and Platt 2003, Clark 2005, Ibáñez et al. 2006, Sacks et al. 2007). However, accounting for various sources of uncertainty is by no means a simple task.

Ecological data are almost always observed incompletely with large and unknown amounts of measurement error or data uncertainty, and often the expense of data collection prohibits collecting as much data as might be desirable. How much and where to sample are important design questions (e.g., Stevens and Olsen 2004). In addition, most ecological phenomena of interest can only be studied by combining various sources of data; aligning these data properly presents interesting statistical challenges. While data play a large role in most ecological analyses, incorporating scientific knowledge through substantive modeling of ecological processes is essential. Often such process modeling is based on competing scientific theories and simplifications of reality. This results in an additional source of uncertainty, termed model or process uncertainty. Furthermore, substantive models should acknowledge parameter uncertainty. Parameter uncertainty can be handled either by estimating the unknown parameters (empirical-Bayesian analysis) or by expressing that uncertainty via a prior probability distribution (Bayesian analysis); see, for example, Ver Hoef (1996), Carlin and Louis (2000), and Gelman and Hill (2006), where the two approaches are presented. An empirical-Bayesian analysis looks for plug-in estimates and may avoid more realistic and flexible specifications that can include variation over space and time. The Bayesian analysis can use such variation to help with the choice of the prior distribution.

Manuscript received 4 May 2007; revised 14 September 2007; accepted 1 November 2007; final version received 5 December 2007. Corresponding Editor: J. A. Powell. For reprints of this Forum, see footnote 1, p. 551.
[5] E-mail: ncressie@stat.asu.edu

In April 2006, a workshop on "Uncertainty in Ecological Analysis" was held at the Mathematical Biosciences Institute (MBI), The Ohio State University. The workshop organizers, who are the authors of this paper, believe that the workshop demonstrated how hierarchical statistical modeling is a powerful approach for dealing with uncertainty in ecology. As will be exposited in the sections that follow, it is a statistical methodology for handling complex (ecological) problems by building a hierarchy of statistical models: Broadly, the first level of the hierarchy is a data model, the second is a process model, and the third (optional) level is a parameter model. It also provides a framework for how a team of scientists might work together, and it partitions variability/uncertainty in a way that can suggest study designs where uncertainty can be controlled. Both the Bayesian and non-Bayesian (i.e., empirical-Bayesian) versions of the hierarchical statistical model are presented, reflecting the diversity of workshop participants' views.

This paper does not introduce new methodology, but it is meant to reach, and engender discussion from, readers of this journal. To illustrate the themes of the article, we use an ecological study of harbor seals and their abundance at haul-out sites in Prince William Sound, Alaska, as reported by Ver Hoef and Frost (2003). Harbor seals haul out to rest (da Silva and Terhune 1988), molt (Boily 1995), and escape predators (Watts 1992), among other reasons. They are monitored for trend (e.g., Ver Hoef and Frost 2003) and abundance (e.g., Boveng et al. 2003) in northern latitudes around the world, and they are of particular interest in this study because of potential long-term impacts from the *Exxon Valdez* oil spill in 1989.

The organization of the paper is as follows. *Modeling in the presence of uncertainty* addresses the general notion of modeling in the presence of uncertainty and features hierarchical statistical modeling. *Design for data collection in ecological studies* addresses how hierarchical modeling can contribute to experimental design and data collection. *Statistical inference in ecological analyses* discusses statistical inference in ecological analyses, including treatment of computational aspects and model choice. The paper finishes with a discussion of challenges for hierarchical statistical modeling, presented in *Challenges for hierarchical statistical modeling*.

## Modeling in the Presence of Uncertainty

Why does one build models? Is it simply to organize information? Does the model depend on the goal, whether that is estimation, prediction, forecasting, explanation, or simplification? Ultimately, modeling is about the synthesis of information, whether that comes from observations, or from the collective wisdom of a team of scientists, or more broadly from diverse corners of the relevant literature. We present a basic framework here that accommodates the synthesis of information in a coherent manner. The key to this framework is thinking conditionally, something we argue comes naturally to scientists. For example, we observe $D$ conditional upon $E$, $E$ happens conditional upon $P$, and so forth. Not only is this true in an observational context, but it is appropriate when we think about processes as well. Think of $E$ as an ecological process; then $E$ might behave in one way conditioned on one set of environmental conditions and in another way under a different set of conditions. The conditional behavior is endemic to the process's behavior, regardless of whether observations are taken or not. Formally, we can link such thinking within the context of conditional probabilities.

Let $E$ denote our ecological process of interest. We also observe some data that may, in some sense, be relevant to this process; call these data $D$. A traditional approach that has often been considered is simply to fit a curve to different parts of the data $D$ and to interpret the parameters of that curve in some scientific context. For example, in studies of population dynamics, population counts are often modeled using standard nonlinear growth curves, such as the Ricker or the Gompertz curves (e.g., Wright 1926, Medawar 1940, Ricker 1954, Bjornstad and Grenfell 2001) that are functions of ecologically interpretable parameters. The goal is not to find an exact fit to the data but, rather, to find parameter values so that the curve "best" fits the data. To find these best values for the unknown parameters, it is typically assumed that the data $D$ are generated according to the growth curve up to additive error. However, such considerations do not necessarily make a distinction between the errors of observations (e.g., measurement error) and the errors in modeling the process (e.g., error due to model misspecification). For example, the model may not account for the true underlying process occurring on a much smaller scale (in space and/or time) than that of the observations, which themselves have error due to the particular measuring device. Conceptually, it is important to separate out these different sources of uncertainty.

An alternative to this "curve-fitting" approach is formal statistical modeling. Here, we may wish to specify a probability distribution for $D$ that depends on some parameters, say $\Theta$. Thus, we might think about estimating the parameters of this distribution, $[D \mid \Theta]$. That is, can we find an estimate of the parameters $\Theta$ that maximizes the likelihood (defined by the distributional assumption made about $D$) of observing the data $D$? Note that we are using the brackets "[ ]" to refer to a distribution and the vertical bar "|" is read as "conditioned upon." So, $[A \mid B]$ would be read as "the distribution of $A$ conditioned upon $B$," or "the distribution of $A$ given $B$." In this framework, the parameters $\Theta$ are assumed to summarize the ecological process $E$ appropriately, and it is assumed that there is no uncertainty in this summarization. Although greatly simplified, this is the idea behind much of classical statistical inference in science. In this setting, the observations on the process of interest are directly

modeled without explicit reference to a statistical model for the process $E$. The focus here is on a data model; it is assumed that the uncertainty lies with the data and is due to sampling and measurement.

Alternatively, we may sometimes be interested in developing process models for $E$ directly. Often these models are deterministic. However, our knowledge of $E$ is always limited in some fashion, which suggests that randomness should also play a role in process modeling. Indeed, such thinking has long been crucial to ecological modeling, where it has been incorporated into the science (e.g., in the fire-simulation models of Catchpole et al. [1989] or the forest-simulation models of Botkin et al. [1972]). These random processes can, in turn, be characterized by distributions that have associated parameters, say $P_E$. That is, we know the distribution of $[E \mid P_E]$. If we do not distinguish between observations (which are called $D$ in this paper) and their true values ($E$), then the usual likelihood analysis ignores uncertainties about the relationship between $E$ and $D$, resulting in incorrect statistical inferences.

In hierarchical statistical modeling, both the randomness in the data and in the process are acknowledged. This is achieved by specifying a model for $D$ through a series of conditional distributions. Instead of specifying $[D \mid \Theta]$ directly, we decompose $\Theta$ into the ecological process $E$ and a set of unknown parameters $P_D$ that describe uncertainty in the relationship between $D$ and $E$ (e.g., Calder et al. 2003, Wikle 2003*a*). Then, after accounting for the uncertainty in $E$, we have two conditional distributions, $[D \mid E, P_D]$ and $[E \mid P_E]$, that together comprise a hierarchical statistical model. We now illustrate this conditional-probabilistic approach to separating out the sources of uncertainty, with a case study of harbor seals in Prince William Sound, Alaska.

The introduction given above to hierarchical statistical modeling is quite general, motivated by generic problems of curve fitting or population growth in a random environment. For the rest of the paper we shall use a specific ecological study to illustrate the strengths and limitations of hierarchical statistical modeling. Consider the harbor seals censuses from Ver Hoef and Frost (2003) in Prince William Sound in Alaska, which were collected as part of routine monitoring following the *Exxon Valdez* oil spill in 1989 (see Plate 1). These data were obtained by counting seals at a set of haul-out sites photographed during aerial surveys. Surveys were flown repeatedly for 7–10 days in August or September each year from 1990 to 1999. A map showing the geographical context and the sampling plan (sites were surveyed in either the order of the sites' identifying indices or in the reverse order) is given in Fig. 1.

We now establish notation and a hierarchical statistical model for the harbor seals study. Let $Y_{ij}$ denote the $j$th count for the $i$th site, $i = 1, 2, \ldots, n$. In terms of our previous discussion, all of the counts for all surveys and all sites form a set $D = \{Y_{ij}\}$. A Poisson (e.g., Ver Hoef and Frost 2003) or negative binomial

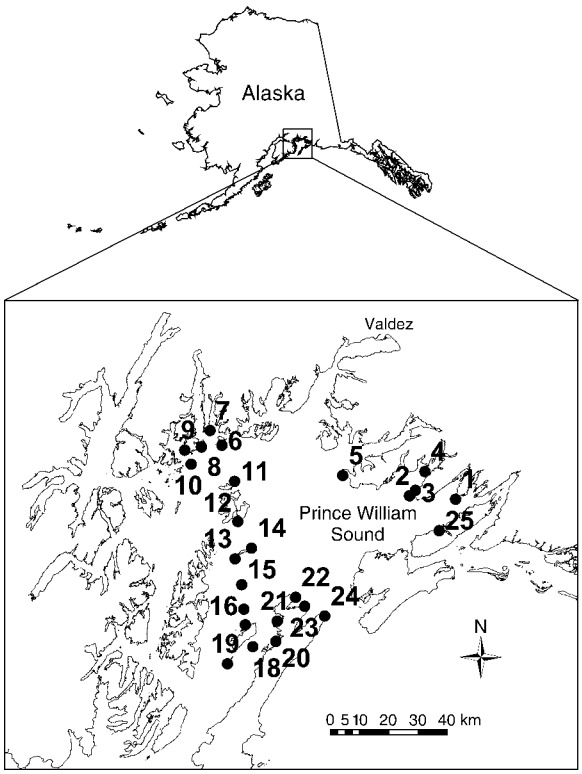


FIG. 1. Map of Prince William Sound, Alaska, USA, with the locations of $n = 25$ harbor seal haul-out sites superimposed. The sites were chosen for aerial surveys to monitor harbor seal trends following the *Exxon Valdez* oil spill in 1989. (Reprinted from Ver Hoef and Frost [2003] with permission.)

(e.g., Boveng et al. 2003) is a natural probability distribution to assume for count data. We can imagine that for any given site at any given time, there is a true number of harbor seals that "should" haul out, but we are not able to observe that number. That is, harbor seals are constantly sliding in and out of the water so that at any given time the actual number that hauls out is more or less than expected. Let this expected number be $\lambda_{ij}$, so we might consider $Y_{ij} \sim f(\lambda_{ij}, \kappa)$, where "$\sim$" is read as "is distributed as" and $f$ is the "Poisson" probability distribution or the "negative binomial" probability distribution with mean parameter $\lambda_{ij}$ and variance parameter $\kappa$ (in the case of the negative binomial). Write $E = \{\lambda_{ij}\}$ and $P_D = \{\kappa\}$, and hence the joint distribution of all the data is $[D \mid E, P_D]$, where this data model is conditional on the ecological process $E$ and a parameter $\kappa$ expressing variability in the data.

