

1 have also planted some seeds for discussion and tried to show where the practical difficulties lie.
2 Our thesis is that hierarchical statistical modeling is a powerful way of approaching ecological
3 analysis in the presence of inevitable but quantifiable uncertainties, even if practical issues
4 sometimes require pragmatic compromises.

5

6 **Key words:** Bayesian modeling; data model; design; empirical Bayes; harbor seals;
7 MCMC; prior; process model; spatial process; spatio-temporal process

1 **1. Introduction**

2

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

10

11 Ecological data are almost always observed incompletely with large and unknown amounts of
12 measurement error or *data uncertainty*, and often the expense of data collection prohibits
13 collecting as much data as might be desirable. How much and where to sample are important
14 design questions (e.g., Stevens and Olsen, 2004). In addition, most ecological phenomena of
15 interest can only be studied by combining various sources of data; aligning these data properly
16 presents interesting statistical challenges. While data play a large role in most ecological
17 analyses, incorporating scientific knowledge through substantive modeling of ecological
18 processes is essential. Often such process modeling is based on competing scientific theories
19 and simplifications of reality. This results in an additional source of uncertainty, termed *model*
20 *or process uncertainty*. Furthermore, substantive models should acknowledge *parameter*
21 *uncertainty*. Parameter uncertainty can be handled either by estimating the unknown parameters
22 (empirical-Bayesian analysis) or by expressing that uncertainty via a prior probability
23 distribution (Bayesian analysis); see, for example, Ver Hoef (1996), Carlin and Louis (2000),

1 and Gelman and Hill (2006), where the two approaches are presented. An empirical-Bayesian
2 analysis looks for plug-in estimates and may avoid more realistic and flexible specifications that
3 can include variation over space and time. The Bayesian analysis can use such variation to help
4 with the choice of the prior distribution.

5
6 In April 2006, a workshop on “Uncertainty in Ecological Analysis” was held at the Mathematical
7 Biosciences Institute (MBI), The Ohio State University. The workshop organizers, who are the
8 authors of this paper, believe that the workshop demonstrated how hierarchical statistical
9 modeling is a powerful approach for dealing with uncertainty in ecology. As will be explicated in
10 the sections that follow, it is a statistical methodology for handling complex (ecological)
11 problems by building a hierarchy of statistical models: Broadly, the first level of the hierarchy is
12 a data model, the second is a process model, and the third (optional) level is a parameter model.
13 It also provides a framework for how a team of scientists might work together, and it partitions
14 variability/uncertainty in a way that can suggest study designs where uncertainty can be
15 controlled. Both the Bayesian and non-Bayesian (i.e., empirical-Bayesian) versions of the
16 hierarchical statistical model are presented, reflecting the diversity of workshop participants’
17 views.

18
19 This paper does not introduce new methodology, but it is meant to reach, and engender
20 discussion from, readers of this journal. To illustrate the themes of the article, we use an
21 ecological study of harbor seals and their abundance at haul-out sites in Prince William Sound,
22 Alaska, as reported by Ver Hoef and Frost (2003). Harbor seals haul out to rest (da Silva and
23 Terhune, 1988), molt (Boily, 1995), and escape predators (Watts, 1992), among other reasons.

1 They are monitored for trend (e.g., Ver Hoef and Frost, 2003) and abundance (e.g., Boveng et
2 al., 2003) in northern latitudes around the world, and they are of particular interest in this study
3 because of potential long-term impacts from the Exxon Valdez oil spill in 1989.

4
5 The organization of the paper is as follows. Section 2 addresses the general notion of modeling
6 in the presence of uncertainty and features hierarchical statistical modeling. Section 3 addresses
7 how hierarchical modeling can contribute to experimental design and data collection. Section 4
8 discusses statistical inference in ecological analyses, including treatment of computational
9 aspects and model choice. The paper finishes with a discussion of challenges for hierarchical
10 statistical modeling, presented in Section 5.

11 12 **2. Modeling in the presence of uncertainty**

13
14 Why does one build models? Is it simply to organize information? Does the model depend on
15 the goal, whether that is estimation, prediction, forecasting, explanation, or simplification?

16 Ultimately, modeling is about the synthesis of information, whether that comes from
17 observations, or from the collective wisdom of a team of scientists, or more broadly from diverse
18 corners of the relevant literature. We present a basic framework here that accommodates the
19 synthesis of information in a coherent manner. The key to this framework is thinking
20 conditionally, something we argue comes naturally to scientists. For example, we observe D
21 conditional upon E , E happens conditional upon P , and so forth. Not only is this true in an
22 observational context, but it is appropriate when we think about processes as well. Think of E as
23 an ecological process; then E might behave in one way conditioned on one set of environmental

1 conditions and in another way under a different set of conditions. The conditional behavior is
2 endemic to the process' behavior, regardless of whether observations are taken or not. Formally,
3 we can link such thinking within the context of conditional probabilities.

4

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

21

22 An alternative to this “curve-fitting” approach is formal *statistical modeling*. Here, we may wish
23 to specify a probability distribution for D that depends on some parameters, say Θ . Thus, we

1 might think about estimating the parameters of this distribution, $[D | \Theta]$. That is, can we find an
2 estimate of the parameters Θ that maximizes the likelihood (defined by the distributional
3 assumption made about D) of observing the data D ? Note that we are using the brackets "[]" to
4 refer to a distribution and the vertical bar "|" is read as "conditioned upon". So, $[A | B]$ would be
5 read as "the distribution of A conditioned upon B ", or "the distribution of A given B ". In this
6 framework, the parameters Θ are assumed to summarize the ecological process E appropriately,
7 and it is assumed that there is no uncertainty in this summarization. Although greatly simplified,
8 this is the idea behind much of classical statistical inference in science. In this setting, the
9 observations on the process of interest are directly modeled without explicit reference to a
10 statistical model for the process E . The focus here is on a data model; it is assumed that the
11 uncertainty lies with the data and is due to sampling and measurement.

12
13 Alternatively, we may sometimes be interested in developing process models for E directly.
14 Often these models are deterministic. However, our knowledge of E is always limited in some
15 fashion, which suggests that randomness should also play a role in process modeling. Indeed,
16 such thinking has long been crucial to ecological modeling, where it has been incorporated into
17 the science (e.g., in the fire-simulation models of Catchpole et al., 1989, or the forest-simulation
18 models of Botkin et al., 1972). These random processes can, in turn, be characterized by
19 distributions that have associated parameters, say P_E . That is, we know the distribution of $[E |$
20 $P_E]$. If we do not distinguish between *observations on E* (which are called D in this paper) and
21 their true values, then the usual likelihood analysis ignores uncertainties about the relationship
22 between E and D , resulting in incorrect statistical inferences.

23

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

10

11 The introduction given above to hierarchical statistical modeling is quite general, motivated by
12 generic problems of curve fitting or population growth in a random environment. For the rest of
13 the paper we shall use a specific ecological study to illustrate the strengths and limitations of
14 hierarchical statistical modeling . Consider the harbor-seals censuses from Ver Hoef and Frost
15 (2003) in Prince William Sound in Alaska, which were collected as part of routine monitoring
16 following the Exxon Valdez oil spill in 1989. These data were obtained by counting seals at a
17 set of haul-out sites photographed during aerial surveys. Surveys were flown repeatedly for 7-10
18 days in August or September each year from 1990-1999. A map showing the geographical
19 context and the sampling plan (sites were surveyed in either the order of the sites' identifying
20 indices or in the reverse order) is given in Figure 1.

21

22

Figure 1 here

23

1 We now establish notation and a hierarchical statistical model for the harbor-seals study. Let Y_{ij}
2 denote the j^{th} count for the i^{th} site, $i = 1, 2, \dots, n$. In terms of our previous discussion, all of the
3 counts for all surveys and all sites form a set $D = \{Y_{ij}\}$. A Poisson (e.g., Ver Hoef and Frost,
4 2003) or negative binomial (e.g., Boveng et al., 2003) is a natural probability distribution to
5 assume for count data. We can imagine that for any given site at any given time, there is a true
6 number of harbor seals that “should” haul out, but we are not able to observe that number. That
7 is, harbor seals are constantly sliding in and out of the water so that at any given time the actual
8 number that hauls out is more or less than expected. Let this expected number be λ_{ij} , so we
9 might consider $Y_{ij} \sim f(\lambda_{ij}, \kappa)$, where f is the “Poisson” probability distribution or the “negative
10 binomial” probability distribution with mean parameter λ_{ij} and variance parameter κ (in the case
11 of the negative binomial). Write $E = \{\lambda_{ij}\}$ and $P_D = \{\kappa\}$, and hence the joint distribution of all
12 the data is $[D | E, P_D]$, where this data model is conditional on the ecological process E and a
13 parameter κ expressing variability in the data.

