

**Accounting for Uncertainty in Ecological Analysis: The Strengths
And Limitations of Hierarchical Statistical Modeling**

N. Cressie, *The Ohio State University*

C. A. Calder, *The Ohio State University*

J. S. Clark, *Duke University*

J. M. Ver Hoef, *National Marine Mammal Laboratory, NOAA*

C. K. Wikle, *University of Missouri-Columbia*

Technical Report No. 798

July, 2007

**Department of Statistics
The Ohio State University
1958 Neil Avenue
Columbus, OH 43210-1247**

Accounting for Uncertainty in Ecological Analysis: The Strengths and Limitations of Hierarchical Statistical Modeling

Noel Cressie¹, Catherine A. Calder¹, James S. Clark², Jay M. Ver Hoef³,
And Christopher K. Wikle⁴

July 11, 2007

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 error are often the only sources of uncertainty confronted when addressing complex ecological problems, yet analyses need to 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 for accounting 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 of notation and technology from which discussion on hierarchical statistical modeling in ecology can emanate. 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 uncertainties, even if practicalities sometimes necessitate pragmatic compromises.

1. 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 (Daszak et al., 2000; Clark et al., 2001; Beckage and Platt, 2003; Clark, 2005; Ibenez 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. 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,

¹ Program in Spatial Statistics and Environmental Sciences, Department of Statistics, The Ohio State University.

² Nicholas School of the Environment, Duke University.

³ National Marine Mammal Laboratory, NOAA.

⁴ Department of Statistics, University of Missouri-Columbia.

incorporating scientific knowledge through substantive modeling of ecological processes is essential. Often such theoretical contributions are 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*. For example, more realistic descriptions of ecosystems might allow parameters to vary over space and time. 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, Carlin and Louis (2000), where the two approaches are presented.

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 was path-breaking in the way that hierarchical statistical modeling became central to dealing with uncertainty in ecology. As will be explicated in the sections that follow, it is a statistical methodology for handling complex (ecological) problems; it is a framework for working in science teams; and it partitions variability/uncertainty in a way that can suggest study designs where uncertainty can be controlled. While there was not a consensus among the participants on how parameter uncertainty should be handled, the hierarchical statistical approach emerged as an essential technology in modern analyses of ecological data.

The quantitative rough diamonds that fell out of the presentations and discussions of the workshop run the risk of being lost to those in the ecology community who did not attend; the conference organizers have polished the rough diamonds from their own perspective. This paper does not introduce new methodology, but it is meant to reach and engender discussion from readers of this journal. The Bayesian and non-Bayesian versions of the hierarchical statistical model are presented. As the themes of the article are developed, we use an ecological study of abundance of haul-out sites in Alaska, as reported by Ver Hoef and Frost (2003), to illustrate the accompanying notation and theory.

The organization of the paper is as follows. Section 2 addresses the general notion of modeling in the presence of uncertainty and features hierarchical statistical modeling. Section 3 addresses the important aspect of design for data collection. Section 4 discusses statistical inference for ecological analysis. Challenges for hierarchical statistical modeling are presented for discussion in Section 5.

2. Modeling in the presence of uncertainty

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. To illustrate the general notions that follow, we shall consider harbor-seals data from Ver Hoef and Frost (2003) in Prince William Sound in Alaska, where aerial surveys were flown over haul-out sites and all seals were counted. Surveys were flown repeatedly for 7-10 days in August or September each year from 1990-1999. The data were collected as part of routine monitoring

following the Exxon Valdez oil spill in 1989. We return to this study throughout the article as different notions are discussed.

Why does one build models? Is it simply to organize information? Does it 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 group 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 given E happens, E happens conditionally on 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' behavior, regardless of whether observations are taken or not. Formally, we can link such thinking in 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 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 P . Thus, we might think about estimating the parameters of this distribution, $[D|P]$. That is, can we find an estimate of the parameters P 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, $[A|B]$ would be read as “the distribution of A conditioned upon B ”, or “the distribution of A given B ”. In this framework, the parameters P are assumed to summarize the ecological process E appropriately, and it is assumed there is no uncertainty at this stage. 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 is on a data model, where it is assumed the uncertainty lies 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 $[E | P_E]$. If we do not distinguish between *observations on E* (which we call D) and their true values, then the usual likelihood analysis ignores uncertainties about the relationship between E and D , resulting in incorrect statistical inferences.

Recall the harbor-seals study introduced earlier in this section, where aerial surveys were flown over haul-out sites and all seals were counted; surveys were flown repeatedly for 7-10 days. A map showing the geographical context and the sampling plan is given in Figure 1.

Figure 1 here

Let Y_{ij} denote the j^{th} count for the i^{th} site, $i = 1, 2, \dots, n$. 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 (e.g., Boveng et al., 2003) is a natural 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 hauled is more or less than expected. Let this expected number be λ_{ij} , so we might consider $Y_{ij} \sim f(\lambda_{ij}, \kappa)$, where f is a Poisson or negative binomial distribution with mean parameter λ_{ij} and variance parameter κ (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 $[D | E, P_D]$, where this data model is conditional on the ecological process E and a parameter κ expressing variability in the data.