This model has very little "ecology" in it and it would be very difficult to carry out inference because we have one parameter ($\lambda_{ij}$) per datum ($Y_{ij}$). However, we know more about harbor seals; for example, we know that counts are highest around low tide and midday, and we know that they change seasonally (date, within year) and there may be temporal trend (year). These factors reflect our understanding of harbor seal biology, such as

PLATE 1. (Top) Aerial photo of harbor seals hauled out, and (bottom) a close-up of a harbor seal. Photo credits: (top) NOAA National Marine Mammal Laboratory, and (bottom) Dave Withrow, NOAA National Marine Mammal Laboratory.

substrate availability and physiology, and we would like to include them in the model. If we were omniscient, we would have a perfect deterministic model for the "expected" harbor seal abundance, without error, on every haul-out site for every instant of time. Lacking that, we use a "smooth curve," which is a function of a set of unknown parameters and the available information about the process (i.e., year, date, tide level, and time of day), to describe the ecological process that controls haul-out abundance. Since we acknowledge that this curve does not accurately reflect the complexity of this process, we account for errors from our ecological process model by introducing an additional source of randomness. This can be achieved by letting $\log(\lambda_{ij})$ be distributed according to a normal (Gaussian) distribution with mean $\mu_{ij} \equiv \mu_i(\text{year}_{ij}, \text{date}_{ij}, \text{tide}_{ij}, \text{time}_{ij}; \boldsymbol{\theta}_i)$ and variance $\sigma^2$ for the $i^{\text{th}}$ site, where we assume that $\{\log(\lambda_{ij})\}$ are conditionally independent given $\{\mu_i\}$ and $\sigma^2$. For example, the smooth curve could be modeled as

$$\begin{aligned}\mu_i(\text{year}_{ij}, \text{date}_{ij}, \text{tide}_{ij}, \text{time}_{ij}; \boldsymbol{\theta}_i) &= \theta_{0i} + \theta_{1i}\text{year}_{ij} + \theta_{2i}\text{date}_{ij} + \theta_{3i}(\text{date}_{ij})^2 + \theta_{4i}\text{tide}_{ij} \\ &\quad + \theta_{5i}(\text{tide}_{ij})^2 + \theta_{6i}\text{time}_{ij} + \theta_{7i}(\text{time}_{ij})^2\end{aligned}$$

where $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i}, \theta_{5i}, \theta_{6i}, \theta_{7i})$; $i = 1, \ldots, n$. The parameter $\sigma^2$ could be viewed as describing the random effect $\{\varepsilon_{ij}\}$ in the equivalent process-model formulation:

$$\log(\lambda_{ij}) = \mu_{ij} + \varepsilon_{ij}$$

where $\{\varepsilon_{ij}\}$ are independent and identically distributed normal random variables with $E(\varepsilon_{ij}) = 0$ and $\text{var}(\varepsilon_{ij}) = \sigma^2$. In terms of our previous discussion, recall $E = \{\lambda_{ij}\}$ and write $P_E = \{\boldsymbol{\theta}, \sigma^2\}$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$. Hence, we can obtain the conditional distribution of $E$, $[E \mid P_E]$, conditional on the parameters $P_E$.

In the conditional modeling described above, we have accounted for the uncertainty in the process $E$ by modeling $[E \mid P_E]$. In the data-generation mechanism, we have accounted for the uncertainty by modeling $[D \mid E, P_D]$ (for the harbor seals data, this is $[\{Y_{ij}\} \mid \{\lambda_{ij}\}]$ for a Poisson data model, or $[\{Y_{ij}\} \mid \{\lambda_{ij}\}, \kappa]$ for a negative-binomial data model).

Returning to our general discussion of hierarchical statistical modeling, we can combine the data model, $[D \mid E, P_D]$, with the process model, $[E \mid P_E]$, to obtain the joint uncertainties in the data $D$ and the ecological process $E$, as follows:

$$[D, E|P_D, P_E] = [D|E, P_D][E|P_E] \tag{1}$$

which is a result from probability theory that shows how the joint uncertainties can be expressed hierarchically using conditional probabilities. We note that this is the essence of Gaussian, linear, state-space models for temporal processes, that result in Kalman filtering for $E$ at a current time based on data at current and past times (e.g., Meinhold and Singpurwalla 1983, West and Harrison 1997). In terms of the harbor seals study, Eq. 1 says that we obtain a joint model of the data and the underlying ecological process, $[\{Y_{ij}\}, \{\lambda_{ij}\} \mid \kappa, \boldsymbol{\theta}, \sigma^2]$, now conditional on fewer parameters. Fig. 2A shows this hierarchical model schematically, with conditional dependencies shown by arrows linking the relevant boxes. The first level is the data model, $[D \mid E, P_D] = [\{Y_{ij}\} \mid \{\lambda_{ij}\}, \kappa]$, and the second level is the process model, $[E \mid P_E] = [\{\lambda_{ij}\} \mid \boldsymbol{\theta}, \sigma^2]$. From Eq. 1, the joint model of the data and the underlying ecological process is simply the product of the data model and the process model. (The more complicated hierarchical model shown in Fig. 2B will be discussed later in this section.)

Note that, so far, we simply have a Poisson regression model for each site, and the sites are not assumed related. This hierarchical model is a special case of a generalized linear mixed model, where the ecological process is a linear model with normal errors, and the data model comes from the exponential family of distributions (e.g., binomial, Poisson, gamma, Gaussian, and so forth).

The formula given by Eq. 1 uses a general result from probability theory that we present here for three random variables, $A$, $B$, and $C$. We might try to specify directly the joint probabilistic behavior of these three variables, $[A, B,$

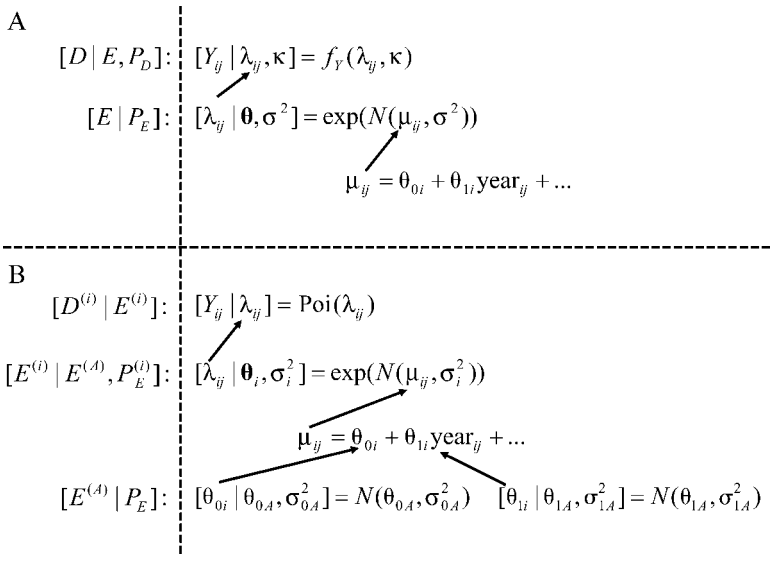


FIG. 2. A schematic of the hierarchical statistical models for the harbor seals study. (A) The data model $[D \mid E, P_D]$ is the top level with the process model $[E \mid P_E]$ at the second level. A Bayesian hierarchical statistical model would have a parameter model at a third level. (B) As in panel A, except the linear model parameters from each site $i$ are given a distribution $[E^{(i)} \mid E^{(A)}, P_E^{(i)}]$. Here, a Bayesian hierarchical statistical model would have a parameter model at the fourth level. All symbols shown in the figure are defined in the text and equations preceding Eq. 1.

$C$]. Equivalently, we could use the following result based on conditional probabilities: $[A, B, C] = [A \mid B, C][B \mid C][C]$. That is, the joint distribution can be factored into a series of three distributions, namely two conditional distributions, $[A \mid B, C]$ and $[B \mid C]$, and one marginal distribution, $[C]$. We refer to such a representation as a hierarchical statistical model; Eq. 1 is such a representation for two random quantities, $D$ and $E$, conditional on parameters $P \equiv \{P_D, P_E\}$. Hierarchical statistical modeling simply means that we decompose the joint distribution into a probabilistically valid series of conditional models: $[D, E \mid P] = [D \mid E, P][E \mid P]$. We note that a conditional-probability decomposition is not unique; for example, we could write equivalently, $[A, B, C] = [B \mid C, A][C \mid A][A]$, and so forth. Part of scientific modeling in the presence of uncertainty is to use a decomposition that reflects the causal mechanisms: If $C$ causes $B$ which causes $A$, then the original decomposition is appropriate. Of course, getting the causal mechanism right means that the results from the hierarchical statistical analysis are scientifically meaningful and interpretable.

We now use this framework to put the hierarchical statistical model (Eq. 1) into a more general context. Berliner (1996) gives the joint distribution of data, process, and parameters using the following decomposition:

$$\begin{aligned}&[\text{data, process, parameters}]\\ &= [\text{data}|\text{process, parameters}] \times [\text{process}|\text{parameters}]\\ &\quad \times [\text{parameters}]\end{aligned}$$

which we have just seen defines a hierarchical statistical model. Note that in this case the parameters are assumed to be random and thus they have a distribution. Although there may be many cases in which one really believes that this is scientifically plausible, it can also simply serve as a mechanism for accounting for uncertainty in our knowledge about the parameters. It is also possible to build a simpler hierarchical model that conditions on the parameters

$$\begin{aligned}&[\text{data, process}|\text{parameters}]\\ &= [\text{data}|\text{process, parameters}] \times [\text{process}|\text{parameters}]\end{aligned}$$

which is Eq. 1. The parameters could then be estimated in a separate inference step (empirical-Bayesian analysis; in the ecology context; see e.g., Ver Hoef 1996) based on [data | parameters]. Lele et al. (2007) show how this can be achieved using a Markov chain Monte Carlo method for the (non-Bayesian) hierarchical statistical model given by Eq. 1.

The analogous model to (1) that incorporates uncertainty in the parameters $P = \{P_D, P_E\}$, is

$$[D, E, P][D|E, P][E|P][P] = [D|E, P_D][E|P_E][P_D, P_E] \quad (2)$$

where we have used the obvious relationships, $[D \mid E, P] = [D \mid E, P_D]$ and $[E \mid P] = [E \mid P_E]$ (e.g., Wikle 2003*a*). The utility of such a decomposition is that it allows us to account formally for uncertainty within each stage, where the stages are linked in a probabilistically consistent fashion, resulting in a Bayesian analysis (e.g., Link et al. 2002).

Both types of analyses, empirical-Bayesian and Bayesian, result in inferences on process values ($E$) and parameter values ($P$). When making inference on $E$, an empirical-Bayesian analysis substitutes in an estimate

$\hat{P}$ of $P$ into whatever summary of $[E \mid D, P]$ was chosen for inference. Typically, the variability of the estimator $\hat{P}$ is not accounted for, leading to credible intervals for values of $E$ that are too liberal. One could try to account for the variability in $\hat{P}$, but formal inference is usually approximate (e.g., Morris 1983*a*, *b*); one commonly used approximation is based on multivariable differential calculus and Taylor-series expansions of estimators (e.g., Prasad and Rao 1990, Cressie 1992, Rao 2003:section 6.2). In contrast, the Bayesian analysis bases its inferences directly on $[E \mid D]$ and, provided it can be computed, the resulting credible intervals are accurate. Comparisons of this sort are made by Carlin and Louis (2000) and Gelman and Hill (2006).