14

15 This model has very little “ecology” in it and it would be very difficult to carry out inference
16 because we have one parameter (λ_{ij}) per datum (Y_{ij}). However, we know more about harbor
17 seals; for example, we know that counts are highest around low tide and midday, and we know
18 that they change seasonally (date, within year) and there may be temporal trend (year). These
19 factors reflect our understanding of harbor-seal biology, such as substrate availability and
20 physiology, and we would like to include them in the model. If we were omniscient, we would
21 have a perfect deterministic model for the “expected” harbor-seal abundance, without error, on
22 every haul-out site for every instant of time. Lacking that, we use a “smooth curve”, which is a
23 function of a set of unknown parameters and the available information about the process (i.e.,

1 year, date, tide level, and time of day), to describe the ecological process that controls haul-out
 2 abundance. Since we acknowledge that this curve does not accurately reflect the complexity of
 3 this process, we account for errors from our ecological process model by introducing an
 4 additional source of randomness. This can be achieved by letting $\log(\lambda_{ij})$ be distributed according
 5 to a normal (Gaussian) distribution with mean $\mu_{ij} \equiv \mu_i(\text{year}_{ij}, \text{date}_{ij}, \text{tide}_{ij}, \text{time}_{ij}; \theta_i)$ and variance
 6 σ^2 for the i^{th} site, where we assume that $\{\log(\lambda_{ij})\}$ are conditionally independent given $\{\mu_i\}$ and
 7 σ^2 . For example, the smooth curve could be modeled as,

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

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

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

14 where $\{\varepsilon_{ij}\}$ are independent and identically distributed normal random variables with $E(\varepsilon_{ij}) = 0$
 15 and $\text{var}(\varepsilon_{ij}) = \sigma^2$. In terms of our previous discussion, recall $E = \{\lambda_{ij}\}$ and write $P_E = \{\theta, \sigma^2\}$,
 16 where $\theta = (\theta_1, \dots, \theta_n)$. Hence, we can obtain the conditional distribution of E , $[E | P_E]$,
 17 conditional on the parameters P_E .

18 In the conditional modeling described above, we have accounted for the uncertainty in the
 19 process E by modeling $[E | P_E]$. In the data-generation mechanism, we have accounted for the

1 uncertainty by modeling $[D | E, P_D]$ (for the harbor-seals data, this is $[\{Y_{ij}\} | \{\lambda_{ij}\}]$ for a Poisson
2 data model, or $[\{Y_{ij}\} | \{\lambda_{ij}\}, \kappa]$ for a negative-binomial data model).

3
4 Returning to our general discussion of hierarchical statistical modeling, we can combine the data
5 model, $[D | E, P_D]$, with the process model, $[E | P_E]$, to obtain the joint uncertainties in the data D
6 and the ecological process E , as follows:

7
8
$$[D, E | P_D, P_E] = [D | E, P_D][E | P_E], \tag{1}$$

9
10 which is a result from probability theory that shows how the joint uncertainties can be expressed
11 hierarchically using conditional probabilities. We note that this is the essence of Gaussian,
12 linear, state-space models for temporal processes, that result in Kalman filtering for E at a
13 current time based on data at current and past times (e.g., Meinhold and Singpurwalla, 1983;
14 West and Harrison, 1997). In terms of the harbor-seals study, (1) says that we obtain a joint
15 model of the data and the underlying ecological process, $[\{Y_{ij}\}, \{\lambda_{ij}\} | \kappa, \theta, \sigma^2]$, now conditional
16 on fewer parameters. Figure 2(A) shows this hierarchical model schematically, with conditional
17 dependencies shown by arrows linking the relevant boxes. The first level is the data model, $[D |$
18 $E, P_D] = [\{Y_{ij}\} | \{\lambda_{ij}\}, \kappa]$, and the second level is the process model, $[E | P_E] = [\{\lambda_{ij}\} | \theta, \sigma^2]$. From
19 (1), the joint model of the data and the underlying ecological process is simply the *product* of the
20 data model and the process model. (The more complicated hierarchical model shown in Figure
21 2(B) will be discussed later in this section.)

22
23

 Figure 2 here

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

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 (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 $P \equiv \{P_D, P_E\}$. Hierarchical statistical modeling simply means that we decompose the joint distribution into a probabilistically valid series of conditional models: $[D, E | P] = [D | E, P][E | P]$. 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. Of course, getting the causal mechanism right means that the results from the hierarchical statistical analysis are scientifically meaningful and interpretable.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

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:

$$[data, process, parameters] = [data | process, parameters] \times [process | parameters] \times [parameters] ,$$

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:

$$[data, process | parameters] = [data | process, parameters] \times [process | parameters] ,$$

which is equation (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 equation (1).

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

$$[D, E, P] = [D | E, P][E | P][P] = [D | E, P_D][E | P_E][P_D, P_E] , \tag{2}$$

1
2 where we have used the obvious relationships, $[D | E, P] = [D | E, P_D]$ and $[E | P] = [E | P_E]$ (e.g.,
3 Wikle, 2003a). The utility of such a decomposition is that it allows us to account formally for
4 uncertainty within each stage, where the stages are linked in a probabilistically consistent
5 fashion, resulting in a Bayesian analysis (e.g., Link et al., 2002).

6
7 Both types of analyses, empirical-Bayesian and Bayesian, result in inferences on process values
8 (E) and parameter values (P). When making inference on E , an empirical-Bayesian analysis
9 substitutes in an estimate \hat{P} of P into whatever summary of $[E | D, P]$ was chosen for inference.
10 Typically, the variability of the estimator \hat{P} is not accounted for, leading to credible intervals for
11 values of E that are too liberal. One could try to account for the variability in \hat{P} , but formal
12 inference is usually approximate (e.g., Morris, 1983a, 1983b); one commonly used
13 approximation is based on multivariable differential calculus and Taylor-series expansions of
14 estimators (e.g., Prasad and Rao, 1990; Cressie, 1992; Rao, 2003, Sec. 6.2). In contrast, the
15 Bayesian analysis bases its inferences directly on $[E | D]$ and, provided it can be computed, the
16 resulting credible intervals are accurate. Comparisons of this sort are made by Carlin and Louis
17 (2000).

18
19 Hierarchical statistical modeling leads to computationally intensive inference, regardless of
20 whether the empirical Bayesian or the Bayesian analysis is chosen. These are discussed in
21 Section 4, including the use of Markov chain Monte Carlo and bootstrap procedures to account
22 for the variability in \hat{P} .

23

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

11
 12 The real power of this approach for complicated problems comes from the fact that each of the
 13 component distributions in (1) or (2) can be decomposed further, if necessary, and they may be
 14 simplified with modeling assumptions. For example, say we are interested in a process E for
 15 which we have several different data sets, $D^{(1)}, D^{(2)}, D^{(3)}$, all of which measure the process E with
 16 uncertainty and perhaps at different spatial or temporal scales. Then, it is often possible in such
 17 cases to make the following modeling assumption:

18
 19
$$[D^{(1)}, D^{(2)}, D^{(3)} | E, P_D] = [D^{(1)} | E, P_D^{(1)}][D^{(2)} | E, P_D^{(2)}][D^{(3)} | E, P_D^{(3)}],$$

20
 21 where $P_D = \{P_D^{(1)}, P_D^{(2)}, P_D^{(3)}\}$. Such an assumption is appropriate in the harbor-seals study for
 22 repeat counts of the same set of sites on successive days. That is, we might assume that the
 23 different datasets are independent, conditional upon the true process. Although such an

1 assumption must be justified, it is often plausible and provides a very convenient approach for
 2 synthesizing various types of observations. For the harbor-seals study, a further assumption,
 3 $P_D^{(1)} = P_D^{(2)} = P_D^{(3)} \equiv P_D$, was made; that is, it was assumed that each successive daily count
 4 follows the same data model. The parameters in each of the component distributions can
 5 accommodate changes of resolution and alignment, as well as different measurement-error
 6 characteristics (e.g., Wikle et al., 1998; Mugglin et al., 2000; Gelfand et al., 2001; Wikle et al.,
 7 2001; Gotway and Young, 2002; Banerjee et al., 2004; Clark et al., 2004; Wikle and Berliner,
 8 2005; LaDeau and Clark, 2006; Hille Ris Lambers et al., 2006).

9
 10 Decomposition can also be considered for the probability distribution of the ecological process.
 11 For example, consider a process E made up of two subprocesses, $E^{(1)}$ and $E^{(2)}$. We can often
 12 make use of conditional modeling in this context as well:

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

13
 14
 15
 16 For example, one could build hierarchical models for multiple species to examine species
 17 diversity and distribution patterns (Gelfand et al., 2005a, 2005b; Latimer et al., 2006). In terms
 18 of the harbor-seals study, the ideas of multiple datasets and multiple processes were used by Ver
 19 Hoef and Frost (2003), yielding a model,

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

1 That is, a joint model for the data and the linear-regression parameters was built conditionally.

2 Here, $[D^{(i)} | E^{(i)}] = [Y_{ij} | \lambda_{ij}]$, is the Poisson-regression formulation at the i^{th} site, given earlier.

3 Assume,

4

$$5 \quad [E^{(1)}, \dots, E^{(n)} | E^{(A)}, P_E] = [\{\lambda_{1j}\} | \theta_1, \sigma_1^2] \dots [\{\lambda_{nj}\} | \theta_n, \sigma_n^2],$$

6

7 where there are now potentially n different variances $\{\sigma_i\}$, one for each site. Furthermore,

8 suppose that all intercepts across sites have a common distribution with parameters associated

9 with the area A , all slope parameters across sites have a common distribution ..., and so forth.

10 Let $E^{(A)}$ denote the set of all these parameters associated with the area A . This model is shown

11 in Figure 2(B), which extends the model shown in Figure 2(A). For example, let the intercept

12 mean parameter be normally distributed. This and specifications for the slope parameters and

13 the variance parameters determine $[E^{(A)} | P_E]$. This makes clear one of the real strengths of

14 hierarchical modeling. Without it, we might model each site separately but then lack the ability

15 to say anything about all sites within the area, or we might ignore sites and model all data with a

16 single Poisson regression. The hierarchical model allows inference for $E^{(i)}$ at the i^{th} site or

17 for $E^{(A)}$ over the area that contains all sites, through their respective posterior distributions.