This model has very little “ecology” in it and would be very difficult to estimate because we have one parameter (λ_{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 that they change seasonally (date within year) and yearly (trend). These factors reflect our understanding of harbor-seal biology, such as 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 on the “expected” harbor-seal abundance, without error, on every haul-out site for every instant of time. Lacking that, we use a “smooth curve” of the ecological process that controls haul-out abundance, and we account for errors from that model using a random process. The curve includes information on year, date (within year), tide (height), and time (of day). We let $\log(\lambda_{ij})$ be distributed as a normal distribution with mean $\mu_{ij} \equiv \mu_i(\text{year}_{ij}, \text{date}_{ij}, \text{tide}_{ij}, \text{time}_{ij}; \theta_i)$ and variance σ^2 for the i -th site. Hence, $\exp(\mu_{ij})$ could be used to predict abundance, but it should account for the variability (i.e., the uncertainty) in the process model. For example, the smooth curve could be modeled as,

$$\mu_i(\text{year}_{ij}, \text{date}_{ij}, \text{tide}_{ij}, \text{time}_{ij}; \theta_i) = \theta_{0i} + \theta_{1i}\text{year}_{ij} + \theta_{2i} \text{date}_{ij} + \theta_{3i} (\text{date}_{ij})^2 + \theta_{4i}\text{tide}_{ij} + \theta_{5i}(\text{tide}_{ij})^2 + \theta_{6i} \text{time}_{ij} + \theta_{7i} (\text{time}_{ij})^2,$$

where $\theta_i = (\theta_{0i}, \theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{4i}, \theta_{5i}, \theta_{6i}, \theta_{7i})$. In terms of our previous discussion, recall $E = \{\lambda_{ij}\}$ and write $P_E = \{\theta, \sigma^2\}$. Hence, we can obtain the conditional distribution of E , $[E | P_E]$, conditional on the parameters P_E .

It is then natural to link *hierarchically* the conditional modeling described above so that one may account for the uncertainty in the process E separately from the uncertainty in the observations D (Calder et al., 2003; Wikle, 2003a). We do this by considering the distribution, $[D | E, P_D]$, where recall that P_D refers to the parameters that describe the distribution of the observations. Note that in the first formulation, we modeled $[D | P]$; in effect, we recognize that part of P may be the true process E , which deserves special consideration. In principle, one could try to do inference by maximizing this likelihood jointly in terms of E and P_D . However, this can be very complicated in settings with spatial, temporal, and multiple-species dependencies, and it does not allow one to specify directly uncertainties and stochastic components related to the process E .

Combining this data model, $[D | E, P_D]$, with the process model, $[E | P_E]$, implies the joint uncertainties in the data D and the ecological process E can be written as

$$[D, E | P_D, P_E] = [D | E, P_D][E | P_E], \quad (1)$$

which shows how they 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. In terms of the harbor-seal study, (1) says that we obtain a joint model of the data and the underlying ecological process, $[\{Y_{ij}\}, \{\lambda_{ij}\} | \kappa, \theta, \sigma^2]$, now conditional on fewer parameters. Figure 2(A) shows this hierarchical model schematically, with conditional dependencies shown by arrows linking the relevant boxes.

Figure 2 here

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 measurement error comes from the exponential family of distributions (e.g., binomial, Poisson, gamma, Gaussian, and so forth).

The formula given by (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, $[A, B, C]$. Equivalently, we could use the following result based on conditional probabilities: $[A, B, C] = [A | B, C][B | C][C]$. That is, the joint distribution can be factored into a series of three distributions, namely two conditional distributions, $[A | B, C]$ and $[B | C]$, and one marginal distribution, $[C]$. We refer to such a representation as a *hierarchical statistical model*; equation (1) is such a representation for two random quantities, D

and E , conditional on parameters $P = \{P_D, P_E\}$. Hierarchical statistical modeling simply means that we decompose the joint distribution into a probabilistically valid series of conditional models: $[D, E | P] = [D | E, P][E | P]$. We note that a conditional-probability decomposition is not unique; for example, we could write equivalently, $[A, B, C] = [B | C, A][C | A][A]$, 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.

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

$$[\text{data, process, parameters}] = [\text{data} | \text{process, parameters}] \times [\text{process} | \text{parameters}] \times [\text{parameters}] .$$

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:

$$[\text{data, process} | \text{parameters}] = [\text{data} | \text{process, parameters}] \times [\text{process} | \text{parameters}] ,$$

which is equation (1). The parameters could then be estimated in a separate inference step (empirical Bayesian analysis; e.g., Ver Hoef, 1996) based on $[\text{data} | \text{parameters}]$, or their uncertainty could be quantified distributionally (Bayesian analysis; e.g., Link et al., 2002).

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, $[D | E, P] = [D | E, P_D]$ and $[E | P] = [E | P_E]$ (e.g., Wikle, 2003a). 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. The real power of this approach for complicated problems comes from the fact that each of the component distributions in (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 $P_D = \{P_D^{(1)}, P_D^{(2)}, P_D^{(3)}\}$. 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. 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; Mugglin et al., 2000; Gelfand et al., 2001; Wikle et al., 2001; Banerjee et al., 2004; Clark et al., 2004; Wikle and Berliner, 2005; LaDeau and Clark, 2006; Hille Ris Lambers et al., 2006).

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

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

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 a model,

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

That is, a joint model for the data and linear regression parameters was built conditionally. Here, $[D^{(i)} | E^{(i)}]$ is the Poisson regression formulation at each site, given earlier. But now suppose that all intercepts across sites have a distribution with a common mean for the area, all slope parameters across sites have a distribution with a common mean for the area, and so forth, which we write as $[E^{(1)}, \dots, E^{(n)} | E^{(A)}, P_E]$. This model is shown in Figure 2(B), which extends the model in Figure 2(A). Further, let the collection of common means be normally distributed and independent for each parameter, which we write as $[E^{(A)} | P_E]$. 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.

Figure 3 shows the posterior density of the individual slopes $\{\theta_i : i = 1, \dots, n\}$ for year (trend) in the linear model for μ_i given above (Figure 2(B)), along with the posterior density of the slope θ_{1A} , where recall that $\theta_{11}, \dots, \theta_{1n}$ are conditionally distributed with mean θ_{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 (Figure 2(A)), is shown with solid circles along the x-axis of Figure 3. The one-at-a-time site means are linked by the dashed line to their counterparts under the conditional model with mean θ_{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.

Figure 3 here

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, with Markov assumptions. Specifically, if one has the time series, $E_1, E_2, \dots, E_{T-1}, E_T$, 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:

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

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 (2003b) uses a reaction-diffusion partial differential equation to motivate a stochastic process model for invasive species.

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 stages of the hierarchy, there will typically be unknown parameters that could be estimated or whose uncertainty could be captured probabilistically by (prior) statistical distributions.