Hierarchical statistical modeling leads to computationally intensive inference, regardless of whether the empirical-Bayesian or the Bayesian analysis is chosen. These are discussed in *Statistical inference in ecological analyses*, including the use of Markov chain Monte Carlo and bootstrap procedures to account for the variability in $\hat{P}$.

The evaluation of a statistical-inference procedure will depend on its purpose. If the goal is to describe the process $E$, then different procedures could be evaluated based on some summary of the variability in $[E \mid D]$, for a Bayesian analysis. However, if the goal is management of the ecological process $E$, then we might want a summary that averages over $D$ as well. With many management decisions to make under like circumstances, we would not want to use an inference procedure that is specific to the one dataset we happened to observe. Hence, inference is often designed to minimize the expected loss, where the loss is $L(E, \delta(D))$, with respect to the estimator $\delta(D)$ of $E$. Importantly, that expected loss involves expectations over both $E$ and $D$. Likewise, for an empirical-Bayesian analysis, expectation of the loss is taken over both $E$ and $D$, but conditional on $P$.

The real power of this approach for complicated problems comes from the fact that each of the component distributions in Eqs. 1 or 2 can be decomposed further, if necessary, and they may be simplified with modeling assumptions. For example, say we are interested in a process $E$ for which we have several different data sets, $D^{(1)}$, $D^{(2)}$, $D^{(3)}$, all of which measure the process $E$ with uncertainty and perhaps at different spatial or temporal scales. Then, it is often possible in such cases to make the following modeling assumption:

$$[D^{(1)}, D^{(2)}, D^{(3)} | E, P_D] = [D^{(1)} | E, P_D^{(1)}][D^{(2)} | E, P_D^{(2)}][D^{(3)} | E, P_D^{(3)}]$$

where $P_D = \{P_D^{(1)}, P_D^{(2)}, P_D^{(3)}\}$. Such an assumption is appropriate in the harbor seals study for repeat counts of the same set of sites on successive days. That is, we might assume that the different datasets are independent, conditional upon the true process. Although such an assumption must be justified, it is often plausible and provides a very convenient approach for synthesizing various types of observations. For the harbor seals study, a further assumption, $P_D^{(1)} = P_D^{(2)} = P_D^{(3)} \equiv P_D$, was made; that is, it was assumed that each successive daily count follows the same data model. The parameters in each of the component distributions can accommodate changes of resolution and alignment, as well as different measurement-error characteristics (e.g., Wikle et al. 1998, 2001, Mugglin et al. 2000, Gelfand et al. 2001, Gotway and Young 2002, Banerjee et al. 2004, Clark et al. 2004, Hille Ris Lambers et al. 2005, Wikle and Berliner 2005, LaDeau and Clark 2006).

Decomposition can also be considered for the probability distribution of the ecological process. For example, consider a process $E$ made up of two subprocesses, $E^{(1)}$ and $E^{(2)}$. We can often make use of conditional modeling in this context as well:

$$[E^{(1)}, E^{(2)} | P_E] = [E^{(1)} | E^{(2)}, P_E][E^{(2)} | P_E].$$

For example, one could build hierarchical models for multiple species to examine species diversity and distribution patterns (Gelfand et al. 2005*a*, *b*, Latimer et al. 2006). In terms of the harbor seals study, the ideas of multiple data sets and multiple processes were used by Ver Hoef and Frost (2003), yielding the following model:

$$[D^{(1)}, \ldots, D^{(n)}, E^{(1)}, \ldots, E^{(n)}, E^{(A)} | P_E] = [D^{(1)} | E^{(1)}] \cdots [D^{(n)} | E^{(n)}][E^{(1)}, \ldots, E^{(n)} | E^{(A)}, P_E][E^{(A)} | P_E].$$

That is, a joint model for the data and the linear-regression parameters was built conditionally. Here, $[D^{(i)} \mid E^{(i)}] = [Y_{ij} \mid \lambda_{ij}]$, is the Poisson-regression formulation at the $i^{\text{th}}$ site, given earlier. Assume

$$[E^{(1)}, \ldots, E^{(n)} | E^{(A)}, P_E] = [\{\lambda_{1j}\} | \boldsymbol{\theta}_1, \sigma_1^2] \cdots [\{\lambda_{nj}\} | \boldsymbol{\theta}_n, \sigma_n^2]$$

where there are now potentially $n$ different variances $\{\sigma_i\}$, one for each site. Furthermore, suppose that all intercepts across sites have a common distribution with parameters associated with the area $A$, all slope parameters across sites have a common distribution, and so forth. Let $E^{(A)}$ denote the set of all these parameters associated with the area $A$. This model is shown in Fig. 2B, which extends the model shown in Fig. 2A. For example, let the intercept mean parameter be normally distributed. This and specifications for the slope parameters and the variance parameters determine $[E^{(A)} \mid P_E]$. This makes clear one of the real strengths of hierarchical modeling. Without it, we might model each site separately but then lack the ability to say anything about all sites within the area, or we might ignore sites and model all data with a single Poisson regression. The hierarchical model allows inference for $E^{(i)}$ at the $i$th site or for $E^{(A)}$ over the area that contains all sites, through their respective posterior distributions.

Fig. 3 shows the posterior density of the individual slopes $\{\theta_{1i}: i = 1, \ldots, n\}$ for year (temporal trend) in the

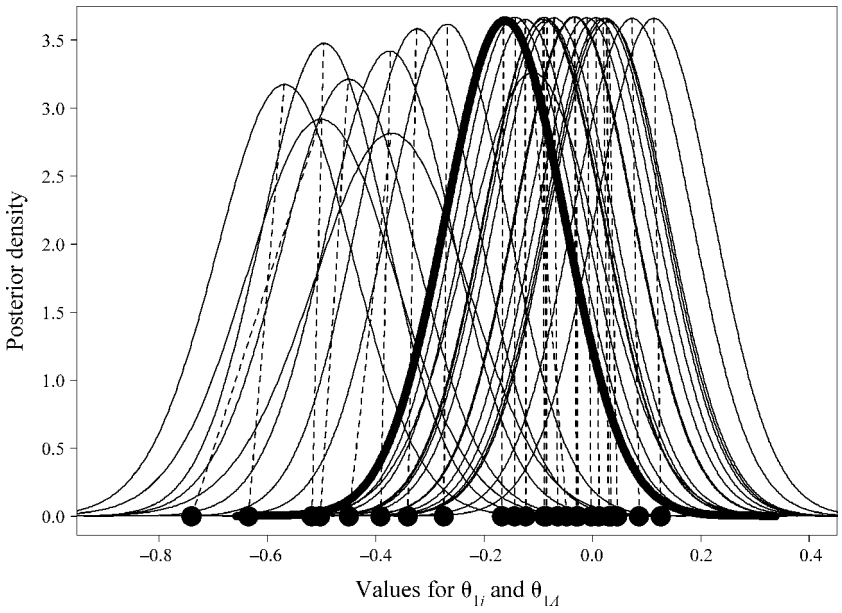


FIG. 3. Posterior densities of slopes $\{\theta_{1i}: i = 1, \ldots, 25\}$, corresponding to the year (trend) covariate, with the posterior density of $\theta_{1A}$ superimposed (heavy solid line). These posterior densities correspond to the model in Fig. 2B. The solid circles along the $x$-axis correspond to the peak densities for the model in Fig. 2A. The dashed lines link the peaks for each site and show how the Bayesian hierarchical model "shrinks" one-at-a-time site estimates toward the global trend estimate.

linear model for $\mu_i$ given above (Fig. 2B), along with the posterior density of the slope $\theta_{1A}$, where recall that $\theta_{11}$, $\ldots$, $\theta_{1n}$ are conditionally distributed with mean $\theta_{1A}$. Consider the mean of each posterior density, and notice the distribution of the individual sites' mean slopes around the areal mean slope. The mean of each posterior density, considering each site separately (Fig. 2A), is shown with solid circles along the $x$-axis of Fig. 3. The one-at-a-time site means are linked by the dashed line to their counterparts under the conditional model with mean $\theta_{1A}$. It is generally true that one-at-a-time estimates will "shrink" towards the global mean under the hierarchical model; the extent of shrinkage depends on the (estimated) variability within site versus among sites. From Fig. 3, the majority of the slopes' means are negative, indicating a decline in harbor seal abundance over years. We also see that the mean of $\theta_{1,A}$, the regression coefficient for the area, is negative.

It is often possible to simplify the joint interaction of one component of the process with another, by using a conditional probabilistic relationship. A well known example of this occurs in time series, where a Markov assumption is made. Specifically, if one has the time series, $E_1$, $E_2$, $\ldots$, $E_{T-1}$, $E_T$, it is often very difficult to specify the joint distribution of the entire series. However, under the first-order Markov assumption, the joint distribution is

$$[E_T, E_{T-1}, \ldots, E_2, E_1] = [E_T|T_{T-1}][E_{T-1}|E_{T-2}] \cdots [E_2|E_1][E_1].$$

The key to these process decompositions is to model the process in stages that are scientifically plausible. In that way, very complicated joint distributions can be modeled by relatively simple conditional distributions. Often, deterministic process models for $E$ (given $P_E$) can be reformulated with stochastic components; for example, Wikle (2003*b*) uses a reaction–diffusion partial differential equation to motivate a stochastic process model for invasive species. The Markov assumption is the basis of the Kalman filter (e.g., West and Harrison 1997); Meinhold and Singpurwalla (1983) demonstrate that the Kalman filter can be derived from a non-Bayesian hierarchical statistical model. When the individual processes in the series $E_1$, $E_2$, $\ldots$, are spatial processes, spatiotemporal hierarchical models can be built that result in a spatiotemporal Kalman filter (Huang and Cressie 1996, Wikle and Cressie 1999).

In this section, we have provided a baseline for the concept of, consistent notation for, and the flexibility of, hierarchical statistical modeling to deal with uncertainty in ecological studies. An important part of the exposition has been to present the role of *both* the ecological process $E$ and the parameters $P$. The observant reader will see that the distinction between $E$ and $P$ is not always precise, but this is not probabilistically important in a hierarchical statistical model. At the last stage of the hierarchy, there will typically be unknown parameters that could be estimated or whose uncertainty could be captured probabilistically by (prior) statistical distributions.

## Design for Data Collection in Ecological Studies

Before moving to discussion of statistical inference for hierarchical models (*Statistical inference in ecological analyses*), we consider the issue of design for data collection from a statistical modeling perspective. This is a topic that deserves more exposure in the ecology literature, where it has hitherto not been connected with hierarchical statistical modeling. Hence, we aim here for a complete presentation, where we highlight the importance of collecting data using sampling designs based on the principles of randomization, stratification, and replication, no matter how the data are analyzed. We show how sampling designs can be incorporated into (hierarchical) statistical models. This should be contrasted with model-assisted, design-based inference (e.g., Särndal et al. 1992), where a model-based estimator is assessed based on distributions implied by survey-sampling probabilities. We believe that design is the medium through which team members with different scientific expertise can communicate with each other, and that its principles are founded on the uncertainties that are present in all scientific studies.

To many people who have taken a course on design in a statistics program, the topic brings back memories of factorial designs, randomized block designs, partially balanced incomplete block designs, latin-square designs, and so forth. These are relatively dry topics (although variants of latin-square designs have surfaced with the popularity of the game of Sudoku, found in most daily newspapers). In fact, design is a much broader concept, and appearances of it can be found throughout a typical Statistics graduate curriculum, particularly in survey-sampling courses (sampling design) and design-of-experiments courses (experimental design).

Every sampling design should start with a specification of the population and the subpopulations under study. For the harbor seals study, the population is the continuous abundance of harbor seals on all haul-out sites for a given study area and time span.