18

19 Figure 3 shows the posterior density of the individual slopes $\{\theta_{1i} : i = 1, \dots, n\}$ for year

20 (temporal trend) in the linear model for μ_i given above (Figure 2(B)), along with the posterior

21 density of the slope θ_{1A} , where recall that $\theta_{11}, \dots, \theta_{1n}$ are conditionally distributed with mean

22 θ_{1A} . Consider the mean of each posterior density, and notice the distribution of the individual

23 sites' mean slopes around the areal mean slope. The mean of each posterior density, considering

1 each site separately (Figure 2(A)), is shown with solid circles along the x-axis of Figure 3. The
2 one-at-a-time site means are linked by the dashed line to their counterparts under the conditional
3 model with mean $\theta_{1,A}$. It is generally true that one-at-a-time estimates will “shrink” towards the
4 global mean under the hierarchical model; the extent of shrinkage depends on the (estimated)
5 variability within site versus among sites. From Figure 3, the majority of the slopes’ means are
6 negative, indicating a decline in harbor-seal abundance over years. We also see that the mean of
7 $\theta_{1,A}$, the regression coefficient *for the area*, is negative.

8 _____
9 Figure 3 here

10
11 It is often possible to simplify the joint interaction of one component of the process with another,
12 by using a conditional probabilistic relationship. A well known example of this occurs in time
13 series, where a Markov assumption is made. Specifically, if one has the time series, $E_1, E_2, \dots,$
14 E_{T-1}, E_T , it is often very difficult to specify the joint distribution of the entire series. However,
15 under the first-order Markov assumption, the joint distribution is:

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

16
17
18
19 The key to these process decompositions is to model the process in stages that are scientifically
20 plausible. In that way, very complicated joint distributions can be modeled by relatively simple
21 conditional distributions. Often, deterministic process models for E (given P_E) can be
22 reformulated with stochastic components; for example, Wikle (2003b) uses a reaction-diffusion
23 partial differential equation to motivate a stochastic process model for invasive species. The

1 Markov assumption is the basis of the Kalman filter (e.g., West and Harrison, 1997); Meinhold
2 and Singpurwalla (1983) demonstrate that the Kalman filter can be derived from a non-Bayesian
3 hierarchical statistical model. When the individual processes in the series E_1, E_2, \dots , are spatial
4 processes, spatio-temporal hierarchical models can be built that result in a spatio-temporal
5 Kalman filter (Huang and Cressie, 1996; Wikle and Cressie, 1999).

6
7 In this section, we have provided a baseline for the concept of, consistent notation for, and the
8 flexibility of, hierarchical statistical modeling to deal with uncertainty in ecological studies. An
9 important part of the exposition has been to present the role of *both* the ecological process E and
10 the parameters P . The observant reader will see that the distinction between E and P is not
11 always precise, but this is not probabilistically important in a hierarchical statistical model. At
12 the last stage of the hierarchy, there will typically be unknown parameters that could be
13 estimated or whose uncertainty could be captured probabilistically by (prior) statistical
14 distributions.

15

16 **3. Design for data collection in ecological studies**

17

18 Before moving to discussion of statistical inference for hierarchical models (Section 4), in this
19 section we consider the issue of design for data collection from a statistical modeling
20 perspective. This is a topic that deserves more exposure in the ecology literature, where it has
21 hitherto not been connected with hierarchical statistical modeling. Hence, we aim here for a
22 complete presentation, where we highlight the importance of collecting data using sampling
23 designs based on the principles of randomization, stratification, and replication, no matter how

1 the data are analyzed. We show how sampling designs can be incorporated into (hierarchical)
2 statistical models. This should be contrasted with model-assisted, design-based inference (e.g.,
3 Särndal et al., 1992), where a model-based estimator is assessed based on distributions implied
4 by survey-sampling probabilities. We believe that Design is the medium through which team
5 members with different scientific expertise can communicate with each other, and that its
6 principles are founded on the uncertainties that are present in all scientific studies.

7
8 To many people who have taken a course on design in a Statistics program, the topic brings back
9 memories of factorial designs, randomized block designs, partially balanced incomplete block
10 designs, latin-square designs, and so forth. These are relatively dry topics (although variants of
11 latin-square designs have surfaced with the popularity of the game of Sudoku, found in most
12 daily newspapers). In fact, design is a much broader concept, and appearances of it can be found
13 throughout a typical Statistics graduate curriculum, particularly in survey-sampling courses
14 (sampling design) and design-of-experiments courses (experimental design).

15
16 Every sampling design should start with a specification of the population and the subpopulations
17 under study. For the harbor-seals study, the population is the continuous abundance of harbor
18 seals on all haul-out sites for a given study area and time span.

19
20 The main principles of design are *stratification*, *randomization*, and *replication*, which are
21 intended to allow the scientist to get closer to the elusive goal of establishing causation.
22 *Stratification* controls variability by splitting populations into subpopulations that are internally
23 comparable. Strata (or blocks) should be chosen so that important subpopulations are included

1 and so that the subpopulations will span the variability in responses expected. Often strata are
2 chosen according to a combination of (pre-specified) *factors*. Suppose each factor is split up into
3 levels (e.g., two levels, “high” and “low”). A factorial design is where every possible
4 combination of levels appears once, and each combination of levels represents a stratum. By
5 ensuring that the levels of each factor are at least at the “high end” and at the “low end” of
6 possible choices, strata in a factorial design will span the variability in responses expected.

7
8 The goal of the harbor-seals study was to monitor trend. Hence, it would be best to standardize,
9 as much as possible, the timing of counts each year. As indicated earlier, counts are primarily
10 affected by date, tide, and time of day. To see how stratification would work, the time relative to
11 low tide could be divided into three categories: more than 1/2 hour before low tide, 1/2 hour
12 before to 1/2 hour after low tide, and more than 1/2 hour after low tide. Likewise, the time of
13 day could be divided into three categories and the date into three categories. Then, one could
14 decrease variability across years by sampling only within a single cross-classified stratum, for
15 example, low tide in the middle of the day in the middle of August. As is often the case, real-
16 world factors do not allow such a design; low tide varies considerably among days and weeks
17 within each year, and using a single aircraft to observe all sites every day does not allow
18 observers to be everywhere at once. The strategy of finding a constant set of environmental
19 conditions is not possible in this example. The alternative is to embrace the environmental
20 variation and create a design that models variations in counts due to environmental conditions.
21 Stratification was based on cross-classified levels of factors, and as much as possible the
22 numbers of samples in strata were balanced. For example, samples from late in the day, more
23 than 1/2 hour after low tide, and in August, were obtained in about the same number as samples

1 from the stratum with the near-optimal conditions of early morning, low tide, and in early
2 August.

3
4 *Randomization* offers protection against a biased (unintentional or intentional) choice of
5 sampling sites (or of organisms) in the study. If it is at all possible, ecologists should avoid
6 choosing a so-called “representative site”. In the harbor-seals study, Ver Hoef and Frost (2003)
7 analyzed data from 25 sites in Prince William Sound, Alaska, but economical considerations did
8 not permit sampling these 25 sites randomly from the hundreds of possible sites in the study
9 area. It is very expensive to contract aircraft, and there are only a few possible ways to fly
10 economically from site to site. Thus, to avoid spending a lot of time and money flying back and
11 forth to sites selected at random, or even a spatially clustered set of sites in a random order, the
12 harbor-seals study did not use random sampling. As a result, purely design-based statistical
13 inference cannot extend beyond these 25 sites, and how representative they are of a larger region
14 has been criticized by Hoover-Miller et al. (2002).

15
16 A sampling design becomes a *probability sampling design* when there is a known non-zero
17 inclusion probability for every member of the population (Overton, 1993). The sampling
18 weights, which should be given as part of the design, are simply one divided by the inclusion
19 probabilities. Without randomization, the harbor-seals study clearly does not use a probability
20 sampling design.

21
22 *Replication* is a way to decrease (e.g., measurement, spatial, temporal) variability within strata.
23 The general idea is that an average of responses within a stratum has variance proportional to one

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

4

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

16

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

1

2 Sampling design has traditionally eschewed the modeling approach given in Section 2, but while
3 design-based inference can be robust, model-based inference can be very efficient (e.g.,
4 Aldworth and Cressie, 1999), where efficiency is characterized by mean squared prediction
5 error. In this section, we show how traditional sampling design can be included in the
6 hierarchical-statistical-modeling approach.

7

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

20

21 The ecological process E is of interest, and one wishes to design a sampling plan through which
22 knowledge about E can be obtained. For the moment, we make the (unrealistic) assumption that

1 when E or parts of it are observed, they are observed without error. We discard this assumption
2 later in the section.

3
4 In the rest of this section, we write

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

7
8 where A is an index set that describes the complete population of scientific interest; s may or
9 may not index a spatial location. In traditional sampling design, the unknown population is
10 sampled randomly. That is, a probability distribution is put on $\{S : S \subset A\}$, the set of all subsets
11 of A ; then inference on E is based on the stochasticity in the random samples. This is called
12 design-based inference and, in the terminology developed in Section 2, it depends on $[S | E]$.
13 Because of the practical difficulties mentioned for the harbor-seals study, appropriate
14 stratification and random sampling was not carried out. Hence, design-based inference is not
15 possible in the study, and it cannot be used to illustrate the methodology given below.