3. Design for data collection in ecological studies

Before moving to discussion of statistical inference for hierarchical models (Section 4), in this section we consider the issue of design for data collection from a statistical modeling perspective. Specifically, we highlight the importance of collecting data using sampling designs based on the principles of randomization, stratification, and replication, even when data are analyzed using the hierarchical statistical methods discussed in the previous section. In addition, we show how sampling designs can be incorporated into statistical models, resulting in what is termed model-assisted design-based inferences.

To many people who have taken a course in Design in a Statistics program, the topic brings back memories of 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 Su Doku, 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). We begin with a general discussion of design concepts and then relate it to the formalization given in Section 2.

The key features 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. An instance of poor design would be to study how soil moisture affects seed production but to fail to stratify on elevation. The study would miss different soil-moisture/seed-production relationships for the same plant species but at different elevations. In that instance, “elevation” has been confounded (or not controlled for) in the 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. A sampling design becomes a

probability sampling design when there is a known non-zero 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.

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. (2002). Absent the practical difficulties mentioned above, a well designed study should stratify by date, tide, and time of day, and should randomly sample sites within those strata over the whole region of interest.

Even though design-based inference on the population is not permissible without random sampling, the study can still provide understanding about the relationship between abundance (hauling out) and important covariates, like tide, date, and time of day. Stratification can help ensure sampling over a range of conditions for these covariates. For example, we could divide the time relative to low tide into three categories: more than 1/2 hour before low tide, 1/2 hour before to 1/2 hour after low tide, and more than 1/2 hour after low tide. Likewise, we could divide time of day into three categories and date into three categories. Then, one should at least try to get samples from each stratum for each site, and, if the budget permits, from some well chosen cross-classified strata. Unfortunately, the dates the seals were counted came earlier in the year over the 10 years of the study. As a result, the date and year covariates are confounded. That is, without making strong assumptions, it is not possible to separate how abundance varied by date from how it varied over time, the latter effect being of primary interest in assessing the impact of the oil spill. Even with the most sophisticated statistical modeling tools, this study-design limitation is difficult to overcome.

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 science team, whose goal is to avoid confounding 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. Even those with a basic knowledge of statistics realize that

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. Evidently, replication can be increased by spending more money or reducing the number of strata.

Good design specifies, in advance, inference thresholds and determines the number of observations per stratum needed to achieve those thresholds. 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; 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).

Sampling design has traditionally eschewed the modeling approach given in Section 2. However, by using conditional-probability models, we show how traditional sampling design can be related to the approach we recommend below, that of model-assisted design-based inference. Given that many ecological processes evolve dynamically through space and time, purely spatial designs are typically not as efficient as those that consider spatio-temporal dependence. For the most part, the construction of optimal designs in the spatio-temporal 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). However, as demonstrated in Wikle and Royle (1999, 2005), even more efficient designs in the spatio-temporal context can be obtained by allowing the design to change with time. 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, it is 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. In traditional sampling design, the unknown population is sampled randomly. That is, a probability distribution is put on $\{S : S \subset A\}$, 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 Section 2, it depends on $[S | E]$.

Suppose that $S = S^{(1)}$ is observed, and hence $E^{(1)} \equiv \{E_s : s \in S^{(1)}\}$ 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 , $\bar{E}^{(1)} \equiv \{E_s : s \in A \setminus S^{(1)}\}$, from the observed part $E^{(1)}$, where $A \setminus S^{(1)}$ means the index set of A minus those in $S^{(1)}$. Traditional

quantities of interest are the population total and the population mean: $T_E \equiv \sum_{s \in A} E_s$, and $\mu_E \equiv \sum_{s \in A} E_s / \sum_{s \in A} 1$, respectively, as well as the population variance: $\sigma_E^2 \equiv \sum_{s \in A} (E_s - \mu_E)^2 / \sum_{s \in A} 1$ (e.g., Cochran, 1977).

Estimators of population quantities are functions of $S^{(1)}$ and $E^{(1)}$. For example, consider estimating the total T_E . We can write it as, $T_E = \sum_{s \in S^{(1)}} E_s + \sum_{s \in A \setminus S^{(1)}} E_s$. 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, 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 $[S^{(1)} | E]$. Notice that sampling moments (e.g., mean, variance) are calculated with respect to the $S^{(1)}$ that *could have* occurred, through the probability $[S^{(1)} | E]$.

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 $\bar{E}^{(0)} = E \setminus E^{(0)}$, the unsampled part of E , requires further assumptions that may be difficult to justify.

Surprisingly, much of the past sampling-design literature assumes that E can be observed without error (as we did above). From the discussion in Section 2, 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:

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

Model-assisted design-based inference (e.g., Särndal et al., 1992) 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 that $P = \{P_D, P_E\}$. If parameters P are assumed 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 $[D^{(1)}, S^{(1)} | P]$.

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 $[S^{(1)} | E] = [S^{(1)}]$, 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, sampling uncertainty disappears from the inference, yielding the purely model-based inference developed in Section 2. (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), Ch. 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 $[E | P_E]$. 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) assumed that the measurement process $D^{(i)}$ had Poisson variation with means $\{\lambda_{ij}\}$; then inference is based on $[\{\lambda_{ij}\}, \kappa, \theta, \sigma^2 | D^{(1)}, \dots, D^{(n)}]$, where there is now no reference to the sampling scheme in the notation.

We have 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, pp. 316-323), in which case the approach given in Section 2 is appropriate. More generally, the posterior distribution (3), which involves the sampling probabilities, should be used for inference.

4. Statistical inference in ecological analyses

In Section 2, 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, $[D | E, P_D]$ and $[E | P_E]$, 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, Ch. 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 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 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. This 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: $[E, P_E, P_D | D]$. Our interest might be in terms of 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:

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

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.