The main principles of design are stratification, randomization, and replication, which are intended to allow the scientist to get closer to the elusive goal of establishing causation. Stratification controls variability by splitting populations into subpopulations that are internally comparable. Strata (or blocks) should be chosen so that important subpopulations are included and so that the subpopulations will span the variability in responses expected. Often strata are chosen according to a combination of (prespecified) factors. Suppose each factor is split up into levels (e.g., two levels, "high" and "low"). A factorial design is where every possible combination of levels appears once, and each combination of levels represents a stratum. By ensuring that the levels of each factor are at least at the "high end" and at the "low end" of possible choices, strata in a factorial design will span the variability in responses expected.

The goal of the harbor seals study was to monitor trend. Hence, it would be best to standardize, as much as possible, the timing of counts each year. As indicated earlier, counts are primarily affected by date, tide, and time of day. To see how stratification would work, the time relative to low tide could be divided into three categories: more than one-half hour before low tide, one-half hour before to one-half hour after low tide, and more than one-half hour after low tide. Likewise, the time of day could be divided into three categories and the date into three categories. Then, one could decrease variability across years by sampling only within a single cross-classified stratum, for example, low tide in the middle of the day in the middle of August. As is often the case, real-world factors do not allow such a design; low tide varies considerably among days and weeks within each year, and using a single aircraft to observe all sites every day does not allow observers to be everywhere at once. The strategy of finding a constant set of environmental conditions is not possible in this example. The alternative is to embrace the environmental variation and create a design that models variations in counts due to environmental conditions. Stratification was based on cross-classified levels of factors, and as much as possible the numbers of samples in strata were balanced. For example, samples from late in the day, more than one-half hour after low tide, and in August, were obtained in about the same number as samples from the stratum with the near-optimal conditions of early morning, low tide, and in early August.

Randomization offers protection against a biased (unintentional or intentional) choice of sampling sites (or of organisms) in the study. If it is at all possible, ecologists should avoid choosing a so-called "representative site." In the harbor seals study, Ver Hoef and Frost (2003) analyzed data from 25 sites in Prince William Sound, Alaska, but economical considerations did not permit sampling these 25 sites randomly from the hundreds of possible sites in the study area. It is very expensive to contract aircraft, and there are only a few possible ways to fly economically from site to site. Thus, to avoid spending a lot of time and money flying back and forth to sites selected at random, or even a spatially clustered set of sites in a random order, the harbor seals study did not use random sampling. As a result, purely design-based statistical inference cannot extend beyond these 25 sites, and how representative they are of a larger region has been criticized by Hoover-Miller et al. (2001).

A sampling design becomes a probability sampling design when there is a known nonzero inclusion probability for every member of the population (Overton 1993). The sampling weights, which should be given as part of the design, are simply one divided by the inclusion probabilities. Without randomization, the harbor seals study clearly does not use a probability sampling design.

Replication is a way to decrease (e.g., measurement, spatial, temporal) variability within strata. The general idea is that an average of responses within a stratum has

variance proportional to one divided by the number of replicates. The larger the number of replicates, the smaller the variance and the more precise the inferences. For the harbor seals study, 7–10 replicate flights per year were flown to decrease variation.

In the harbor seals study, as is the case with most ecological studies, a limiting factor that influences the total number of observations is the amount of money available. Generally, the number of strata that make up the stratification is determined by the team of scientists, whose goal is to avoid confounding of factors and to increase the precision of estimates as much as possible. To illustrate, consider the simple case where there are equal numbers in each stratum. Then, the number of replications per stratum is simply obtained by dividing the total number of observations by the number of strata. But, a small number of replicates typically does not allow us to find a regression parameter significant when it should be, or it results in a hypothesis test that has weak power. There is clearly a tension between appropriate stratification/sufficient replication and cost of the scientific study; statistical design shows how to express this tension mathematically.

A good design will specify, in advance, inference thresholds and determine the number of observations per stratum needed to achieve those thresholds. It creates a rational basis for the inevitable compromise between the cost of the study and the ability to make scientific inferences from incomplete and noisy data (e.g., Cressie [1998] relates cost and inference in an ice-core sampling design for a transect across Antarctica). Equally, a good design will involve random sampling, from which robust, design-based inference is possible. Additionally, a good design will stratify to ensure sampling over a range of levels (values) of factors (covariates).

Sampling design has traditionally eschewed the modeling approach given in *Design for data collection in ecological studies*, but while design-based inference can be robust, model-based inference can be very efficient (e.g., Aldworth and Cressie 1999), where efficiency is characterized by mean squared prediction error. In this section, we show how traditional sampling design can be included in the hierarchical-statistical-modeling approach.

Given that many ecological processes evolve dynamically through space and time, purely spatial designs are typically not as efficient as those that consider spatiotemporal dependence. For the most part, the construction of optimal designs in the spatiotemporal context is similar in principle to the optimal-design problem in the purely spatial case (Le and Zidek 1994, Federov and Nachtsheim 1995, Arbia and Lafratta 1997, Zimmerman 2006). However, as demonstrated in Wikle and Royle (1999, 2005), even more efficient designs in the spatiotemporal context can be obtained by allowing the design to change with time, where they characterize efficiency by mean squared prediction error averaged over the current spatial domain of interest. Although related to the adaptive-sampling approaches for purely spatial designs (e.g., Thompson and Seber 1996, Chao and Thompson 2001), these "dynamic designs" are fundamentally different. They explicitly account for temporal changes of the process caused by the underlying dynamics.

The ecological process $E$ is of interest, and one wishes to design a sampling plan through which knowledge about $E$ can be obtained. For the moment, we make the (unrealistic) assumption that when $E$ or parts of it are observed, they are observed without error. We discard this assumption later in the section.

In the rest of this section, we write

$$E \equiv \{E_s : s \in A\}$$

where $A$ is an index set that describes the complete population of scientific interest; $s$ may or may not index a spatial location. In traditional sampling design, the unknown population is sampled randomly. That is, a probability distribution is put on $\{S\colon S \subset A\}$, the set of all subsets of $A$; then inference on $E$ is based on the stochasticity in the random samples. This is called design-based inference and, in the terminology developed in *Modeling in the presence of uncertainty*, it depends on $[S \mid E]$, where $S$ is a random subset of $A$ that defines the random sample. Because of the practical difficulties mentioned for the harbor seals study, appropriate stratification and random sampling was not carried out. Hence, design-based inference is not possible in the study, and it cannot be used to illustrate the methodology given below.

Suppose that $S = S^{(1)}$ is observed, and hence $E^{(1)} \equiv \{E_s\colon s \in S^{(1)}\}$ is the observed part of $E$. Recall that we are momentarily assuming that there is no measurement error; hence we do not distinguish between $D_s$ (data value) and $E_s$ (process value). In this simple formulation, one wishes to make inference on the unobserved parts of $E$, $\bar{E}^{(1)} \equiv \{E_s\colon s \in A\backslash S^{(1)}\}$, from the observed part $E^{(1)}$, where $A\backslash S^{(1)}$ means the index set of $A$ with those in $S^{(1)}$ removed. Traditional quantities of interest are the population total and the population mean: $T_E \equiv \Sigma_{s\in A} E_s$, and $\mu_E \equiv \Sigma_{s\in A} E_s/\Sigma_{s\in A} 1$, respectively, as well as the population variance: $\sigma_E^2 \equiv \Sigma_{s\in A} (E_s - \mu_E)^2/\Sigma_{s\in A} 1$ (e.g., Cochran 1977).

Estimators of population quantities are functions of $S^{(1)}$ and $E^{(1)}$. For example, consider estimating the total $T_E$, which we can write as, $T_E = \Sigma_{s\in S^{(1)}} E_s + \Sigma_{s\in A\backslash S^{(1)}} E_s$. One popular estimator of $T_E$ is known as the Horvitz-Thompson estimator:

$$\hat{T}_E = \sum_{s\in S^{(1)}} (E_s/p_s)$$

where $p_s$ is the probability that $s$ belongs to $S^{(1)}$; that is,

$$p_s = \sum_{S^{(1)}: s\in S^{(1)}} [S^{(1)}|E].$$

Recall that for a probability sampling design (Overton 1993), $\{p_s\}$ is known. The Horvitz-Thompson estimator is unbiased (e.g., Cochran 1977); that is, the mean of the estimator is equal to the population quantity:

$$\sum_{S^{(1)}} \hat{T}_E \times [S^{(1)}|E] = T_E$$

and its sampling variance can be calculated in a like manner, again based on $[S^{(1)} \mid E]$. Notice that sampling moments (e.g., mean, variance) are calculated with respect to the $S^{(1)}$ that *could have* occurred, through the probability $[S^{(1)} \mid E]$.

In other words, for design-based inference, the source of randomness is an externally imposed probability distribution that refers to which parts of the fixed $E$ will be sampled. A sample is purposive if a particular $S^{(0)}$ is used and hence data on $E^{(0)} \equiv \{E_s: s \in S^{(0)}\}$ is obtained; as mentioned above, ecologists should beware of choosing a "representative site" (i.e., purposive sample) for their study, because statistical inference on $\bar{E}^{(0)} = E \backslash E^{(0)}$, the unsampled part of $E$, requires further assumptions that may be difficult to justify. Furthermore, inferences can be very inefficient (e.g., Aldworth and Cressie 1999, Ver Hoef 2002).

Surprisingly, much of the past sampling-design literature assumes that $E$ can be observed without error (as we did above). From the discussion in *Modeling in the presence of uncertainty*, this is an unrealistic assumption; uncertainty in the measurement process should always be accounted for. Let $D^{(1)} \equiv \{D_s: s \in S^{(1)}\}$ denote the data resulting from observing $E^{(1)}$. Then the joint uncertainties are expressed through

$$[D^{(1)}, S^{(1)}|E, P_D] = [D^{(1)}|S^{(1)}, E, P_D][S^{(1)}|E].$$

A hierarchical-statistical-modeling approach takes the extra step of including uncertainties in the process $E$:

$$[D^{(1)}, S^{(1)}, E|P] = [D^{(1)}|S^{(1)}, E, P_D][S^{(1)}|E][E|P_E]$$

where, recall from Eq. 2, $P = \{P_D, P_E\}$. *Modeling in the presence of uncertainty* has already discussed the importance of accounting for uncertainty in $E$. The joint distribution just above shows how this can be done in conjunction with the uncertainty in sampling.

If parameters $P$ are assumed to be fixed, inference on $E$ is based on the posterior distribution:

$$[E|D^{(1)}, S^{(1)}, P] \propto [D^{(1)}|S^{(1)}, E, P_D][S^{(1)}|E][E|P_E]$$

where the proportionality constant is $[D^{(1)}, S^{(1)} \mid P]$. Now suppose that uncertainty in $P = \{P_D, P_E\}$ is captured by the prior $[P]$; then the appropriate posterior distribution is

$$[E, P|D^{(1)}, S^{(1)}] \propto [D^{(1)}|S^{(1)}, E, P_D][S^{(1)}|E][E|P_E][P]. \quad (3)$$

Many designs have the property that they are ignorable (e.g., Gelman et al. 2004), namely, $[S^{(1)} \mid E] = [S^{(1)}]$, independent of the process $E$. In this case, an extraordinary simplification takes place:

$$[E, P|D^{(1)}, S^{(1)}] \propto [D^{(1)}|S^{(1)}, E, P_D][E|P_E][P].$$

That is, uncertainty from sampling disappears from the inference, yielding the purely model-based inference developed in *Modeling in the presence of uncertainty*. (A similar result for ignorable designs occurs when $P$ is assumed fixed.) Further details on incorporating sampling designs (ignorable and non-ignorable) in a hierarchical statistical model can be found in Gelman et al. (2004: Chapter 7).