16
17 Suppose that $S = S^{(1)}$ is observed, and hence $E^{(1)} \equiv \{E_s : s \in S^{(1)}\}$ is the observed part of E .
18 Recall that we are momentarily assuming that there is no measurement error; hence we do not
19 distinguish between D_s (data value) and E_s (process value). In this simple formulation, one
20 wishes to make inference on the unobserved parts of E , $\bar{E}^{(1)} \equiv \{E_s : s \in A \setminus S^{(1)}\}$, from the
21 observed part $E^{(1)}$, where $A \setminus S^{(1)}$ means the index set of A with those in $S^{(1)}$ removed. Traditional
22 quantities of interest are the population total and the population mean: $T_E \equiv \sum_{s \in A} E_s$, and $\mu_E \equiv$

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

3
 4 Estimators of population quantities are functions of $S^{(1)}$ and $E^{(1)}$. For example, consider
 5 estimating the total T_E , which we can write as, $T_E = \sum_{s \in S^{(1)}} E_s + \sum_{s \in A \setminus S^{(1)}} E_s$. One popular
 6 estimator of T_E is known as the Horvitz-Thompson estimator,

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

7
 8
 9
 10 where p_s is the probability that s belongs to $S^{(1)}$; that is,

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

11
 12
 13
 14 Recall that for a probability sampling design (Overton, 1993), $\{p_s\}$ is *known*. The Horvitz-
 15 Thompson estimator is unbiased (e.g., Cochran, 1977); that is, the mean of the estimator is equal
 16 to the population quantity:

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

17
 18
 19
 20 and its sampling variance can be calculated in a like manner, again based on $[S^{(1)} | E]$. Notice
 21 that sampling moments (e.g., mean, variance) are calculated with respect to the $S^{(1)}$ that *could*
 22 *have* occurred, through the probability $[S^{(1)} | E]$.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

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. 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 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:

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

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

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

1 where recall from equation (2) that $P = \{P_D, P_E\}$. Section 2 has already discussed the
 2 importance of accounting for uncertainty in E . The joint distribution just above shows how this
 3 can be done in conjunction with the uncertainty in sampling.

4
 5 If parameters P are assumed fixed, inference on E is based on the posterior distribution:

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

8
 9 where the proportionality constant is $[D^{(1)}, S^{(1)} | P]$. Now suppose that uncertainty in $P = \{P_D,$
 10 $P_E\}$ is captured by the prior $[P]$; then the appropriate posterior distribution is:

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

13
 14 Many designs have the property that they are *ignorable* (e.g., Gelman et al., 2004), namely
 15 $[S^{(1)} | E] = [S^{(1)}]$, independent of the process E . In this case, an extraordinary simplification takes
 16 place:

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

19
 20 That is, uncertainty from sampling disappears from the inference, yielding the purely model-
 21 based inference developed in Section 2. (A similar result for ignorable designs occurs when P is
 22 assumed fixed.) Further details on incorporating sampling designs (ignorable and non-ignorable)
 23 in a hierarchical statistical model can be found in Gelman et al. (2004), Ch. 7.

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

9
10 In conclusion, this section has presented a unified way to treat sampling design and statistical
11 modeling together, using hierarchical statistical models. Many sampling designs are chosen to
12 be ignorable (such as simple random sampling, systematic sampling; e.g., see Cressie, 1993, pp.
13 316-323), in which case the approach given in Section 2 is appropriate. More generally, the
14 posterior distribution (3), which involves the sampling probabilities, should be used for
15 inference, but often is not.

16

17 **4. Statistical inference in ecological analyses**

18

19 In Section 2, we have shown how conditional probabilities can be used to build complex models
20 of ecological processes that account for uncertainty. Even if we agree that this approach is
21 reasonable, there remains the question of how one does inference in this setting. An empirical-
22 Bayesian approach could be taken. That is, consider only the first two stages, $[D | E, P_D]$ and

1 [E | P_E], and take the parameters $P = \{P_D, P_E\}$ to be fixed, but unknown. Depending on the
2 complexity of the component models in this case, it is often possible to use classical statistical-
3 estimation approaches to obtain estimates of the parameters P_D and P_E , and hence to use plug-in
4 predictions for E . For example, the usual linear mixed model can be thought of in this context
5 (Christensen, 1991). In addition, spatial prediction (kriging) fits into this framework (Cressie,
6 1993, Ch. 3), as do sequential time-series methods such as Kalman filtering. Common
7 approaches for estimation of P include the Expectation-Maximization (E-M) algorithm,
8 conditional and pseudo-likelihood methods, and estimating equations (Hardin and Hilbe, 2003).
9 Although such methods do not explicitly account for the uncertainty in estimating the
10 parameters, that uncertainty can, if deemed necessary, often be accounted for by Taylor-series
11 approximations (e.g., Rao, 2003, Sec. 6.2) and resampling and bootstrap procedures (Efron and
12 Tibshirani, 1993). For example, prior to using the Bayesian hierarchical model for harbor seals
13 given in Ver Hoef and Frost (2003), Frost et al. (1999) used a bootstrap approach. They modeled
14 variation in counts with Poisson regression for each site, and then they summed up the model
15 predictions across sites for a standardized set of covariate conditions for each year. Linear
16 regression was then used on these yearly sums to estimate trend. When doing “statistics on
17 statistics”, uncertainty from one analysis (parameter estimation) was passed to the next
18 (prediction) through resampling (here a bootstrap).

19

20 We turn now to the (fully) Bayesian approach. Here, we focus on discussing how hierarchical
21 models are fitted in the Bayesian paradigm, rather than on comparative inference. For a
22 comparison of Bayesian and likelihood-based methods of inference, see Browne and Draper
23 (2006) and references therein. The Bayesian paradigm has a conceptually holistic approach to

1 inference, where the parameters are also given distributions, and we are interested in the
 2 distribution of the process and parameters given the data: $[E, P_E, P_D | D]$. Our interest might be in
 3 inference on the parameters P_E and P_D , or in prediction or forecasting of the process E . In each
 4 case, we seek distributions of these variables, given the data that were actually observed. We use
 5 Bayes' Theorem from probability theory:

$$6 \quad [E, P_E, P_D | D] = [D | E, P_E, P_D][E, P_E, P_D]/[D] \quad (4)$$

$$7 \quad = [D | E, P_D][E | P_E][P_E, P_D]/ [D]$$

$$8 \quad \propto [D | E, P_D][E | P_E][P_E, P_D],$$

9
 10
 11
 12
 13 where we have already named the left-hand side the *posterior distribution*. The numerator on the
 14 right-hand side of (4) is the hierarchical decomposition developed above, and the proportionality
 15 constant, $1/[D]$, is the inverse of the marginal distribution of the data.

16
 17 Recall that the joint distribution of all uncertainties is $[D, E, P_E, P_D] = [D | E, P_E, P_D][E, P_E, P_D]$.
 18 We could view $[E, P_E, P_D]$ as the *prior distribution* on all “unknowns” (process and parameters)
 19 and the posterior distribution, $[E, P_E, P_D | D]$, represents what has been learned about the
 20 unknowns from the data D . Should new data come along, the posterior distribution can be
 21 viewed as an updated prior distribution, and an *updated posterior distribution* can be computed
 22 by considering the expression for the joint distribution of all uncertainties and replacing D with

1 the new data and using the updated prior. This is easily seen to be a coherent way of updating
2 information that arrives sequentially.

3
4 For the harbor-seals study, if we assume that the measurement error defining $D = \{Y_{ij}\}$ has
5 negative binomial variation, then the measurement-error process has a single parameter $P_D =$
6 $\{\kappa\}$. The parameters in P_E are mostly linear-regression coefficients of $\log\{\lambda_{ij}\}$, along with one
7 variance parameter σ^2 for a random effect. Typically, either κ or σ^2 have to have a peaked prior.
8 If it is κ , the prior $[P_E]$ can be chosen to be very diffuse and noninformative. Finally, inference is
9 based on $[E, P | D] = [\{\lambda_{ij}\}, P_D, P_E | \{Y_{ij}\}]$, where $P_E = \{\theta, \sigma^2\}$. For example, Figure 3 shows the
10 distribution of $[\theta_{1i} | \{Y_{ij}\}]$; $i = 1, \dots, n$, where $\{\theta_{1i}\}$ are the regression coefficients of abundance
11 regressed on years (temporal trend). This is one summary of the ecological process that could
12 alert managers to a decline in abundance. Inference on the parameter σ^2 based on $[\sigma^2 | \{Y_{ij}\}]$
13 would show how variable the abundances are from site to site, which would serve as a warning
14 that a “one-size-fits-all” management practice would be unsuccessful.

15
16 In principle, $[D]$ in (4) is obtained by integrating out (in the continuous-distribution case) the
17 process and the parameters in the numerator, but in practice it is seldom possible to obtain the
18 constant analytically. As a result, numerical methods must be used. The realization that Markov
19 Chain Monte Carlo (MCMC) methods could be used efficiently and generally for Bayesian
20 hierarchical models (Gelfand and Smith, 1990) revolutionized such computation, and it extended
21 the applicability of these models to ever-more-complicated modeling scenarios. MCMC is a
22 simulation-based method for drawing samples from probability distributions, where a Markov
23 chain is constructed such that its stationary, or long-run, distribution coincides with the

1 distribution from which random samples are desired. This distribution is sometimes called the
2 “target” distribution which, for a Bayesian analysis, is the posterior distribution. It follows that
3 after a sufficient number of realizations, or a “burn-in”, the generated realizations of the chain
4 comprise a random sample from the posterior distribution.

5

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

9

$$\begin{aligned} 10 & [E | P_E, P_D, D] \\ 11 & [P_E | E, P_D, D] \\ 12 & [P_D | E, P_E, D], \end{aligned}$$

13

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

17

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

22

$$23 \quad g(\cdot | P_E, P_D, D) / \int g(E | P_E, P_D, D) dE ,$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

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

1 George, 1992; Chib et al., 1995; Gilks et al., 1996; Chen et al., 2000; Gelman et al., 2004). For
2 an introduction to MCMC in the ecological literature, see Link (2002).