For the harbor-seals study, if we assume that the measurement error defining D 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 for a random effect; in the absence of other information, the prior $[P_E]$ can be chosen to be very diffuse and noninformative. Then inference is based on $[\{\lambda_{ij}\}, P_D, P_E | D]$, where $P_E = \{\theta, \sigma^2\}$.

In principle, $[D]$ in (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; Robert and Casella, 2005). To sample from $[E, P_E, P_D / D]$, we simulate successively from the steps:

$$\begin{aligned} & [E / P_E, P_D, D] \\ & [P_E / E, P_D, D] \\ & [P_D / E, P_E, D], \end{aligned}$$

and repeat; at each step we condition on all 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; Robert and Casella, 2005). For example, consider the first step and suppose that $[E / P_E, P_D, D]$ is given by the density,

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

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 E^* is easy to simulate (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, $[E / P_E, P_D, D]$, $[P_E / E, P_D, D]$, and $[P_D / E, P_E, D]$.

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 et al., 1995; Gilks et al., 1996; Chen et al., 2000; Gelman et al., 2004; and for an introduction 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 $E(E_1 | D)$, where $E(\cdot)$ 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 2.5 and 97.5 percentiles of the posterior samples. The interval defined by these two percentiles approximates the 95-percent credible interval of each unknown. The interpretation of a 95-percent credible interval for an unknown E_1 is that the posterior probability that E_1 falls inside its 95-percent credible interval is 0.95. Thus, the width of the credible interval summarizes the uncertainty in inferences on E_1 . For the harbor-seals study (Ver Hoef and Frost, 2003) E_1 contains the trend parameters $\{\theta_{1i}\}$ and θ_{1A} , and the marginal distributions of the MCMC samples are given in Figure 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; Spiegelhalter et al., 2004) and an open-source version (OpenBUGS; Thomas, 2004). Another open-source software package for posterior simulation is Just Another Gibbs Sampler (JAGS; Plummer, 2005). 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 can be considerably more tedious to implement and debug. For the 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). 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 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 three covariates, tide, date, and time in our model, but we want to compare a Poisson versus negative binomial distribution for the measurement errors. Then a stepwise approach is not appropriate, however the information-theoretic approach handles it 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 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.

5. Discussion

Hierarchical models address complex problems for which there may be several sources of information and hidden variables (Wikle et al., 1998; Clark, 2005; Ogle et al., 2005). Goals may include estimation of parameters for a process model (Ibanez et al., 2006; Hille Ris Lambers et al., 2005), inference on hidden states (Dupris, 1995; Ver Hoef and Cressie, 1997; deValpine, 2003; Stenseth et al., 2003; Dorazio and Royle, 2003; Clark and Bjornstad, 2004; Gelfand et al., 2004; Clark et al., 2005), quantifying the importance of interactions (Coulson et al., 2001; Cam et al., 2002; Clark et al., 2003, 2004; Mohan et al., 2007; LaDeau and Clark, 2006), and prediction (Beckage and Platt, 2003; Clark et al., 2007). In this section, we begin a discussion of the issues surrounding hierarchical statistical modeling. It is expected that in 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, and Clark and Gelfand, 2006, 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. However, there is 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. However, it is not always clear what "subjective" means in this context. 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! 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., Brooks and Roberts, 1998; Cowels and Carlin, 1995; Gelman et al., 2004). Software for assessing convergence of MCMC algorithms include 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; Smith, 2005).

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 only saving 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, \dots, y_n be independent samples from a normal distribution with unknown mean μ and variance σ^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 μ and σ^2 , respectively. Instead of parameterizing the sampling distribution in terms of μ and σ^2 , assume that the data were generated from a normal distribution with unknown mean $\mu_1 + \mu_2$ and variance σ^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 μ_1 and μ_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 μ_1 , μ_2 , and σ^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 μ_1 and μ_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). When running MCMC algorithms, the sample paths of nonidentifiable parameters will “trade off” their values, leading to numerical and convergence problems. If identifiability problems go undiagnosed, inferences on these model parameters and possibly others can be misleading.

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^{rep} , conditional on the observed data, D (see Gelman et al., 2004, Ch. 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 $[E, P_E, P_D | D]$ and plugging it into $[D^{\text{rep}} | E, P_E, P_D]$ to then generate a sample from $[D^{\text{rep}} | D]$.

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 data D_{-i} to predict what should have been observed assuming the model is correct. A Bayesian model allows the posterior distribution $[D_i^{\text{rep}} | D_{-i}]$ 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 $[D_i^{\text{rep}} | D_{-i}]$, 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 $[D^{\text{rep}} | D]$ 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, 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 $[D^{\text{rep}} | D]$ to average $(\text{Observed}_{ij} - \text{Expected}_{ij})^2 / \text{Expected}_{ij}$, over j , for each site i , which should be near one. They found lack of fit for models without a random effect to account for overdispersion.

Bayesian inference in space and time

In a spatio-temporal setting, process and parameter models are less influential on smoothing and filtering, because the data redeem us from misspecification. But they 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, Sec. 6.2) are available for simple cases, but a dynamic, non-stationary, spatio-temporal hierarchical model is far from simple. In contrast, Bayesian inference on the spatio-temporal process accounts for parameter variability coherently.

In this article, we have featured the hierarchical statistical modeling approach for dealing with uncertainty in ecological analysis. The last section is representative of the type of discussion we hope our presentation will engender.

Acknowledgments

This paper resulted from a workshop, "Uncertainty in Ecological Analysis", organized by The Ohio State University's Mathematical Biosciences Institute and Program in Spatial Statistics and Environmental Sciences and held in Columbus, OH from April 3-6, 2006. Major financial support for the workshop came from the National Science Foundation under Agreement No. 0112050. The workshop was attended by fifty U.S. and international quantitative ecologists and applied statisticians, as well as about twenty Ohio State faculty, postdocs, and graduate students. Presentations were relatively short and, in almost every session, as much time was devoted to discussion as to talks. Session themes were almost all ecological, the idea being that science problems would drive the workshop, and 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 on the MBI's website (<http://www.mbi.osu.edu/2005/ws5abstracts.html>).