It is important to realize that for purely observational studies without randomization, the sample is purposive; statistical inference on $\bar{E}^{(0)}$ is achieved by making ecological modeling assumptions about all of $E$ through $[E \mid P_E]$. Such is the case for the harbor seals study, where recall that there was *no* randomization in the data acquisition. Ver Hoef and Frost (2003) assume that the measurement process $\{D^{(i)}\}$ has Poisson variation with means $\{\lambda_{ij}\}$; then inference is based on $[\{\lambda_{ij}\}, \boldsymbol{\theta}, \sigma^2 \mid D^{(1)}, \ldots, D^{(n)}]$, where there is now no reference to the sampling scheme in the posterior distribution.

In conclusion, this section has presented a unified way to treat sampling design and statistical modeling together, using hierarchical statistical models. Many sampling designs are chosen to be ignorable (such as simple random sampling, systematic sampling; e.g., see Cressie [1993:316–323]), in which case the approach given in *Modeling in the presence of uncertainty* is appropriate. More generally, the posterior distribution (Eq. 3), which involves the sampling probabilities, should be used for inference, but often is not.

## Statistical Inference in Ecological Analyses

In *Modeling in the presence of uncertainty*, we have shown how conditional probabilities can be used to build complex models of ecological processes that account for uncertainty. Even if we agree that this approach is reasonable, there remains the question of how one does inference in this setting. An empirical-Bayesian approach could be taken. That is, consider only the first two stages, $[D \mid E, P_D]$ and $[E \mid P_E]$, and take the parameters $P = \{P_D, P_E\}$ to be fixed, but unknown. Depending on the complexity of the component models in this case, it is often possible to use classical statistical-estimation approaches to obtain estimates of the parameters $P_D$ and $P_E$, and hence to use plug-in predictions for $E$. For example, the usual linear mixed model can be thought of in this context (Christensen 1991). In addition, spatial prediction (kriging) fits into this framework (Cressie 1993: Chapter 3), as do sequential time-series methods such as Kalman filtering. Common approaches for estimation of $P$ include the expectation–maximization (E–M) algorithm, conditional and pseudo-likelihood methods, and estimating equations (Hardin and Hilbe 2003). Although such methods do not explicitly account for the uncertainty in estimating the parameters, that uncertainty can, if deemed necessary, often be accounted for by Taylor-

series approximations (e.g., Rao 2003: section 6.2) and resampling and bootstrap procedures (Efron and Tibshirani 1993). For example, prior to using the Bayesian hierarchical model for harbor seals given in Ver Hoef and Frost (2003), Frost et al. (1999) used a bootstrap approach. They modeled variation in counts with Poisson regression for each site, and then they summed up the model predictions across sites for a standardized set of covariate conditions for each year. Linear regression was then used on these yearly sums to estimate trend. When doing "statistics on statistics," uncertainty from one analysis (parameter estimation) was passed to the next (prediction) through resampling (here a bootstrap).

We turn now to the (fully) Bayesian approach. Here, we focus on discussing how hierarchical models are fitted in the Bayesian paradigm, rather than on comparative inference. For a comparison of Bayesian and likelihood-based methods of inference, see Browne and Draper (2006) and references therein. The Bayesian paradigm has a conceptually holistic approach to inference, where the parameters are also given distributions, and we are interested in the distribution of the process and parameters given the data: $[E, P_E, P_D \mid D]$. Our interest might be in inference on the parameters $P_E$ and $P_D$, or in prediction or forecasting of the process $E$. In each case, we seek distributions of these variables, given the data that were actually observed. We use Bayes' Theorem from probability theory:

$$\begin{aligned}[E, P_E, P_D|D] &= [D|E, P_E, P_D][E, P_E, P_D]/[D] \\ &= [D|E, P_D][E|P_E][P_E, P_D]/[D] \\ &\propto [D|E, P_D][E|P_E][P_E, P_D] \qquad (4)\end{aligned}$$

where we have already named the left-hand side the posterior distribution. The numerator on the right-hand side of (4) is the hierarchical decomposition developed above, and the proportionality constant, $1/[D]$, is the inverse of the marginal distribution of the data.

Recall that the joint distribution of all uncertainties is $[D, E, P_E, P_D] = [D \mid E, P_E, P_D][E, P_E, P_D]$. We could view $[E, P_E, P_D]$ as the prior distribution on all "unknowns" (process and parameters) and the posterior distribution, $[E, P_E, P_D \mid D]$, represents what has been learned about the unknowns from the data $D$. Should new data come along, the posterior distribution can be viewed as an updated prior distribution, and an updated posterior distribution can be computed by considering the expression for the joint distribution of all uncertainties and replacing $D$ with the new data and using the updated prior. This is easily seen to be a coherent way of updating information that arrives sequentially.

For the harbor seals study, if we assume that the measurement error defining $D = \{Y_{ij}\}$ has negative binomial variation, then the measurement-error process has a single parameter $P_D = \{\kappa\}$. The parameters in $P_E$ are mostly linear-regression coefficients of $\log\{\lambda_{ij}\}$, along with one variance parameter $\sigma^2$ for a random effect. Typically, either $\kappa$ or $\sigma^2$ have to have a peaked prior. If it is $\kappa$, the prior $[P_E]$ can be chosen to be very diffuse and noninformative. Finally, inference is based on $[E, P \mid D] = [\{\lambda_{ij}\}, P_D, P_E \mid \{Y_{ij}\}]$, where $P_E = \{\boldsymbol{\theta}, \sigma^2\}$. For example, Fig. 3 shows the distribution of $[\theta_{1i} \mid \{Y_{ij}\}]$; $i = 1, \ldots, n$, where $\{\theta_{1i}\}$ are the regression coefficients of abundance regressed on years (temporal trend). This is one summary of the ecological process that could alert managers to a decline in abundance. Inference on the parameter $\sigma^2$ based on $[\sigma^2 \mid \{Y_{ij}\}]$ would show how variable the abundances are from site to site, which would serve as a warning that a "one-size-fits-all" management practice would be unsuccessful.

In principle, $[D]$ in Eq. 4 is obtained by integrating out (in the continuous-distribution case) the process and the parameters in the numerator, but in practice it is seldom possible to obtain the constant analytically. As a result, numerical methods must be used. The realization that Markov chain Monte Carlo (MCMC) methods could be used efficiently and generally for Bayesian hierarchical models (Gelfand and Smith 1990) revolutionized such computation, and it extended the applicability of these models to ever-more-complicated modeling scenarios. MCMC is a simulation-based method for drawing samples from probability distributions, where a Markov chain is constructed such that its stationary, or long-run, distribution coincides with the distribution from which random samples are desired. This distribution is sometimes called the "target" distribution which, for a Bayesian analysis, is the posterior distribution. It follows that after a sufficient number of realizations, or a "burn-in," the generated realizations of the chain comprise a random sample from the posterior distribution.

The easiest MCMC algorithm to describe is the Gibbs sampler (e.g., Gelfand and Smith 1990, Roberts and Casella 2005). To sample from $[E, P_E, P_D \mid D]$, we simulate successively from the steps:

$$[E|P_E, P_D, D]$$
$$[P_E|E, P_D, D]$$
$$[P_D|E, P_E, D]$$

and repeat; at each step, we condition on the latest values we obtained from the previous steps. The conditional distributions just above are referred to as the full-conditional distributions.

When one of these full-conditional distributions can only be calculated up to a normalizing constant, we can carry out the simulation in that step by performing a Metropolis-type simulation (e.g., Tierney 1994, Roberts and Casella 2005). For example, consider the first step and suppose that $[E \mid P_E, P_D, D]$ is given by the density:

$$g(\cdots|P_E, P_D, D) \Big/ \int g(E|P_E, P_D, D)dE$$

where $g$ is known but its integral is not. Let $E'$ be the current value of $E$ and suppose that $E^*$ is a simulated random variable of the same dimension as $E$ (from a

distribution centered at $E'$ with certain symmetry properties), where it is easy to simulate $E^*$ (e.g., from a normal distribution). Define

$$E'' \equiv \begin{cases} E^* & \text{with probability, } \min[1, g(E^*)/g(E')] \\ E' & \text{with probability, } 1 - \min[1, g(E^*)/g(E')]. \end{cases}$$

Then $E''$ is the update of $E'$ (given $P_E$, $P_D$, $D$) in that step of the Gibbs sampler. The Metropolis algorithm can slow up the MCMC procedure if the acceptance probability for $E^*$ is not chosen carefully, so where possible we avoid it in the Gibbs sampler. There is much judgment involved in constructing an MCMC algorithm that burns in quickly and yields stable samples from the posterior distribution. Furthermore, given the large number of variables whose posteriors we seek, MCMC can be a challenge. In the harbor seals study, Ver Hoef and Frost (2003) had to use Metropolis-Hastings algorithms for each of the Gibbs sampling steps, $[E \mid P_E, P_D, D]$, $[P_E \mid E, P_D, D]$, and $[P_D \mid E, P_E, D]$.

We note that MCMC differs from standard Monte Carlo integration in that the samples are dependent, since they are realized paths of a Markov chain. Some care should be taken when doing data analysis on the samples to produce the desired summaries of the posterior distribution. Instead of describing how an MCMC algorithm can be set up for a particular model, we refer the reader to some of the excellent overviews in the literature (e.g., Casella and George 1992, Chib and Greenberg 1995, Gilks et al. 1996, Chen et al. 2000, Gelman et al. 2004). For an introduction to MCMC in the ecological literature, see Link (2002).

In order to summarize inferences on the unknowns in a hierarchical Bayesian model using the output from an MCMC simulation, the sampled values from the posterior distribution are used to calculate common distributional summary statistics, such as histograms, means, and variances, of the marginal distributions of the unknowns, conditional on the observations $D$. For example, the samples of a particular unknown $E_1$, from the posterior distribution, can be averaged to approximate the center of the marginal posterior distribution of $E_1$, which is $\mathrm{E}(E_1 \mid D)$, where $\mathrm{E}(\cdot)$ denotes expectation. In addition to the mean of the samples, it is also common to characterize the uncertainty in the marginal posterior distribution of each unknown using the 2.5th and 97.5th percentiles of the posterior samples. The interval defined by these two percentiles approximates the 95% credible interval of each unknown. The interpretation of a 95% credible interval for an unknown $E_1$ is: the posterior probability that $E_1$ falls inside its 95% credible interval is 0.95. Thus, the widths of the credible intervals summarize the uncertainty in inferences on $E_1$. For the harbor seals study (Ver Hoef and Frost 2003), $E_1$ contains the temporal-trend parameters $\{\theta_{1i}\}$ and $\theta_{1A}$, and the marginal distributions of the MCMC samples are given in Fig. 3, from which means, credible intervals, and so forth, can be computed.

There are a variety of different software packages available for fitting Bayesian hierarchical models using MCMC methods. Several of these packages are based on the Bayesian inference Using Gibbs sampling (BUGS) language, including its Windows implementation (WinBUGS; *available online*)[6] and an open-source version (OpenBUGS; *available online*).[7] Another open-source software package for posterior simulation is Just Another Gibbs Sampler (JAGS; *available online*).[8] The JAGS model-specification syntax is nearly identical to BUGS, however the implementation is different. While all of these MCMC software packages can greatly facilitate the fitting of Bayesian hierarchical models, many researchers choose to write their own MCMC code in languages such as R/S, C/C++, and FORTRAN. This option provides increased flexibility over the various packages, but it can be considerably more tedious to implement and debug. For their analysis of the harbor seals data, Ver Hoef and Frost (2003) used WinBUGS to obtain samples from the posterior distribution.