3
4 In order to summarize inferences on the unknowns in a hierarchical Bayesian model using the
5 output from an MCMC simulation, the sampled values from the posterior distribution are used to
6 calculate common distributional summary statistics, such as histograms, means, and variances, of
7 the marginal distributions of the unknowns, conditional on the observations D . For example, the
8 samples of a particular unknown E_1 , from the posterior distribution, can be averaged to
9 approximate the center of the marginal posterior distribution of E_1 , which is $E(E_1 | D)$, where
10 $E(\cdot)$ denotes expectation. In addition to the mean of the samples, it is also common to
11 characterize the uncertainty in the marginal posterior distribution of each unknown using the 2.5
12 and 97.5 percentiles of the posterior samples. The interval defined by these two percentiles
13 approximates the 95-percent credible interval of each unknown. The interpretation of a 95-
14 percent credible interval for an unknown E_1 is: the posterior probability that E_1 falls inside its
15 95-percent credible interval is 0.95. Thus, the widths of the credible intervals summarize the
16 uncertainty in inferences on E_1 . For the harbor-seals study (Ver Hoef and Frost, 2003), E_1
17 contains the temporal-trend parameters $\{\theta_{1t}\}$ and θ_{1A} , and the marginal distributions of the
18 MCMC samples are given in Figure 3, from which means, credible intervals, and so forth, can be
19 computed.

20
21 There are a variety of different software packages available for fitting Bayesian hierarchical
22 models using MCMC methods. Several of these packages are based on the Bayesian inference
23 Using Gibbs Sampling (BUGS) language, including its Windows implementation (WinBUGS;

1 Spiegelhalter et al., 2004) and an open-source version (OpenBUGS; Thomas, 2004). Another
2 open-source software package for posterior simulation is Just Another Gibbs Sampler (JAGS;
3 Plummer, 2005). The JAGS model-specification syntax is nearly identical to BUGS, however
4 the implementation is different. While all of these MCMC software packages can greatly
5 facilitate the fitting of Bayesian hierarchical models, many researchers choose to write their own
6 MCMC code in languages such as R/S, C/C++, and FORTRAN. This option provides increased
7 flexibility over the various packages, but can be considerably more tedious to implement and
8 debug. For their analysis of the harbor-seals data, Ver Hoef and Frost (2003) used WinBUGS to
9 obtain samples from the posterior distribution.

10
11 A considerable amount of research in quantitative ecology has been devoted to model selection.
12 We now reconsider this part of inference in the context of hierarchical statistical modeling. For a
13 simple example of model selection, consider again the general problem of “curve fitting”. Our
14 goal is to *select* the “best” curve. How do we define best, and how do we decide which is best?
15 Most methods try to find a balance or can be viewed as a balance between parsimony (simpler is
16 better) and goodness-of-fit (we want enough complexity to capture the essential features of the
17 data). For prediction, we can consider “averaging” over models, rather than trying to select the
18 “best” one (e.g., Hoeting et al., 1999), which has the effect of combining the advantages from all
19 models under consideration. These same considerations apply when trying to capture
20 uncertainty in process models and parameter models. Textbook treatments of model selection
21 can be found in Linhart and Zucchini (1986), McQuarrie and Tsai (1998), and Burnham and
22 Anderson (2002).

23

1 For the harbor-seals study, we might want to simplify the model as much as possible by selecting
2 among models that include year and a subset of tide, date, and time. A traditional approach is to
3 use stepwise regression, but more popular recently is an information-theoretic approach, such as
4 AIC (Akaike, 1973) or BIC (Schwarz, 1978). These approaches are especially useful when
5 comparing non-nested models. For example, suppose that we want to use all covariates, year,
6 tide, date, and time in our model, but we want to compare a Poisson distribution versus a
7 negative binomial distribution for the measurement errors. Then a stepwise approach is not
8 appropriate, however the information-theoretic approach handles the comparison easily. The
9 information-theoretic approach uses the notion of maximizing the model likelihood (which can
10 never get worse by adding more parameters), but with a penalty for the number of parameters.
11 However, the counting of parameters can become difficult for hierarchical statistical models, so
12 Spiegelhalter et al. (2002) introduced DIC; DIC would be useful in the harbor-seals study,
13 especially when trend over years is considered. Small, Pendleton, and Pitcher (2003) used an
14 information-theoretic approach on a non-hierarchical analysis of the harbor-seals data.

15

16 **5. Challenges for hierarchical statistical modeling**

17

18 Hierarchical models address complex problems for which there may be several sources of
19 information and hidden variables (Wikle et al., 1998; Abbitt and Breidt, 2002; Brooks et al.,
20 2004; Clark, 2005; Ogle et al., 2005; Royle and Dorazio, 2006). Goals may include estimation
21 of parameters for a process model (Ibanez et al., 2006; Hille Ris Lambers et al., 2005), inference
22 on hidden states (Dupris, 1995; Ver Hoef and Cressie, 1997; deValpine, 2003; Stenseth et al.,
23 2003; Dorazio and Royle, 2003; Clark and Bjornstad, 2004; Gelfand et al., 2004; Maunder,

1 2004; Clark et al., 2005; Thomas et al., 2005), quantifying the importance of interactions
2 (Coulson et al., 2001; Cam et al., 2002; Clark et al., 2003, 2004; LaDeau and Clark, 2006;
3 Mohan et al., 2007), prediction (Beckage and Platt, 2003; Clark et al., 2007; Hooten et al., 2007),
4 and specifying species diversity and patterns (Gelfand et al., 2005a, 2005b; Latimer et al., 2006).
5 In this section, we begin a discussion of the issues surrounding hierarchical statistical modeling.
6 It is expected that in follow-up discussion, these and other issues will be explored more deeply
7 than we are able to do here.

8
9 There are many challenges with building complicated models, with the associated inference, and
10 with computational efficiency (Clark, 2005; Clark and Gelfand, 2006; and Buckland et al., 2007,
11 provide ecological examples). Like any model-building exercise, there must be considerable
12 thought and effort devoted to the specification of the component distributions in the hierarchical
13 framework. In some sense, model complexity is related both to data richness and "scientific
14 richness", which is simply the knowledge one has about the process (and parameters) of interest.
15 The situation is only made more complex when the model is also Bayesian.

16
17 *Subjectivity of a Bayesian analysis*
18 A historical criticism of Bayesian methods is that it requires "subjective" specification of prior
19 information on the parameters. Can a Bayesian get any answer he or she desires by tuning the
20 prior distribution? Technically, the answer to this question is yes; the Bayesian approach can be
21 "gamed", but so can the classical/frequentist approach to inference. In fact, the elicitation of
22 prior distributions can be turned into a (social) science. Subjective judgements within a group of

1 expert ecologists can lead to a subjective prior where both the consensus opinion and the
2 diversity of opinions is recognized (e.g., McCarthy, 2007, Ch. 10).

3
4 There is also a sizable literature on “objective Bayesian” analysis, which develops prior
5 distributions for classes of models that do not impact the posterior distribution. Such
6 noninformative priors are often termed “vague”, “flat”, or “diffuse”. For some models,
7 appropriate noninformative priors are readily available, while for others they are not. For a
8 thorough discussion of issues related to the specification of prior distributions, see Berger
9 (1985). Alternatively, it is common practice to perform sensitivity analyses, or robustness
10 assessments, to ascertain the influence of prior assumptions. This is done by refitting the
11 hierarchical model under different prior assumptions and comparing the resulting inferences.
12 Not only does this exercise provide information on the influence of prior assumptions on
13 inferences, but it gives insight into the amount of information in the data.

14
15 Of course, there is subjectivity in the specification of the likelihood in a classical statistical
16 approach. In fact, a broader perspective is that there is subjectivity involved with the
17 specification of *all* of the model components: data models, process models, and parameter
18 models. For example, it might be "subjective" to specify a stochastic model for tropical winds
19 (e.g., Wikle et al., 2001), but the science comes from Newton's laws of motion! There has
20 always been subjectivity present in the physical sciences; now, instead of having to be certain
21 about a physical model, scientists can quantify their uncertainty about their model in a way that
22 disentangles it from their uncertainty about their data. Thus, from our perspective, it is better to

1 ask about the sensitivity of results to model choices and whether such choices make sense
2 scientifically.

3

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

10

11

12 *Convergence of the MCMC algorithm*

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

18

19 In practice, convergence is usually assessed using a set of diagnostic tools on the sample path of
20 the Markov chain. While this type of output analysis is not able to “prove” that the chain has
21 converged, it can provide insight into the issue. Output analysis can be numerical and graphical
22 and can involve both comparing chains with multiple starting values and assessing differences in
23 different segments of a single chain (e.g., Brooks and Roberts, 1998; Cowles and Carlin, 1995;

1 Gelman et al., 2004). Software for assessing convergence of MCMC algorithms includes the
2 Convergence Diagnosis and Output Analysis Software of Gibbs Sampling Output (CODA; Best
3 et al., 1995; Plummer et al., 2005) and Bayesian Output Analysis (BOA; Smith, 2005).

4

5 *Adequacy of the number of MCMC samples*

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

15

16 *Parameter Identifiability*

17 When building large hierarchical models, it is not always apparent that all model parameters are
18 identifiable. In order to motivate the issue of parameter identifiability, consider the following
19 simple example. Let y_1, \dots, y_n be independent samples from a normal distribution with unknown
20 mean μ and variance σ^2 . Clearly, from a classical/frequentist perspective, the observations
21 contain information about the unknown parameters. In fact, the sample mean and variance
22 provide unbiased estimates of μ and σ^2 , respectively. Instead of parameterizing the sampling
23 distribution in terms of μ and σ^2 , assume that the data were generated from a normal distribution

1 with unknown mean $\mu_1 + \mu_2$ and variance σ^2 . While there is information in the data about the
2 sum, $\mu_1 + \mu_2$, there is no way to identify the components separately, which means that μ_1 and μ_2
3 are not identifiable.