References

- Akaike, H. 1973. Information theory and the extension of the maximum likelihood principle. In *Proceedings of the International Symposium on Information Theory* (Edited by B. N. Petrov and F. Czaki). Akademia Kiado, Budapest, pp. 267-281.
- Arbia, G. and Lafratta, G. 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., Carlin, B.P., and Gelfand, A.E. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC Press, Boca Raton.
- Beckage, B. and Platt, W.J. 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.
- Berliner, L.M. 1996. Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods* (Edited by K. Hanson and R. Silver). Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 15-22.
- Best, N.G., Cowles, M.K., and Vines, S.K. 1995. *CODA Manual version 0.30*. MRC Biostatistics Unit, Cambridge, UK.
<http://www.mrc-bsu.cam.ac.uk/bugs/documentation/coda03/cdaman03.html>.
- Bjornstad, O.N. and Grenfeld, B.F. 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., Janak, J.F., and Wallis, J.R. 1972. Some ecological consequences of a computer model of forest growth. *Journal of Ecology* **60**: 849 – 872.
- Boveng, P.L., Bengtson, J.L., Cesarone, J.C., Simpkins, M.A., Frost, K.J., and Burns, J.J. 2003. The abundance of harbor seals in the Gulf of Alaska. *Marine Mammal Science* **19**: 111 – 127.
- Brooks, S.P. and Roberts, G.O. 1998. Assessing convergence of Markov Chain Monte Carlo algorithms. *Statistics and Computing* **8**: 319-335.
- Burnham, K.P. and Anderson, D.R. 2002. *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach, 2nd Edition*. Springer Verlag, New York.
- Hoover-Miller, A., Parker, K.R., and Burns, J.J. 2002. 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.
- Calder, C.A., Lavine, M., Mueller, P., and Clark, J.S. 2003. Incorporating multiple sources of stochasticity in population dynamic models. *Ecology* **84**: 1395-1402.
- Cam, E., Link, W.A., Cooch, E.G., Monnat, J.-Y., and Danchin, E. 2002. Individual covariation in life-history traits: Seeing the trees despite the forest. *American Naturalist* **159**: 96-105.
- Carlin, B.P. and Louis, T.A. 2000. *Bayes and Empirical Bayes Methods for Data Analysis, 2nd Edition*. Chapman and Hall/CRC, Boca Raton.
- Casella, G., and George, E.I. 1992. Explaining the Gibbs sampler. *The American Statistician* **46**: 167-174.

- Catchpole, E.A., Hatton, E.A., and Catchpole, W.R. 1989. Fire spread through nonhomogeneous fuel modeling as a Markov process. *Ecological Modelling* **48**: 101–112.
- Chao, C.-T. and Thompson, S.K. 2001. Optimal adaptive selection of sampling sites. *Environmetrics* **12**: 517-538.
- Chen, M.-H., Shao, Q.-M., and Ibrahim, J.G. 2000. *Monte Carlo Methods in Bayesian Computation*. Springer Verlag, New York.
- Chib, S., and Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**: 327-335.
- Christensen, R. 1991. *Linear Models for Multivariate, Time Series, and Spatial Data*. Springer-Verlag, New York.
- Clark, J.S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* **8**: 214.
- Clark, J.S. and Bjornstad, O. 2004. Population time series: Process variability, observation errors, missing values, lags, and hidden states. *Ecology* **85**: 3140-3150.
- Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dodson, A., Foley, J., Lodge, D., Pascual, M., Pielke, Jr., R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D., and Wear, D. 2001. Ecological forecasts: An emerging imperative. *Science* **293**: 657-660.
- Clark, J.S., Ferraz, G., Oguge, N., Hayes, H., and DiCostanzo, J. 2005. Hierarchical Bayes for structured and variable populations: From capture-recapture data to life-history prediction. *Ecology* **86**: 2232-2244.
- Clark, J.S. and Gelfand, A.E. 2006. A future for models and data in ecology. *Trends in Ecology and Evolution* **21**: 375-380.
- Clark, J.S., LaDeau, S., and Ibanez, I. 2004. Fecundity of trees and the colonization-competition hypothesis. *Ecological Monographs* **74**: 415-442.
- Clark, J.S., Mohan, J., Dietze, M., and Ibanez, I. 2003. Coexistence: How to identify trophic tradeoffs. *Ecology* **84**: 17-31.
- Clark, J.S., Wolfsin, M., Dietze, M., Ibanez, I., LaDeau, S., Welsh, M., and Koepfel, B. 2007. Tree growth inference and prediction from diameter censuses and ring widths. In review.
- Cochran, W.G. 1977. *Sampling Techniques, 3rd Edition*. Wiley, New York.
- Coulson, T., Catchpole, E.A., Albon, S.D., Morgan, B.J.T., Pemberton, J.M., Clutton-Brock, T.H., Crawley, M.J., and Genfell, B.T. 2001. Age, sex, density, winter weather and population crashes in Soay sheep. *Science* **292**: 1528-1531.
- Cowles, M.K. and Carlin, B.P. 1995. Markov chain Monte Carlo diagnostics: A comparative review. *Journal of the American Statistical Association* **91**: 883-904.
- Cressie, N. 1993. *Statistics for Spatial Data, rev. edition*. Wiley, New York.
- Cressie, N. 1998. Transect-spacing design of ice cores on the Antarctic continent. *Canadian Journal of Statistics* **26**: 405-418.
- da Silva, J. and Terhune, J.M. 1988. Harbour seal grouping as an anti-predator strategy. *Animal Behavior* **36**: 1309 – 1316.
- Daszak, P., Cunningham, A.A., and Hyatt, A.D. 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 Royle, J.A. 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 Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Federov, V.V. and Nachtsheim, C. 1995. Optimal designs for time-dependent responses. In *Proceedings of MODA4* (Edited by C. P. Kitsos, and W.G. Muller). Physica-Verlag, Heidelberg.
- Frost, K.J., Lowry, L.F., and Ver Hoef, J.M. 1999. Monitoring the trend of harbor seals in Prince William Sound, Alaska, after the *Exxon Valdez* oil spill. *Marine Mammal Science* **15**: 494 – 506.
- Gelfand, A.E. and Sahu, S. 1999. Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association* **94**: 247-253.
- Gelfand, A.E., Schmidt, A.M., Wu, S., Silander, J.A.J., Latimer, A., and Rebelo, A.G. 2004. Explaining species diversity through species level hierarchical modeling. *Applied Statistics* **65**: 1-20.
- Gelfand, A.E., and Smith, A.F.M. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**:398-409.
- Gelfand, A.E., Zhu, I., and Carlin, B.P. 2001. On the change of support problem for spatio-temporal data. *Biostatistics* **2**: 31-45.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2004. *Bayesian Data Analysis, 2nd Edition*. Chapman & Hall/CRC, New York.
- Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York.
- Hardin, J.W. and Hilbe, J.M. 2003. *Generalized Estimating Equations*. Chapman and Hall/CRC, Boca Raton, Florida.
- Hille Ris Lambers, J., Clark, J.S., and Lavine, M. 2005. Seed banking in temperate forests: Implications for recruitment limitation. *Ecology* **86**: 85-95.
- Hoover-Miller A., Parker, K.R., and Burns, J.J. 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.
- Ibanez, I., Clark, J.S., Dietze, M.C., Feeley, K., Hersh, M., LaDeau, S., McBride, A., Welch, N.E., and Wolosin, M.S. 2006. Predicting biodiversity change: Outside the climate envelope, beyond the species-area curve. *Ecology* **87**: 1896-1906.
- LaDeau, S.L. and Clark, J.S. 2006. Elevated CO₂ and tree fecundity: The role of tree size, interannual variability, and population heterogeneity. *Global Change Biology* **12**: 822-833.
- Le, N.D. and Zidek, J.V. 1994. Network designs for monitoring multivariate random spatial fields. In *Recent Advances in Statistics and Probability* (Edited by J.P. Vilaplana and M.L. Puri). VSP, Leiden, Netherlands, pp. 191-206.
- Link, W.A., Cam, E., Nichols, J.D., and Cooch, E.G. (2002). Of BUGS and birds: Markov chain Monte Carlo for hierarchical modeling in wildlife research. *Journal of Wildlife Management* **66**: 277-291.
- Linhart, H. and Zucchini, W. 1986. *Model Selection*. Wiley, New York.