A considerable amount of research in quantitative ecology has been devoted to model selection. We now reconsider this part of inference in the context of hierarchical statistical modeling. For a simple example of model selection, consider again the general problem of "curve fitting." Our goal is to select the "best" curve. How do we define best, and how do we decide which is best? Most methods try to find a balance or can be viewed as a balance between parsimony (simpler is better) and goodness of fit (we want enough complexity to capture the essential features of the data). For prediction, we can consider "averaging" over models, rather than trying to select the "best" one (e.g., Hoeting et al. 1999), which has the effect of combining the advantages from all models under consideration. These same considerations apply when trying to capture uncertainty in process models and parameter models. Textbook treatments of model selection can be found in Linhart and Zucchini (1986), McQuarrie and Tsai (1998), and Burnham and Anderson (2002).

For the harbor seals study, we might want to simplify the model as much as possible by selecting among models that include year and a subset of tide, date, and time. A traditional approach is to use stepwise regression, but more popular recently is an information-theoretic approach, such as AIC (Akaike 1973) or BIC (Schwarz 1978). These approaches are especially useful when comparing non-nested models. For example, suppose that we want to use all covariates, year, tide, date, and time in our model, but we want to compare a Poisson distribution vs. a negative binomial distribution for the measurement errors. Then a stepwise approach is not appropriate, however the information-

[6] ⟨http://www.mrc-bsu.cam.ac.uk/bugs/⟩
[7] ⟨http://mathstat.helsinki.fi/openbugs/⟩
[8] ⟨http://www-fis.iarc.fr/~martyn/software/jags/⟩

theoretic approach handles the comparison easily. The information-theoretic approach uses the notion of maximizing the model likelihood (which can never get worse by adding more parameters), but with a penalty for the number of parameters. However, the counting of parameters can become difficult for hierarchical statistical models, so Spiegelhalter et al. (2002) introduced DIC; DIC would be useful in the harbor seals study, especially when trend over years is considered. Small, Pendleton, and Pitcher (2003) used an information-theoretic approach on a non-hierarchical analysis of the harbor seals data.

## Challenges for Hierarchical Statistical Modeling

Hierarchical models address complex problems for which there may be several sources of information and hidden variables (Wikle et al. 1998, Abbitt and Breidt 2002, Brooks et al. 2004, Ogle et al. 2004, Clark 2005, Royle and Dorazio 2006). Goals may include estimation of parameters for a process model (Hille Ris Lambers et al. 2005, Ibanez et al. 2006), inference on hidden states (Dupuis 1995, Ver Hoef and Cressie 1997, de Valpine 2003, Dorazio and Royle 2003, Stenseth et al. 2003, Clark and Bjornstad 2004, Gelfand et al. 2004, Maunder 2004, Clark et al. 2005, Thomas et al. 2005), quantifying the importance of interactions (Coulson et al. 2001, Cam et al. 2002, Clark et al. 2003, 2004, LaDeau and Clark 2006, Mohan et al. 2007), prediction (Beckage and Platt 2003, Clark et al. 2007, Hooten et al. 2007), and specifying species diversity and patterns (Gelfand et al. 2005*a*, *b*, Latimer et al. 2006). In this section, we begin a discussion of the issues surrounding hierarchical statistical modeling. It is expected that in follow-up discussion, these and other issues will be explored more deeply than we are able to do here.

There are many challenges with building complicated models, with the associated inference, and with computational efficiency (Clark 2005, Clark and Gelfand 2006, and Buckland et al. 2007 provide ecological examples). Like any model-building exercise, there must be considerable thought and effort devoted to the specification of the component distributions in the hierarchical framework. In some sense, model complexity is related both to data richness and "scientific richness," which is simply the knowledge one has about the process (and parameters) of interest. The situation is only made more complex when the model is also Bayesian.

### *Subjectivity of a Bayesian analysis*

A historical criticism of Bayesian methods is that it requires "subjective" specification of prior information on the parameters. Can a Bayesian get any answer he or she desires by tuning the prior distribution? Technically, the answer to this question is yes; the Bayesian approach can be "gamed," but so can the classical/frequentist approach to inference. In fact, the elicitation of prior distributions can be turned into a (social) science. Subjective judgments within a group of expert ecologists can lead to a subjective prior where both the consensus opinion and the diversity of opinions is recognized (e.g., McCarthy 2007: Chapter 10).

There is also a sizable literature on "objective Bayesian" analysis, which develops prior distributions for classes of models that do not impact the posterior distribution. Such noninformative priors are often termed "vague," "flat," or "diffuse." For some models, appropriate noninformative priors are readily available, while for others they are not. For a thorough discussion of issues related to the specification of prior distributions, see Berger (1985). Alternatively, it is common practice to perform sensitivity analyses, or robustness assessments, to ascertain the influence of prior assumptions. This is done by refitting the hierarchical model under different prior assumptions and comparing the resulting inferences. Not only does this exercise provide information on the influence of prior assumptions on inferences, but it gives insight into the amount of information in the data.

Of course, there is subjectivity in the specification of the likelihood in a classical statistical approach. In fact, a broader perspective is that there is subjectivity involved with the specification of all of the model components: data models, process models, and parameter models. For example, it might be "subjective" to specify a stochastic model for tropical winds (e.g., Wikle et al. 2001), but the science comes from Newton's laws of motion! There has always been subjectivity present in the physical sciences; now, instead of having to be certain about a physical model, scientists can quantify their uncertainty about their model in a way that disentangles it from their uncertainty about their data. Thus, from our perspective, it is better to ask about the sensitivity of results to model choices and whether such choices make sense scientifically.

There are cases where one tries to make the posterior distribution in a Bayesian hierarchical statistical model less sensitive to model specifications. Central to the Bayesian paradigm is the notion that as the data quantity and quality increase, the posterior is less sensitive to prior assumptions. This feature of Bayesian inference is often compared philosophically to "scientific learning," in that knowledge is continuously updated in light of new information, in a coherent fashion.

### *Convergence of the MCMC algorithm*

Based on the discussion above, before the sample path of a Markov chain produces a sample from the posterior distribution, the MCMC algorithm must be run for a certain number of burn-in iterations. While this number must be finite, the exact number of burn-in iterations is rarely known. The process by which the chain moves from its starting value to values that are representative of the distribution is termed convergence of the algorithm.

In practice, convergence is usually assessed using a set of diagnostic tools on the sample path of the Markov

chain. While this type of output analysis is not able to "prove" that the chain has converged, it can provide insight into the issue. Output analysis can be numerical and graphical and can involve both comparing chains with multiple starting values and assessing differences in different segments of a single chain (e.g., Cowles and Carlin 1995, Brooks and Roberts 1998, Gelman et al. 2004). Software for assessing convergence of MCMC algorithms includes the convergence diagnosis and output analysis software of Gibbs sampling output (CODA; Best et al. 1995, Plummer et al. 2005) and Bayesian output analysis (BOA; *available online*).[9]

### *Adequacy of the number of MCMC samples*

The sequences of samples from the joint posterior distribution generated by an MCMC algorithm are by construction correlated. As a result, inferences based on MCMC samples (assuming the chain has converged) will generally be less precise than independent samples from the posterior. By examining the variance of the sample paths for a single parameter, marginal estimates of the "effective" number of iterations can be obtained. If there is limited availability of storage for the output of an MCMC algorithm, the chain can be "thinned" by saving only every $k$th sample. The autocorrelation of the resulting thinned chain will likely be less than the original chain. Therefore, the precision of posterior inferences based on the thinned chain will be better than those based on an equal-length unthinned portion of the chain.

### *Parameter identifiability*

When building large hierarchical models, it is not always apparent that all model parameters are identifiable. In order to motivate the issue of parameter identifiability, consider the following simple example. Let $y_1, \ldots, y_n$ be independent samples from a normal distribution with unknown mean $\mu$ and variance $\sigma^2$. Clearly, from a classical/frequentist perspective, the observations contain information about the unknown parameters. In fact, the sample mean and variance provide unbiased estimates of $\mu$ and $\sigma^2$, respectively. Instead of parameterizing the sampling distribution in terms of $\mu$ and $\sigma^2$, assume that the data were generated from a normal distribution with unknown mean $\mu_1 + \mu_2$ and variance $\sigma^2$. While there is information in the data about the sum, $\mu_1 + \mu_2$, there is no way to identify the components separately, which means that $\mu_1$ and $\mu_2$ are not identifiable.

From a Bayesian perspective, as long as proper prior distributions are assigned to all model parameters, all of them are technically identifiable. To illustrate, consider the example above, now within the Bayesian paradigm. Assuming proper prior distributions on $\mu_1$, $\mu_2$, and $\sigma^2$, we can determine the posterior distribution of all three parameters, making them all technically identifiable. However, unless there is very strong prior information on $\mu_1$ and $\mu_2$, it will be difficult to see very much difference in the posterior distributions of these parameters. In practice, lack of identifiability can be problematic in the Bayesian setting (Gelfand and Sahu 1999). For example, when running MCMC algorithms, the sample paths of nonidentifiable parameters will "trade off" their values, leading to numerical and convergence problems. Generally speaking, if identifiability problems go undiagnosed, inferences on these model parameters and possibly others can be misleading. This can sometimes be anticipated, in which case a sampling design (*Statistical inference in ecological analyses*) might be formulated that avoids identifiability problems in making inference from Eq. 3.

### *Assessing model fit and diagnosing lack of fit*

We have already discussed the importance of assessing the influence of prior assumptions on inferences in Bayesian models. In addition to performing these sensitivity analyses, it is important to assess the fit of a model. There are both numerical and graphical tools for model checking, which are primarily based on the posterior predictive distribution, specifically the distribution of a new (replicate) observation, $D^{\text{rep}}$, conditional on the observed data, $D$ (see Gelman et al. 2004: Chapter 6). The posterior predictive distribution is given by

$$[D^{\text{rep}}|D] = \int [D^{\text{rep}}|E, P_E, P_D][E, P_E, P_D|D]\, dE\, dP_E\, dP_D$$

where the distribution is obtained numerically by taking an MCMC sample from the posterior distribution, $[E, P_E, P_D \mid D]$, and plugging the sampled values of $E$, $P_E$, and $P_D$ into $[D^{\text{rep}} \mid E, P_E, P_D]$ to then generate a sample from $[D^{\text{rep}} \mid D]$.

With any model-checking procedure comes the opportunity to perform cross-validation. That is, the $i$th observation $D_i$ (or a carefully chosen subset of observations) in $D$ is deleted, leaving the remaining data, $D_{-i}$, to predict what should have been observed assuming the model is correct. A Bayesian model allows the posterior distribution $[D_i^{\text{rep}} \mid D_{-i}]$ to be computed, which can then be compared to the observation $D_i$ that was deleted (e.g., Stern and Cressie 2000). For example, if $D_i$ is in the tails of the distribution of $[D_i^{\text{rep}} \mid D_{-i}]$, there would be cause to doubt the model's ability to fit well to the $i$th datum. Consequently, if cross-validation is used as an informal model-selection procedure, it tends to reject overly complicated models, since such models perform poorly when predicting beyond the observed data.