4
5 From a Bayesian perspective, as long as proper prior distributions are assigned to all model
6 parameters, all of them are technically identifiable. To illustrate, consider the example above,
7 now within the Bayesian paradigm. Assuming proper prior distributions on μ_1 , μ_2 , and σ^2 , we
8 can determine the posterior distribution of all three parameters, making them all technically
9 identifiable. However, unless there is very strong prior information on μ_1 and μ_2 , it will be
10 difficult to see very much difference in the posterior distributions of these parameters. In
11 practice, lack of identifiability can be problematic in the Bayesian setting (Gelfand and Sahu,
12 1999). For example, when running MCMC algorithms, the sample paths of nonidentifiable
13 parameters will “trade off” their values, leading to numerical and convergence problems.
14 Generally speaking, if identifiability problems go undiagnosed, inferences on these model
15 parameters and possibly others can be misleading. This can sometimes be anticipated, in which
16 case a sampling design (Section 4) might be formulated that avoids identifiability problems in
17 making inference from equation (3).

18
19 *Assessing Model Fit and Diagnosing Lack of Fit*

20 We have already discussed the importance of assessing the influence of prior assumptions on
21 inferences in Bayesian models. In addition to performing these sensitivity analyses, it is
22 important to assess the fit of a model. There are both numerical and graphical tools for model
23 checking, which are primarily based on the posterior predictive distribution, specifically the

1 distribution of a new (replicate) observation, D^{rep} , conditional on the observed data, D (see
2 Gelman et al., 2004, Ch. 6). The posterior predictive distribution is given by:

3

$$4 \quad [D^{\text{rep}} | D] = \int [D^{\text{rep}} | E, P_E, P_D][[E, P_E, P_D | D]dE dP_E dP_D,$$

5

6 where the distribution is obtained numerically by taking an MCMC sample from the posterior
7 distribution, $[E, P_E, P_D | D]$, and plugging the sampled values of E, P_E , and P_D into $[D^{\text{rep}} | E, P_E,$
8 $P_D]$ to then generate a sample from $[D^{\text{rep}} | D]$.

9

10 With any model-checking procedure comes the opportunity to perform cross-validation. That is,
11 the i^{th} observation D_i (or a carefully chosen subset of observations) in D is deleted, leaving the
12 remaining data, D_{-i} , to predict what should have been observed assuming the model is correct. A
13 Bayesian model allows the posterior distribution $[D_i^{\text{rep}} | D_{-i}]$ to be computed, which can then be
14 compared to the observation D_i that was deleted (e.g., Stern and Cressie, 2000). For example, if
15 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
16 ability to fit well to the i^{th} datum. Consequently, if cross-validation is used as an informal
17 model-selection procedure, it tends to reject overly complicated models, since such models
18 perform poorly when predicting beyond the observed data.

19

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

1 task. Still, if inferences and forecasts are going to be trusted, these checks must be performed.
2 In the harbor-seals study, Ver Hoef and Frost (2003) used $[D^{\text{rep}} | D]$ to average $(\text{Observed}_{ij} -$
3 $\text{Expected}_{ij})^2 / \text{Expected}_{ij}$, over j , for each site i , which should be near one. This is a generic
4 goodness-of-fit statistic that is not tailor-made for any particular departure from the model for
5 which Expected_{ij} is calculated. They found lack of fit for models without a random effect to
6 account for overdispersion. That is, for the models given in Figure 2, the parameters σ^2 or $\{\sigma_i^2\}$
7 are important for accounting for dispersion about the “curves” $\{\mu_{ij}\}$.

8

9

10 *Bayesian inference in space and time*

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

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. (2007).

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.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

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, which was 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. 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 at the MBI’s website (<http://www.mbi.osu.edu/2005/ws5abstracts.html>). 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 Citations

Abbitt, P.J. and Breidt, F.J. 2002. A hierarchical model for estimating distribution profiles of soil texture. In *Case Studies in Bayesian Statistics, Volume 5* (Edited by C. Gatsonis,

- 1 R.E. Kass, B. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West) Springer-
2 Verlag, New York, pp. 263-278.
- 3 Aldworth, J. and Cressie, N. 1999. Sampling designs and prediction methods for Gaussian
4 spatial processes. In *Mathematical Design and Sampling* (Edited by S. Ghosh). Marcel
5 Dekker, New York, pp. 1-54.
- 6 Akaike, H. 1973. Information theory and the extension of the maximum likelihood principle. In
7 *Proceedings of the International Symposium on Information Theory* (Edited by B. N.
8 Petrov and F. Czaki). Akademia Kiado, Budapest, pp. 267-281.
- 9 Arbia, G. and Lafratta, G. 1997. Evaluating and updating the sample design in repeated
10 environmental surveys: monitoring air quality in Padua. *Journal of Agricultural,
11 Biological, and Environmental Statistics* **2**: 451-466.
- 12 Banerjee, S., Carlin, B.P., and Gelfand, A.E. 2004. *Hierarchical Modeling and Analysis for
13 Spatial Data*. Chapman and Hall/CRC Press, Boca Raton.
- 14 Barber, J.J. and Gelfand, A.E. 2007. Hierarchical spatial modeling for estimation of population
15 size. *Environmental and Ecological Statistics* **14**: 193-205.
- 16 Beckage, B. and Platt, W.J. 2003. Predicting severe wildfire years in the Florida Everglades.
17 *Frontiers in Ecology and the Environment* **1**: 235-239.
- 18 Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New
19 York.
- 20 Berliner, L.M. 1996. Hierarchical Bayesian time series models. In *Maximum Entropy and
21 Bayesian Methods* (Edited by K. Hanson and R. Silver). Kluwer Academic Publishers,
22 Dordrecht, Netherlands, pp. 15-22.

- 1 Berliner, L.M., Wikle, C.K., and Cressie, N. 2000. Long-lead prediction of Pacific SSTs via
2 Bayesian dynamic modeling. *Journal of Climate* **13**: 3953-3968.
- 3 Best, N.G., Cowles, M.K., and Vines, S.K. 1995. *CODA Manual version 0.30*. MRC
4 Biostatistics Unit, Cambridge, UK.
5 <http://www.mrc-bsu.cam.ac.uk/bugs/documentation/coda03/cdaman03.html>.
- 6 Bjornstad, O.N. and Grenfeld, B.F. 2001. Noisy clockwork: Time series analysis of population
7 fluctuations in animals. *Science* **293**: 638-643.
- 8 Boily, P. 1995. Theoretical heat flux in water and habitat selection of phocid seals and beluga
9 whales during the annual molt. *Journal of Theoretical Biology* **172**: 235-244.
- 10 Botkin, D.B., Janak, J.F., and Wallis, J.R. 1972. Some ecological consequences of a computer
11 model of forest growth. *Journal of Ecology* **60**: 849-872.
- 12 Boveng, P.L., Bengtson, J.L., Cesarone, J.C., Simpkins, M.A., Frost, K.J., and Burns, J.J. 2003.
13 The abundance of harbor seals in the Gulf of Alaska. *Marine Mammal Science* **19**: 111-
14 127.
- 15 Brooks, S.P., King, R., and Morgan, B.J.T. 2004. A Bayesian approach to combining animal
16 abundance and demographic data. *Animal Biodiversity and Conservation* **27**: 515-529.
- 17 Brooks, S.P. and Roberts, G.O. 1998. Assessing convergence of Markov Chain Monte Carlo
18 algorithms. *Statistics and Computing* **8**: 319-335.
- 19 Browne, W.J. and Draper, D. 2006. A comparison of Bayesian and likelihood-based methods for
20 fitting multilevel models. *Bayesian Analysis* **1**: 473-514.
- 21 Buckland, S.T., Newman, K.B., Fernández, C., Thomas, L., and Harwood, J. 2007. Embedding
22 population dynamics models in inference. *Statistical Science* **22**: 44-58.

- 1 Burnham, K.P. and Anderson, D.R. 2002. *Model Selection and Multimodel Inference: A*
2 *Practical Information Theoretic Approach, 2nd Edition*. Springer Verlag, New York.
- 3 Calder, C.A. 2007. Dynamic factor process convolution models for multivariate space-time data
4 with application to air quality assessment. *Environmental and Ecological Statistics* **14**:
5 229-247.
- 6 Calder, C.A., Lavine, M., Mueller, P., and Clark, J.S. 2003. Incorporating multiple sources of
7 stochasticity in population dynamic models. *Ecology* **84**: 1395-1402.
- 8 Cam, E., Link, W.A., Cooch, E.G., Monnat, J.-Y., and Danchin, E. 2002. Individual covariation
9 in life-history traits: Seeing the trees despite the forest. *American Naturalist* **159**: 96-
10 105.
- 11 Carlin, B.P. and Louis, T.A. 2000. *Bayes and Empirical Bayes Methods for Data Analysis, 2nd*
12 *Edition*. Chapman and Hall/CRC, Boca Raton.
- 13 Casella, G., and George, E.I. 1992. Explaining the Gibbs sampler. *The American Statistician*
14 **46**: 167-174.
- 15 Catchpole, E.A., Hatton, E.A., and Catchpole, W.R. 1989. Fire spread through
16 nonhomogeneous fuel modeling as a Markov process. *Ecological Modelling* **48**: 101-
17 112.
- 18 Chao, C.-T. and Thompson, S.K. 2001. Optimal adaptive selection of sampling sites.
19 *Environmetrics* **12**: 517-538.
- 20 Chen, M.-H., Shao, Q.-M., and Ibrahim, J.G. 2000. *Monte Carlo Methods in Bayesian*
21 *Computation*. Springer Verlag, New York.
- 22 Chib, S., and Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *The*
23 *American Statistician* **49**: 327-335.