- Martin, A.D., and Quinn, K.M. 2004. *MCMCpack: Markov chain Monte Carlo (MCMC) Package*. <http://mcmcpack.wustl.edu>. R package version 0.6-3.
- McQuarrie, A. D., and Tsai, C.-L. 1998. *Regression and Time Series Model Selection*. World Scientific, Hackensack, New Jersey.
- Medawar, P.B. 1940. Growth, growth energy, and ageing of the chicken's heart. *Proceedings of the Royal Society of London* **129**: 332 – 355.
- Mohan, J.E., Clark, J.S., and Schlesinger, W.H. 2007. Long-term CO₂ enrichment of an intact forest ecosystem: Implications for temperate forest regeneration and succession. *Ecological Applications*, in press.
- Mugglin, A.S., Carlin, B.P., and Gelfand, A.E. 2000. Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association* **95**: 877-887.
- Ogle, K., Wolpert, R.L., and Reynolds, J.F. 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. In *Environmental Modeling with GIS* (Edited by M.R. Goodchild, B.O. Parks, and C.T. Steyaert). Oxford University Press, New York, pp. 470-480.
- Plummer, M. *JAGS: Just Another Gibbs Sampler*. <http://www-fis.iarc.fr/~martyn/software/jags/>.
- Plummer, M., Best, N., Cowles, K., and Vines, K. 2005. *CODA: Output Analysis and Diagnostics for MCMC*. <http://www-fis.iarc.fr/coda/>. R package version 0.9-2.
- Rao, J.N.K. 2003. *Small Area Estimation*. Wiley, New York.
- Ricker, W.E. 1954. Stock and recruitment. *Journal of Fisheries Research Board of Canada* **11**: 559 – 623.
- Roberts, C.P. and Casella, G. 2005. *Monte Carlo Statistical Methods, 2nd Edition*. Springer, New York.
- Sacks, W., Schimel, D., and Monson, R. 2007. Coupling between carbon cycling and climate in a high-elevation, subalpine forest: a model-data fusion analysis. *Oecologia*, submitted.
- Särndal, C., Swensson, B., and Wretman, J. 1992. *Model Assisted Survey Sampling*. Springer-Verlag, New York.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.
- Small, R.J., Pendleton, G.W., and Pitcher, K.W. 2003. Trends in abundance of Alaska harbor seals, 1983-2002. *Marine Mammal Science* **19**: 344 – 362.
- Smith, B.J. 2005. *Bayesian Output Analysis Program (BOA), Version 1.1.5*. The University of Iowa. <http://www.public-health.uiowa.edu/boa>.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society. Series B* **64**: 583–639.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. 2004. *WinBUGS*. <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/>.
- Stenseth, N.C., Viljugrein, H., Saitoh, T., Hansen, T.T., Kittilsen, M.O., Bolviken, E., and Glockner, F. 2003. Seasonality, density dependence, and population cycles in Hokkaido voles. *Proceedings of the National Academy of Sciences* **100**: 11478-11483.
- Stern, H.S. and Cressie, N. 2000. Posterior predictive model checks for disease mapping models. *Statistics in Medicine* **19**: 2377-2397.
- Thomas, A. 2004. *OpenBUGS*. <http://mathstat.helsinki.fi/openbugs/>.
- Thompson, S.K. and Seber, A.F., 1996. *Adaptive Sampling*. Wiley, New York.