While conceptually straightforward, in practice, model checking using the posterior predictive distribution for large hierarchical models can be tedious. The distribution $[D^{\text{rep}} \mid D]$ can be high-dimensional and likely will have a complex dependence structure. Finding distributional summaries that adequately convey the extent to which the data fit the model can be a difficult task. Still,

[9] ⟨http://www.public-health.uiowa.edu/boa⟩

if inferences and forecasts are going to be trusted, these checks must be performed. In the harbor seals study, Ver Hoef and Frost (2003) used $[D^{\text{rep}} \mid D]$ to average $(\text{observed}_{ij} - \text{expected}_{ij})^2/\text{expected}_{ij}$, over $j$, for each site $i$, which should be near one. This is a generic goodness-of-fit statistic that is not tailor-made for any particular departure from the model for which $\text{expected}_{ij}$ is calculated. They found lack of fit for models without a random effect to account for overdispersion. That is, for the models given in Fig. 2, the parameters $\sigma^2$ or $\{\sigma_i^2\}$ are important for accounting for dispersion about the "curves" $\{\mu_{ij}\}$.

### *Bayesian inference in space and time*

In a spatiotemporal setting, process and parameter models are less influential on smoothing and filtering, because the data redeem us from misspecification. This is because spatiotemporal dependencies act to reinforce the information in current and past data about current values of the process and values of the parameters. But process and parameter models can matter a lot when we are forecasting, particularly when there is long-range temporal dependence. The Bayesian part of all this captures the variability in the parameters, whereas if they are estimated and plugged into summaries for inference on the process component, those inferences do not account properly for the variability in the parameter estimates. Frequentist-based approximations (e.g., Rao 2003: section 6.2) are available for simple cases, but a dynamic, non-stationary, spatiotemporal hierarchical model is far from simple. In contrast, Bayesian inference on the spatiotemporal process accounts for parameter variability coherently (e.g., Waller et al. 1997, Wikle et al. 1998, 2001, Berliner et al. 2000, Xu et al. 2005, Calder 2007, Hooten and Wikle 2007, Hooten et al. 2007, Ver Hoef and Jansen 2007).

### *Multivariate hierarchical statistical models*

Ecology is about relationships among natural phenomena that include interactions among species and how species relate to their environment. It would seem, then, that multivariate models would be among the most commonly used. However, such models are notoriously difficult to fit; Ver Hoef and Barry (1998) outline some of the difficulties. Just as a complex univariate problem can be broken into a sequence of simpler ones through conditional distributions, multivariate ecological problems can also benefit from a hierarchical statistical specification. Recent progress in this area can be found in Royle and Berliner (1999), Johnson et al. (2006), Barber and Gelfand (2007), Furrer et al. (2007), Sain and Cressie (2007), and Sims et al. (2008).

To sum up, we have featured the hierarchical statistical modeling approach for dealing with uncertainty in ecological analysis. When modeling the sort of complex problems found in ecology, we present it as a coherent approach to deal with uncertainty in measurement, in sampling, in specification of the process, in knowledge of the parameters, and in initial and boundary conditions. While the approach is very powerful, there remain challenges for its practical implementation. This last section discusses a number of these challenges; it is our hope that our article will engender more discussion and promote further research in hierarchical statistical modeling in ecology.

### Acknowledgments

This paper resulted from a workshop, "Uncertainty in Ecological Analysis," organized by Ohio State University's Mathematical Biosciences Institute and Program in Spatial Statistics and Environmental Statistics, which was held in Columbus, Ohio, 3–6 April 2006. Major financial support for the workshop came from the National Science Foundation under Agreement No. 0112050. The workshop was attended by 50 U.S. and international quantitative ecologists and applied statisticians, as well as about 20 Ohio State faculty, postdocs, and graduate students. Session themes were almost all ecological, the idea being that science problems would drive the workshop, and that quantitative methodology would emerge in a context. Workshop sessions included Population Dynamics, Abundance Estimation, Spatial Patterns and Processes, Community Ecology, and Complexity in Ecosystem Analysis, as well as two New Researchers sessions. Both the presentation slides and videos of the presentations can be found at the MBI's web site.[10] We would like to express our appreciation to the referees, Dave Schimel, and Michelle Sims, whose suggestions and comments led to improvements in the manuscript.

### Literature Cited

Abbitt, P. J., and F. J. Breidt. 2002. A hierarchical model for estimating distribution profiles of soil texture. Pages 263–278 *in* C. Gatsonis, R. E. Kass, B. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West, editors. Case studies in Bayesian statistics. Volume 5. Springer-Verlag, New York, New York, USA.

Akaike, H. 1973. Information theory and the extension of the maximum likelihood principle. Pages 267–281 *in* B. N. Petrov and F. Czaki, editors. Proceedings of the International Symposium on Information Theory. Akademia Kiado, Budapest, Hungary.

Aldworth, J., and N. Cressie. 1999. Sampling designs and prediction methods for Gaussian spatial processes. Pages 1–54 *in* S. Ghosh, editor. Mathematical design and sampling. Marcel Dekker, New York, New York, USA.

Arbia, G., and G. Lafratta. 1997. Evaluating and updating the sample design in repeated environmental surveys: monitoring air quality in Padua. Journal of Agricultural, Biological, and Environmental Statistics 2:451–466.

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2004. Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC Press, Boca Raton, Florida, USA.

Barber, J. J., and A. E. Gelfand. 2007. Hierarchical spatial modeling for estimation of population size. Environmental and Ecological Statistics 14:193–205.

Beckage, B., and W. J. Platt. 2003. Predicting severe wildfire years in the Florida Everglades. Frontiers in Ecology and the Environment 1:235–239.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer-Verlag, New York, New York, USA.

Berliner, L. M. 1996. Hierarchical Bayesian time series models. Pages 15–22 *in* K. Hanson and R. Silver, editors. Maximum entropy and Bayesian methods. Kluwer Academic Publishers, Dordrecht, The Netherlands.

[10] ⟨http://www.mbi.osu.edu/2005/ws5abstracts.html⟩

Berliner, L. M., C. K. Wikle, and N. Cressie. 2000. Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. Journal of Climate 13:3953–3968.

Best, N. G., M. K. Cowles, and S. K. Vines. 1995. CODA Manual version 0. 30. MRC Biostatistics Unit, Cambridge, UK.

Bjornstad, O. N., and B. F. Grenfell. 2001. Noisy clockwork: time series analysis of population fluctuations in animals. Science 293:638–643.

Boily, P. 1995. Theoretical heat flux in water and habitat selection of phocid seals and beluga whales during the annual molt. Journal of Theoretical Biology 172:235–244.

Botkin, D. B., J. F. Janak, and J. R. Wallis. 1972. Some ecological consequences of a computer model of forest growth. Journal of Ecology 60:849–872.

Boveng, P. L., J. L. Bengtson, J. C. Cesarone, M. A. Simpkins, K. J. Frost, and J. J. Burns. 2003. The abundance of harbor seals in the Gulf of Alaska. Marine Mammal Science 19:111–127.

Brooks, S. P., R. King, and B. J. T. Morgan. 2004. A Bayesian approach to combining animal abundance and demographic data. Animal Biodiversity and Conservation 27:515–529.

Brooks, S. P., and G. O. Roberts. 1998. Assessing convergence of Markov Chain Monte Carlo algorithms. Statistics and Computing 8:319–335.

Browne, W. J., and D. Draper. 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis 1:473–514.

Buckland, S. T., K. B. Newman, C. Fernández, L. Thomas, and J. Harwood. 2007. Embedding population dynamics models in inference. Statistical Science 22:44–58.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

Calder, C. A. 2007. Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment. Environmental and Ecological Statistics 14:229–247.

Calder, C. A., M. Lavine, P. Mueller, and J. S. Clark. 2003. Incorporating multiple sources of stochasticity in population dynamic models. Ecology 84:1395–1402.

Cam, E., W. A. Link, E. G. Cooch, J.-Y. Monnat, and E. Danchin. 2002. Individual covariation in life-history traits: seeing the trees despite the forest. American Naturalist 159: 96–105.

Carlin, B. P., and T. A. Louis. 2000. Bayes and empirical Bayes methods for data analysis. Second edition. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Casella, G., and E. I. George. 1992. Explaining the Gibbs sampler. American Statistician 46:167–174.

Catchpole, E. A., E. A. Hatton, and W. R. Catchpole. 1989. Fire spread through nonhomogeneous fuel modeling as a Markov process. Ecological Modelling 48:101–112.

Chao, C.-T., and S. K. Thompson. 2001. Optimal adaptive selection of sampling sites. Environmetrics 12:517–538.

Chen, M.-H., Q.-M. Shao, and J. G. Ibrahim. 2000. Monte Carlo methods in Bayesian computation. Springer Verlag, New York, New York, USA.

Chib, S., and E. Greenberg. 1995. Understanding the Metropolis-Hastings algorithm. American Statistician 49:327–335.

Christensen, R. 1991. Linear models for multivariate, time series, and spatial data. Springer-Verlag, New York, New York, USA.

Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8:214.

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

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

Clark, J. S., G. Ferraz, N. Oguge, H. Hayes, and J. DiCostanzo. 2005. Hierarchical Bayes for structured and variable populations: from capture–recapture data to life-history prediction. Ecology 86:2232–2244.

Clark, J. S., and A. E. Gelfand. 2006. A future for models and data in ecology. Trends in Ecology and Evolution 21:375–380.

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

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

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

Cochran, W. G. 1977. Sampling techniques. Third edition. Wiley, New York, New York, USA.

Coulson, T., E. A. Catchpole, S. D. Albon, B. J. T. Morgan, J. M. Pemberton, T. H. Clutton-Brock, M. J. Crawley, and B. T. Genfell. 2001. Age, sex, density, winter weather and population crashes in Soay sheep. Science 292:1528–1531.

Cowles, M. K., and B. P. Carlin. 1995. Markov chain Monte Carlo diagnostics: a comparative review. Journal of the American Statistical Association 91:883–904.

Cressie, N. 1992. Smoothing regional maps using empirical Bayes predictors. Geographical Analysis 24:75–95.

Cressie, N. 1993. Statistics for spatial data, revised edition. Wiley, New York, New York, USA.

Cressie, N. 1998. Transect-spacing design of ice cores on the Antarctic continent. Canadian Journal of Statistics 26:405–418.

da Silva, J., and J. M. Terhune. 1988. Harbour seal grouping as an anti-predator strategy. Animal Behavior 36:1309–1316.

Daszak, P., A. A. Cummingham, and A. D. Hyatt. 2000. Emerging infectious diseases of wildlife: threats to biodiversity and human health. Science 287:443–449.

de Valpine, P. 2003. Better inferences from population-dynamics experiments using Monte Carlo state-space likelihood methods. Ecology 84:3064–3077.

Dorazio, R. M., and J. A. Royle. 2003. Mixture models for estimating the size of a closed population when capture rates vary among individuals. Biometrics 59:351–364.

Dupuis, J. A. 1995. Bayesian estimation of movement and survival probabilities from capture–recapture data. Biometrika 82:761–772.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York, New York, USA.

Federov, V. V., and C. Nachtsheim. 1995. Optimal designs for time-dependent responses. Pages 3–13 *in* C. P. Kitsos and W. G. Muller, editors. Proceedings of MODA4. Physica-Verlag, Heidelberg, Germany.

Frost, K. J., L. F. Lowry, and J. M. Ver Hoef. 1999. Monitoring the trend of harbor seals in Prince William Sound, Alaska, after the *Exxon Valdez* oil spill. Marine Mammal Science 15:494–506.

Furrer, R., S. R. Sain, D. Nychka, and G. A. Meehl. 2007. Multivariate Bayesian analysis of atmosphere–ocean general circulation models. Environmental and Ecological Statistics 14:249–266.

Gelfand, A. E., and S. Sahu. 1999. Indentifiability, improper priors, and Gibbs sampling for generalized linear models. Journal of the American Statistical Association 94:247–253.