- 1 Christensen, R. 1991. *Linear Models for Multivariate, Time Series, and Spatial Data*. Springer-
2 Verlag, New York.
- 3 Clark, J.S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* **8**:
4 214.
- 5 Clark, J.S. and Bjornstad, O. 2004. Population time series: Process variability, observation
6 errors, missing values, lags, and hidden states. *Ecology* **85**: 3140-3150.
- 7 Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dodson, A., Foley, J., Lodge, D., Pascual,
8 M., Pielke, Jr., R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger,
9 W.H., Wall, D., and Wear, D. 2001. Ecological forecasts: An emerging imperative.
10 *Science* **293**: 657-660.
- 11 Clark, J.S., Ferraz, G., Oguge, N., Hayes, H., and DiCostanzo, J. 2005. Hierarchical Bayes for
12 structured and variable populations: From capture-recapture data to life-history
13 prediction. *Ecology* **86**: 2232-2244.
- 14 Clark, J.S. and Gelfand, A.E. 2006. A future for models and data in ecology. *Trends in Ecology*
15 *and Evolution* **21**: 375-380.
- 16 Clark, J.S., LaDeau, S., and Ibanez, I. 2004. Fecundity of trees and the colonization-
17 competition hypothesis. *Ecological Monographs* **74**: 415-442.
- 18 Clark, J.S., Mohan, J., Dietze, M., and Ibanez, I. 2003. Coexistence: How to identify trophic
19 tradeoffs. *Ecology* **84**: 17-31.
- 20 Clark, J.S., Wolosin, M., Dietze, M., Ibanez, I., LaDeau, S., Welsh, M., and Kloeppel, B. 2007.
21 Tree growth inference and prediction from diameter censuses and ring widths.
22 *Ecological Applications* **17**: 1942-1953.
- 23 Cochran, W.G. 1977. *Sampling Techniques, 3rd Edition*. Wiley, New York.

- 1 Coulson, T., Catchpole, E.A., Albon, S.D., Morgan, B.J.T., Pemberton, J.M., Clutton-Brock,
2 T.H., Crawley, M.J., and Genfell, B.T. 2001. Age, sex, density, winter weather and
3 population crashes in Soay sheep. *Science* **292**: 1528-1531.
- 4 Cowles, M.K. and Carlin, B.P. 1995. Markov chain Monte Carlo diagnostics: A comparative
5 review. *Journal of the American Statistical Association* **91**: 883-904.
- 6 Cressie, N. 1992. Smoothing regional maps using empirical Bayes predictors. *Geographical*
7 *Analysis* **24**: 75-95.
- 8 Cressie, N. 1993. *Statistics for Spatial Data, rev. edition*. Wiley, New York.
- 9 Cressie, N. 1998. Transect-spacing design of ice cores on the Antarctic continent. *Canadian*
10 *Journal of Statistics* **26**: 405-418.
- 11 da Silva, J. and Terhune, J.M. 1988. Harbour seal grouping as an anti-predator strategy. *Animal*
12 *Behavior* **36**: 1309-1316.
- 13 Daszak, P., Cunningham, A.A., and Hyatt, A.D. 2000. Emerging infectious diseases of
14 wildlife – threats to biodiversity and human health. *Science* **287**: 443-449.
- 15 de Valpine, P. 2003. Better inferences from population-dynamics experiments using Monte
16 Carlo state-space likelihood methods. *Ecology* **84**: 3064-3077.
- 17 Dorazio, R.M. and Royle, J.A. 2003. Mixture models for estimating the size of a closed
18 population when capture rates vary among individuals. *Biometrics* **59**: 351-364.
- 19 Dupuis, J.A. 1995. Bayesian estimation of movement and survival probabilities from capture-
20 recapture data. *Biometrika* **82**: 761-772.
- 21 Efron, B. and Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New
22 York.

- 1 Federov, V.V. and Nachtsheim, C. 1995. Optimal designs for time-dependent responses. In
2 *Proceedings of MODA4* (Edited by C. P. Kitsos, and W.G. Muller). Physica-Verlag,
3 Heidelberg.
- 4 Frost, K.J., Lowry, L.F., and Ver Hoef, J.M. 1999. Monitoring the trend of harbor seals in
5 Prince William Sound, Alaska, after the *Exxon Valdez* oil spill. *Marine Mammal Science*
6 **15**: 494-506.
- 7 Furrer, R., Sain, S.R., Nychka, D., and Meehl, G.A. 2007. Multivariate Bayesian analysis of
8 atmosphere-ocean general circulation models. *Environmental and Ecological Statistics*
9 **14**: 249-266.
- 10 Gelfand, A.E. and Sahu, S. 1999. Identifiability, improper priors, and Gibbs sampling for
11 generalized linear models. *Journal of the American Statistical Association* **94**: 247-253.
- 12 Gelfand, A.E., Schmidt, A.M., Wu, S., Silander, J.A., Latimer, A., and Rebelo, A.G. 2005a.
13 Explaining species diversity through species level hierarchical modeling. *Applied*
14 *Statistics* **65**: 1-20.
- 15 Gelfand, A.E., Silander, J.A., Wu, S., Latimer, A., Lewis, P.O., Rebelo, A.G., and Holder, M.
16 2005b. Explaining species distribution patterns through hierarchical modeling. *Bayesian*
17 *Analysis* **1**: 41-92.
- 18 Gelfand, A.E., and Smith, A.F.M. 1990. Sampling based approaches to calculating marginal
19 densities. *Journal of the American Statistical Association* **85**: 398-409.
- 20 Gelfand, A.E., Zhu, I., and Carlin, B.P. 2001. On the change of support problem for spatio-
21 temporal data. *Biostatistics* **2**: 31-45.
- 22 Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2004. *Bayesian Data Analysis, 2nd*
23 *Edition*. Chapman & Hall/CRC, New York.

- 1 Gelman, A. and Hill, J. 2006. *Data Analysis using Regression and Mutlilevel/Hierarchical*
2 *Models*. Cambridge University Press, New York.
- 3 Gilks, W.R., Richarson, S., and Spiegelhalter, D.J. 1996. *Markov Chain Monte Carlo in*
4 *Practice*. Chapman & Hall, New York.
- 5 Gotway, C.A. and Young, L.J. 2002. Combining incompatible spatial data. *Journal of the*
6 *American Statistical Association* **97**: 632-648.
- 7 Hardin, J.W. and Hilbe, J.M. 2003. *Generalized Estimating Equations*. Chapman and
8 Hall/CRC, Boca Raton, Florida.
- 9 Hilborn, R. and Mangel, M. 1997. *The Ecological Detective. Confronting Models with Data*.
10 Princeton University Press, Princeton, New Jersey.
- 11 Hille Ris Lambers, J., Clark, J.S., and Lavine, M. 2005. Seed banking in temperate forests:
12 Implications for recruitment limitation. *Ecology* **86**: 85-95.
- 13 Hoeting, J.A., Madigan, D., Raftery, A., and Volinsky, C.T. 1999. Bayesian model averaging:
14 A tutorial. *Statistical Science* **14**: 382-417.
- 15 Hooten, M.B. and Wikle, C.K. 2007. Shifts in the spatial-temporal growth dynamics of shortleaf
16 pine. *Environmental and Ecological Statistics* **14**: 207-227.
- 17 Hooten, M.B., Wikle, C.K., Dorazio, R.M., and Royle, J.A. 2007. Hierarchical spatiotemporal
18 matrix models for characterizing invasions. *Biometrics* **63**: 558-567.
- 19 Hoover-Miller A., Parker, K.R., and Burns, J.J. 2001. A reassessment of the impact of the
20 *Exxon Valdez* oil spill on harbor seals (*Phoca vitulina richardsi*) in Prince William
21 Sound, Alaska. *Marine Mammal Science* **17**: 111-135.
- 22 Huang, H.-C. and Cressie, N. 1996. Spatio-temporal prediction of snow-water equivalent using
23 the Kalman filter. *Computational Statistics and Data Analysis* **22**: 159-175.

- 1 Ibanez, I., Clark, J.S., Dietze, M.C., Feeley, K., Hersh, M., LaDeau, S., McBride, A., Welch,
2 N.E., and Wolosin, M.S. 2006. Predicting biodiversity change: Outside the climate
3 envelope, beyond the species-area curve. *Ecology* **87**: 1896-1906.
- 4 Johnson, D.S., Hoeting, J.A., and Poff, N.L. 2006. Biological monitoring: A Bayesian model
5 for multivariate compositional data. In *Bayesian Statistics and its Applications* (Edited
6 by S.K. Upadhyay, U. Singh, and D.K. Dey). Anamaya Publishers, New Delhi, pp. 270-
7 289.
- 8 Latimer, A.M., Wu, S., Gelfand, A.E., and Silander, J.A. 2006. Building statistical models to
9 analyze species distributions. *Ecological Applications* **16**: 33-50.
- 10 LaDeau, S.L. and Clark, J.S. 2006. Elevated CO₂ and tree fecundity: The role of tree size,
11 interannual variability, and population heterogeneity. *Global Change Biology* **12**: 822-
12 833.
- 13 Le, N.D. and Zidek, J.V. 1994. Network designs for monitoring multivariate random spatial
14 fields. In *Recent Advances in Statistics and Probability* (Edited by J.P. Vilaplana and
15 M.L. Puri). VSP, Leiden, Netherlands, pp. 191-206.
- 16 Lele, S.R., Dennis, B., and Lutscher, F. 2007. Data cloning: easy maximum likelihood
17 estimation for complex ecological models using Bayesian Markov chain Monte Carlo
18 methods. *Ecology Letters* **10**: 551-563.
- 19 Linhart, H. and Zucchini, W. 1986. *Model Selection*. Wiley, New York.
- 20 Link, W.A., Cam, E., Nichols, J.D., and Cooch, E.G. (2002). Of BUGS and birds: Markov
21 chain Monte Carlo for hierarchical modeling in wildlife research. *Journal of Wildlife*
22 *Management* **66**: 277-291.