- 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. and Cressie, N. 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 Frost, K. 2003. A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska. *Environmental and Ecological Statistics* **10**: 201 – 209.
- Watts, P. 1992. Thermal constraints on hauling out by harbor seals (*Phoca vitulina*). *Canadian Journal of Zoology* **70**: 553 – 560.
- Wikle, C.K. 2003a. Hierarchical models in environmental science. *International Statistical Review* **71**: 181-199.
- Wikle, C.K. 2003b. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **84**: 1382-1394.
- Wikle, C.K., Berliner, L.M., and Cressie, N. 1998. Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* **5**: 117-154.
- Wikle, C.K. and Royle, J.A. 1999. Space-time models and dynamic design of environmental monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics* **4**: 489-507.
- Wikle, C.K., Milliff, R.F., Nychka, D., and Berliner, L.M. 2001. Spatio-temporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association* **96**: 382-397.
- Wikle, C.K. and Berliner, L.M. 2005. Combining information across spatial scales. *Technometrics* **47**: 80-91.
- Wikle, C.K. and Royle, J.A. 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.

Figure Captions

Figure 1: Map of Prince William Sound, Alaska, 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.]

Figure 2: A schematic of the hierarchical statistical models for the harbor-seals study. (A): The data model is the top level with the process model at the second level. A *Bayesian* hierarchical statistical model would have a parameter model at a third level. (B): As in (A), except the linear model parameters from each site are given a distribution. Here, a *Bayesian* hierarchical statistical model would have a parameter model at the fourth level.

Figure 3: Posterior densities of slopes $\{\theta_i : i = 1, \dots, 25\}$, corresponding to the year (trend) covariate, with the posterior density of θ_{1A} superimposed with the heavy solid line. These posterior densities correspond to model in Figure 2(B). The solid circles on the x-axis correspond to the peak densities for the model in Figure 2(A). 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.