Gelfand, A. E., A. M. Schmidt, S. Wu, J. A. Silander, A. Latimer, and A. G. Rebelo. 2005*a*. Explaining species diversity through species level hierarchical modeling. Applied Statistics 65:1–20.

Gelfand, A. E., J. A. Silander, S. Wu, A. Latimer, P. O. Lewis, A. G. Rebelo, and M. Holder. 2005*b*. Explaining species

distribution patterns through hierarchical modeling. Bayesian Analysis 1:41–92.
Gelfand, A. E., and A. F. M. Smith. 1990. Sampling based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398–409.
Gelfand, A. E., I. Zhu, and B. P. Carlin. 2001. On the change of support problem for spatio-temporal data. Biostatistics 2:31–45.
Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis. Second edition. Chapman and Hall/CRC, New York, New York, USA.
Gelman, A., and J. Hill. 2006. Data analysis using regression and mutlilevel/hierarchical models. Cambridge University Press, New York, New York, USA.
Gilks, W. R., S. Richarson, and D. J. Spiegelhalter. 1996. Markov chain Monte Carlo in practice. Chapman and Hall, New York, New York, USA.
Gotway, C. A., and L. J. Young. 2002. Combining incompatible spatial data. Journal of the American Statistical Association 97:632–648.
Hardin, J. W., and J. M. Hilbe. 2003. Generalized estimating equations. Chapman and Hall/CRC, Boca Raton, Florida, USA.
Hilborn, R., and M. Mangel. 1997. The ecological detective. Confronting models with data. Princeton University Press, Princeton, New Jersey, USA.
Hille Ris Lambers, J., J. S. Clark, and M. Lavine. 2005. Implications of seed banking for recruitment of southern Appalachian woody species. Ecology 86:85–95.
Hoeting, J. A., D. Madigan, A. Raftery, and C. T. Volinsky. 1999. Bayesian model averaging: a tutorial. Statistical Science 14:382–417.
Hooten, M. B., and C. K. Wikle. 2007. Shifts in the spatial-temporal growth dynamics of shortleaf pine. Environmental and Ecological Statistics 14:207–227.
Hooten, M. B., C. K. Wikle, R. M. Dorazio, and J. A. Royle. 2007. Hierarchical spatiotemporal matrix models for characterizing invasions. Biometrics 63:558–567.
Hoover-Miller, A., K. R. Parker, and J. J. Burns. 2001. A reassessment of the impact of the *Exxon Valdez* oil spill on harbor seals (*Phoca vitulina richardsi*) in Prince William Sound, Alaska. Marine Mammal Science 17:111–135.
Huang, H.-C., and N. Cressie. 1996. Spatio-temporal prediction of snow-water equivalent using the Kalman filter. Computational Statistics and Data Analysis 22:159–175.
Ibáñez, I., J. S. Clark, M. C. Dietze, K. Feeley, M. Hersh, S. LaDeau, A. McBride, N. E. Welch, and M. S. Wolosin. 2006. Predicting biodiversity change: outside the climate envelope, beyond the species–area curve. Ecology 87:1896–1906.
Johnson, D. S., J. A. Hoeting, and N. L. Poff. 2006. Biological monitoring: a Bayesian model for multivariate compositional data. Pages 270–289 *in* S. K. Upadhyay, U. Singh, and D. K. Dey, editors. Bayesian statistics and its applications. Anamaya Publishers, New Delhi, India.
LaDeau, S. L., and J. S. Clark. 2006. Elevated $CO_2$ and tree fecundity: the role of tree size, interannual variability, and population heterogeneity. Global Change Biology 12:822–833.
Latimer, A. M., S. Wu, A. E. Gelfand, and J. A. Silander. 2006. Building statistical models to analyze species distributions. Ecological Applications 16:33–50.
Le, N. D., and J. V. Zidek. 1994. Network designs for monitoring multivariate random spatial fields. Pages 191–206 *in* J. P. Vilaplana and M. L. Puri, editors. Recent advances in statistics and probability. VSP, Leiden, The Netherlands.
Lele, S. R., B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. Ecology Letters 10:551–563.
Linhart, H., and W. Zucchini. 1986. Model selection. Wiley, New York, New York, USA.
Link, W. A., E. Cam, J. D. Nichols, and E. G. Cooch. 2002. Of BUGS and birds: Markov chain Monte Carlo for hierarchical modeling in wildlife research. Journal of Wildlife Management 66:277–291.
Maunder, M. N. 2004. Population viability analysis based on combining integrated, Bayesian, and hierarchical analysis. Acta Oecologica 26:85–94.
McCarthy, M. A. 2007. Bayesian methods for ecology. Cambridge University Press, Cambridge, UK.
McQuarrie, A. D., and C.-L. Tsai. 1998. Regression and time series model selection. World Scientific, Hackensack, New Jersey, USA.
Medawar, P. B. 1940. Growth, growth energy, and ageing of the chicken's heart. Proceedings of the Royal Society 129: 332–355.
Meinhold, J., and N. Singpurwalla. 1983. Understanding the Kalman filter. American Statistician 37:123–127.
Mohan, J. E., J. S. Clark, and W. H. Schlesinger. 2007. Long-term $CO_2$ enrichment of an intact forest ecosystem: implications for temperate forest regeneration and succession. Ecological Applications 17:1198–1212.
Morris, C. N. 1983*a*. Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association 78:47–55.
Morris, C. N. 1983*b*. Parametric empirical Bayes confidence intervals. Pages 25–50 *in* G. E. P. Box, T. Leonard, and C. F. J. Wu, editors. Scientific inference, data analysis, and robustness. Academic Press, New York, New York, USA.
Mugglin, A. S., B. P. Carlin, and A. E. Gelfand. 2000. Fully model-based approaches for spatially misaligned data. Journal of the American Statistical Association 95:877–887.
Ogle, K., R. L. Wolpert, and J. F. Reynolds. 2004. Reconstructing plant root area and water uptake profiles. Ecology 85:1967–1978.
Overton, W. S. 1993. Probability sampling and population inference in monitoring programs. Pages 470–480 *in* M. R. Goodchild, B. O. Parks, and C. T. Steyaert, editors. Environmental modeling with GIS. Oxford University Press, New York, New York, USA.
Plummer, M., N. Best, K. Cowles, and K. Vines. 2005. CODA: output analysis and diagnostics for MCMC. R package version 0. 9-2. ⟨http://www-fis.iarc.fr/coda/⟩
Prasad, N. G. N., and J. N. K. Rao. 1990. On the estimation of mean square error of small area predictors. Journal of the American Statistical Association 85:163–171.
Rao, J. N. K. 2003. Small area estimation. Wiley, New York, New York, USA.
Ricker, W. E. 1954. Stock and recruitment. Journal of Fisheries Research Board of Canada 11:559–623.
Roberts, C. P., and G. Casella. 2005. Monte Carlo statistical methods. Second edition. Springer, New York, New York, USA.
Royle, J. A., and L. M. Berliner. 1999. A hierarchical approach to multivariate spatial modeling and prediction. Journal of Agricultural, Biological, and Environmental Statistics 4:29–56.
Royle, J. A., and R. M. Dorazio. 2006. Hierarchical models of animal abundance and occurrence. Journal of Agricultural, Biological, and Environmental Statistics 11:249–263.
Sacks, W., D. Schimel, and R. Monson. 2007. Coupling between carbon cycling and climate in a high-elevation, subalpine forest: a model-data fusion analysis. Oecologia 151:54–68.
Sain, S., and N. Cressie. 2007. A spatial model for multivariate lattice data. Journal of Econometrics 140:226–259.
Särndal, C., B. Swensson, and J. Wretman. 1992. Model assisted survey sampling. Springer-Verlag, New York, New York. USA.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.

Sims, M., R. Lewison, and T. Cox. 2008. Identifying spatial patterns in fisheries by catch: using models to improve stability of estimates and aid management. Ecological Applications 18:649–661.

Small, R. J., G. W. Pendleton, and K. W. Pitcher. 2003. Trends in abundance of Alaska harbor seals, 1983–2002. Marine Mammal Science 19:344–362.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, Series B 64:583–639.

Stenseth, N. C., H. Viljugrein, T. Saitoh, T. T. Hansen, M. O. Kittilsen, E. Bolviken, and F. Glockner. 2003. Seasonality, density dependence, and population cycles in Hokkaido voles. Proceedings of the National Academy of Sciences (USA) 100:11478–11483.

Stern, H. S., and N. Cressie. 2000. Posterior predictive model checks for disease mapping models. Statistics in Medicine 19: 2377–2397.

Stevens, D. L., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99:262–278.

Thomas, L., S. T. Buckland, K. B. Newman, and J. Harwood. 2005. A unified framework for modelling wildlife population dynamics. Australian and New Zealand Journal of Statistics 47:19–34.

Thompson, S. K., and A. F. Seber. 1996. Adaptive sampling. Wiley, New York, New York, USA.

Tierney, L. 1994. Markov chains for exploring posterior distributions (with discussion). Annals of Statistics 22:1701–1762.

Ver Hoef, J. M. 1996. Parametric empirical Bayes methods for ecological applications. Ecological Applications 6:1047–1055.

Ver Hoef, J. M. 2002. Sampling and geostatistics for spatial data. Ecoscience 9:152–161.

Ver Hoef, J. M., and R. D. Barry. 1998. Constructing and fitting models for cokriging and multivariable spatial prediction. Journal of Statistical Planning and Inference 69: 275–294.

Ver Hoef, J. M., and N. Cressie. 1997. Using hidden Markov chains and empirical Bayes change-point estimation for transect data. Environmental and Ecological Statistics 4: 247–264.

Ver Hoef, J. M., and K. Frost. 2003. A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska. Environmental and Ecological Statistics 10: 201–209.

Ver Hoef, J. M., and J. K. Jansen. 2007. Space-time zero-inflated count models of harbor seals. Environmetrics 18: 697–712.

Waller, L. A., B. P. Carlin, H. Xia, and A. Gelfand. 1997. Hierarchical spatio-temporal mapping of disease rates. Journal of the American Statistical Association 92:607–617.

Watts, P. 1992. Thermal constraints on hauling out by harbor seals (*Phoca vitulina*). Canadian Journal of Zoology 70:553–560.

West, M., and J. Harrison. 1997. Bayesian forecasting and dynamic models. Springer-Verlag, New York, New York, USA.

Wikle, C. K. 2003*a*. Hierarchical models in environmental science. International Statistical Review 71:181–199.

Wikle, C. K. 2003*b*. Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84:1382–1394.

Wikle, C. K., and L. M. Berliner. 2005. Combining information across spatial scales. Technometrics 47:80–91.

Wikle, C. K., L. M. Berliner, and N. Cressie. 1998. Hierarchical Bayesian space-time models. Environmental and Ecological Statistics 5:117–154.

Wikle, C. K., and N. Cressie. 1999. A dimension-reduced approach to space-time Kalman filtering. Biometrika 86:815–829.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner. 2001. Spatio-temporal hierarchical Bayesian modeling: tropical ocean surface winds. Journal of the American Statistical Association 96:382–397.

Wikle, C. K., and J. A. Royle. 1999. Space-time models and dynamic design of environmental monitoring networks. Journal of Agricultural, Biological, and Environmental Statistics 4:489–507.

Wikle, C. K., and J. A. Royle. 2005. Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. Environmetrics 16:507–522.

Wright, S. 1926. Book reviews. Journal of the American Statistical Association 21:493–497.

Xu, K., C. K. Wikle, and N. I. Fox. 2005. A kernel-based spatio-temporal dynamical model for nowcasting radar precipitation. Journal of the American Statistical Association 100:1133–1144.

Zimmerman, D. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17:635–662.