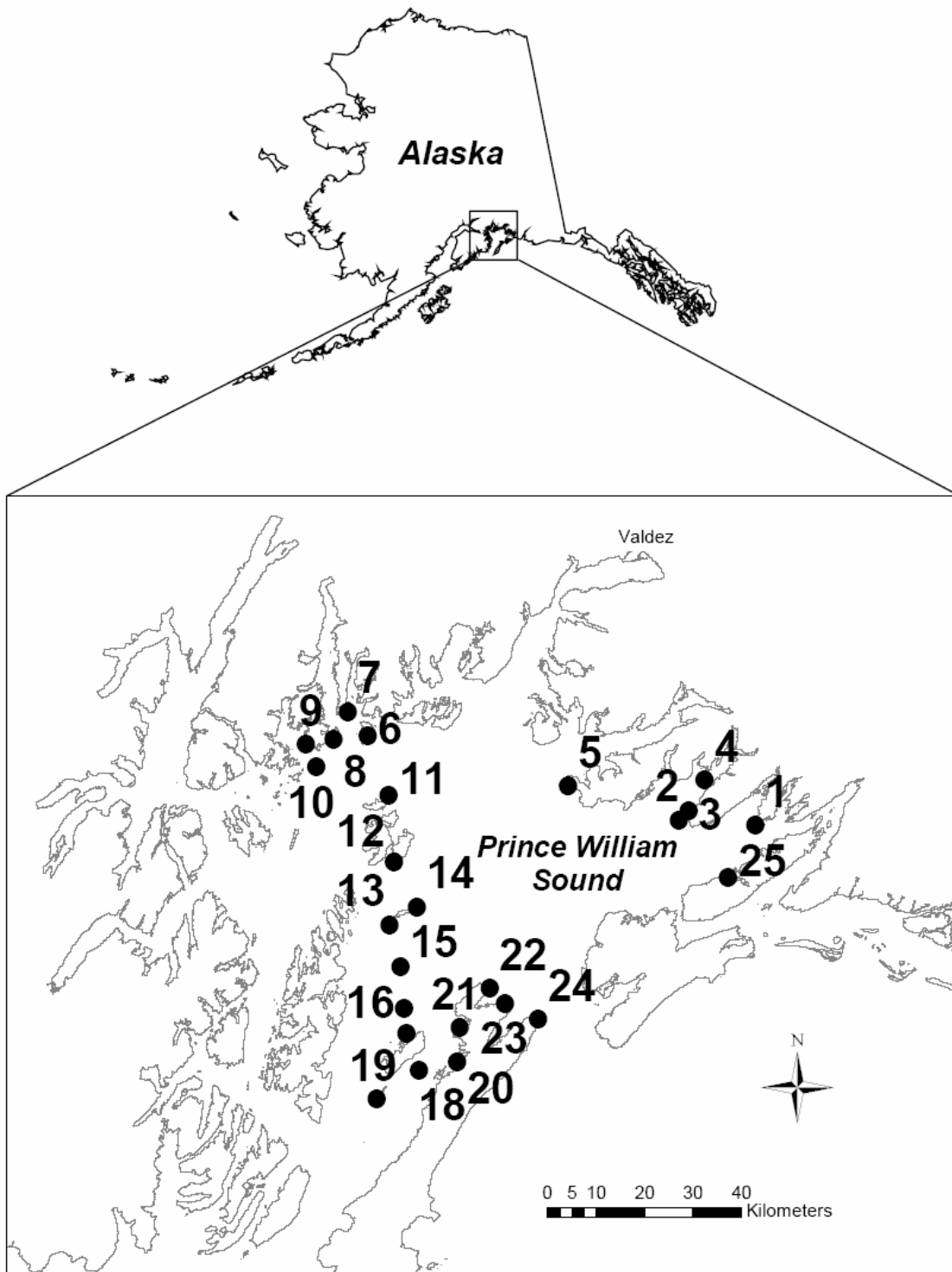


Figure 1

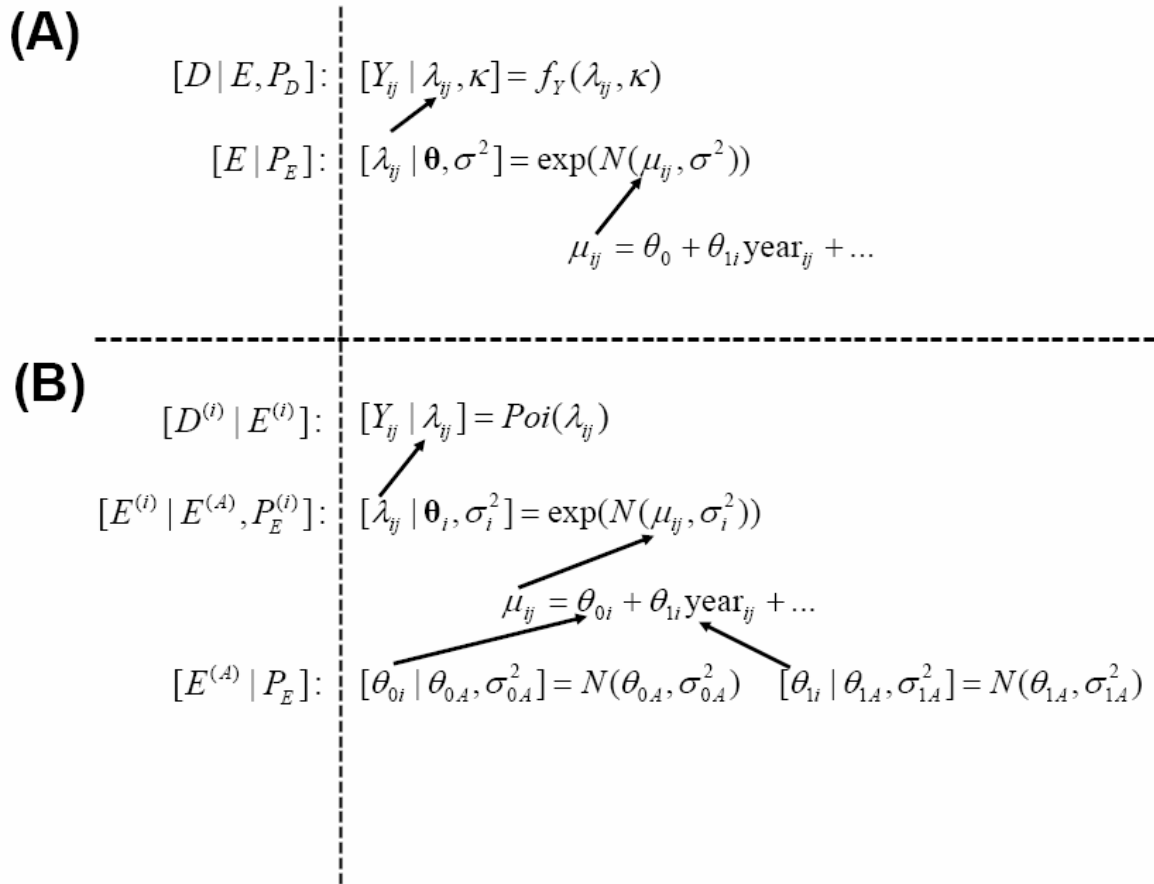


Figure 2

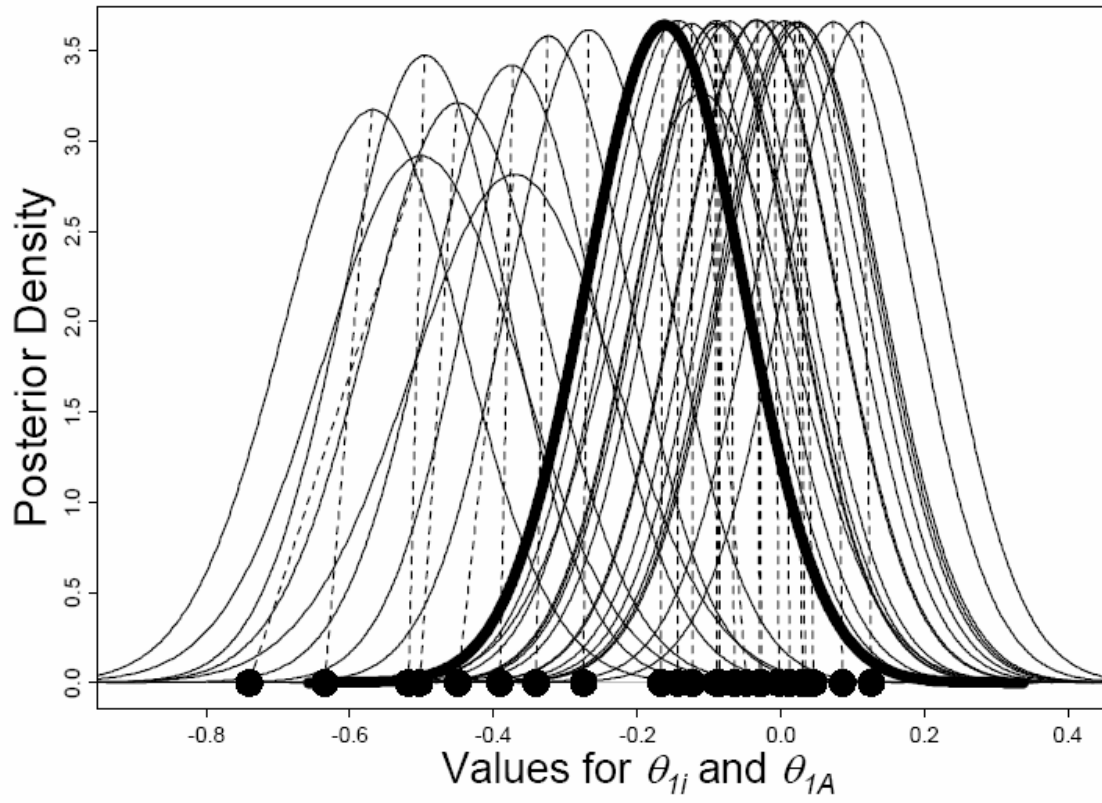


Figure 3