- 1 Martin, A.D., and Quinn, K.M. 2004. *MCMCpack: Markov chain Monte Carlo (MCMC)*
2 *Package*. <http://mcmcpack.wustl.edu>. R package version 0.6-3.
- 3 Maunder, M.N. 2004. Population viability analysis based on combining integrated, Bayesian,
4 and hierarchical analysis. *Acta Oecologica* **26**: 85-94.
- 5 McCarthy, M.A. 2007. *Bayesian Methods for Ecology*. Cambridge University Press,
6 Cambridge, UK.
- 7 McQuarrie, A. D., and Tsai, C.-L. 1998. *Regression and Time Series Model Selection*. World
8 Scientific, Hackensack, NJ.
- 9 Medawar, P.B. 1940. Growth, growth energy, and ageing of the chicken's heart. *Proceedings*
10 *of the Royal Society of London* **129**: 332-355.
- 11 Meinhold, J. and Singpurwalla, N. 1983. Understanding the Kalman filter. *American*
12 *Statistician* **37**: 123-127.
- 13 Mohan, J.E., Clark, J.S., and Schlesinger, W.H. 2007. Long-term CO₂ enrichment of an intact
14 forest ecosystem: Implications for temperate forest regeneration and succession.
15 *Ecological Applications* **17**: 1198-1212
- 16 Morris, C.N. 1983a. Parametric empirical Bayes inference: Theory and applications. *Journal*
17 *of the American Statistical Association* **78**: 47-55.
- 18 Morris, C.N. 1983b. Parametric empirical Bayes confidence intervals. In *Scientific Inference,*
19 *Data Analysis, and Robustness* (Edited by G.E.P. Box, T. Leonard, and C.F.J. Wu).
20 Academic Press, New York, pp. 25-50.
- 21 Mugglin, A.S., Carlin, B.P., and Gelfand, A.E. 2000. Fully model-based approaches for
22 spatially misaligned data. *Journal of the American Statistical Association* **95**: 877-887.

- 1 Ogle, K., Wolpert, R.L., and Reynolds, J.F. 2004. Reconstructing plant root area and water
2 uptake profiles. *Ecology*, **85**: 1967-1978.
- 3 Overton, W.S. 1993. Probability sampling and population inference in monitoring programs. In
4 *Environmental Modeling with GIS* (Edited by M.R. Goodchild, B.O. Parks, and C.T.
5 Steyaert). Oxford University Press, New York, pp. 470-480.
- 6 Plummer, M. 2005. *JAGS: Just Another Gibbs Sampler*.
7 <http://www-fis.iarc.fr/~martyn/software/jags/>.
- 8 Plummer, M., Best, N., Cowles, K., and Vines, K. 2005. *CODA: Output Analysis and*
9 *Diagnostics for MCMC*. <http://www-fis.iarc.fr/coda/>. R package version 0.9-2.
- 10 Prasad, N.G.N. and Rao, J.N.K. 1990. On the estimation of mean square error of small area
11 predictors. *Journal of the American Statistical Association* **85**: 163-171.
- 12 Rao, J.N.K. 2003. *Small Area Estimation*. Wiley, New York.
- 13 Ricker, W.E. 1954. Stock and recruitment. *Journal of Fisheries Research Board of Canada* **11**:
14 559-623.
- 15 Roberts, C.P. and Casella, G. 2005. *Monte Carlo Statistical Methods, 2nd Edition*. Springer,
16 New York.
- 17 Royle, J.A. and Berliner, L.M. 1999. A hierarchical approach to multivariate spatial modeling
18 and prediction. *Journal of Agricultural, Biological, and Environmental Statistics* **4**: 29-
19 56.
- 20 Royle, J.A. and Dorazio, R.M. (2006). Hierarchical models of animal abundance and
21 occurrence. *Journal of Agricultural, Biological, and Environmental Statistics* **11**: 249-
22 263.

- 1 Sacks, W., Schimel, D., and Monson, R. 2007. Coupling between carbon cycling and climate in
2 a high-elevation, subalpine forest: a model-data fusion analysis. *Oecologia*, submitted.
- 3 Sain, S. and Cressie, N. 2007. A spatial model for multivariate lattice data. *Journal of*
4 *Econometrics* **140**: 226-259.
- 5 Särndal, C., Swensson, B., and Wretman, J. 1992. *Model Assisted Survey Sampling*. Springer-
6 Verlag, New York.
- 7 Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.
- 8 Sims, M., Lewison, R., and Cox, T. 2007. Identifying spatial patterns in fisheries by catch:
9 Using models to improve stability of estimates and aid management. *Ecological*
10 *Applications*, in press.
- 11 Small, R.J., Pendleton, G.W., and Pitcher, K.W. 2003. Trends in abundance of Alaska harbor
12 seals, 1983-2002. *Marine Mammal Science* **19**: 344-362.
- 13 Smith, B.J. 2005. *Bayesian Output Analysis Program (BOA), Version 1.1.5*. The University of
14 Iowa. <http://www.public-health.uiowa.edu/boa>.
- 15 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. 2002. Bayesian measures
16 of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*.
17 *Series B* **64**: 583-639.
- 18 Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. 2004. *WinBUGS*.
19 <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/>.
- 20 Stenseth, N.C., Viljugrein, H., Saitoh, T., Hansen, T.T., Kittilsen, M.O., Bolviken, E., and
21 Glockner, F. 2003. Seasonality, density dependence, and population cycles in Hokkaido
22 voles. *Proceedings of the National Academy of Sciences* **100**: 11478-11483.

- 1 Stern, H.S. and Cressie, N. 2000. Posterior predictive model checks for disease mapping
2 models. *Statistics in Medicine* **19**: 2377-2397.
- 3 Stevens, D.L. and Olsen, A.R. 2004. Spatially balanced sampling of natural resources. *Journal*
4 *of the American Statistical Association* **99**: 262-278.
- 5 Thomas, A. 2004. *OpenBUGS*. <http://mathstat.helsinki.fi/openbugs/>.
- 6 Thomas, L., Buckland, S.T., Newman, K.B., and Harwood, J. 2005. A unified framework for
7 modelling wildlife population dynamics. *Australian and New Zealand Journal of*
8 *Statistics* **47**: 19-34.
- 9 Thompson, S.K. and Seber, A.F. 1996. *Adaptive Sampling*. Wiley, New York.
- 10 Tierney, L. 1994. Markov chains for exploring posterior distributions (with discussion). *Annals*
11 *of Statistics* **22**: 1701-1762.
- 12 Ver Hoef, J.M. 1996. Parametric empirical Bayes methods for ecological applications.
13 *Ecological Applications* **6**: 1047-1055.
- 14 Ver Hoef, J.M. 2002. Sampling and geostatistics for spatial data. *Ecoscience* **9**: 152-161.
- 15 Ver Hoef, J.M. and Barry, R.D. 1998. Constructing and fitting models for cokriging and
16 multivariable spatial prediction. *Journal of Statistical Planning and Inference* **69**: 275-
17 294.
- 18 Ver Hoef, J.M. and Cressie, N. 1997. Using hidden Markov chains and empirical Bayes change-
19 point estimation for transect data. *Environmental and Ecological Statistics* **4**: 247-264.
- 20 Ver Hoef, J.M. and Frost, K. 2003. A Bayesian hierarchical model for monitoring harbor seal
21 changes in Prince William Sound, Alaska. *Environmental and Ecological Statistics* **10**:
22 201-209.

- 1 Ver Hoef, J.M. and Jansen, J.K. 2007. Space-time zero-inflated count models of harbor seals.
2 *Environmetrics* **18**: 697-712.
- 3 Waller, L.A., Carlin, B.P., Xia, H., and Gelfand, A. 1997. Hierarchical spatio-temporal mapping
4 of disease rates. *Journal of the American Statistical Association* **92**: 607-617.
- 5 Watts, P. 1992. Thermal constraints on hauling out by harbor seals (*Phoca vitulina*). *Canadian*
6 *Journal of Zoology* **70**: 553-560.
- 7 West, M. and Harrison, J. 1997. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag,
8 New York.
- 9 Wikle, C.K. 2003a. Hierarchical models in environmental science. *International Statistical*
10 *Review* **71**: 181-199.
- 11 Wikle, C.K. 2003b. Hierarchical Bayesian models for predicting the spread of ecological
12 processes. *Ecology* **84**: 1382-1394.
- 13 Wikle, C.K. and Berliner, L.M. 2005. Combining information across spatial scales.
14 *Technometrics* **47**: 80-91.
- 15 Wikle, C.K., Berliner, L.M., and Cressie, N. 1998. Hierarchical Bayesian space-time models.
16 *Environmental and Ecological Statistics* **5**: 117-154.
- 17 Wikle, C.K. and Cressie, N. 1999. A dimension-reduced approach to space-time Kalman
18 filtering. *Biometrika* **86**: 815-829.
- 19 Wikle, C.K., Milliff, R.F., Nychka, D., and Berliner, L.M. 2001. Spatio-temporal hierarchical
20 Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical*
21 *Association* **96**: 382-397.

- 1 Wikle, C.K. and Royle, J.A. 1999. Space-time models and dynamic design of environmental
2 monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics*
3 **4**: 489-507.
- 4 Wikle, C.K. and Royle, J.A. 2005. Dynamic design of ecological monitoring networks for non-
5 Gaussian spatio-temporal data. *Environmetrics* **16**: 507-522.
- 6 Wright, S. 1926. Book Reviews. *Journal of the American Statistical Association* **21**: 493-497.
- 7 Xu, K., Wikle, C.K., and Fox, N.I. 2005. A kernel-based spatio-temporal dynamical model for
8 nowcasting radar precipitation. *Journal of the American Statistical Association* **100**:
9 1133-1144.
- 10 Zimmerman, D. 2006. Optimal network design for spatial prediction, covariance parameter
11 estimation, and empirical prediction. *Environmetrics* **17**: 635-662.

12

1 **Figure Legends**

2

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

7

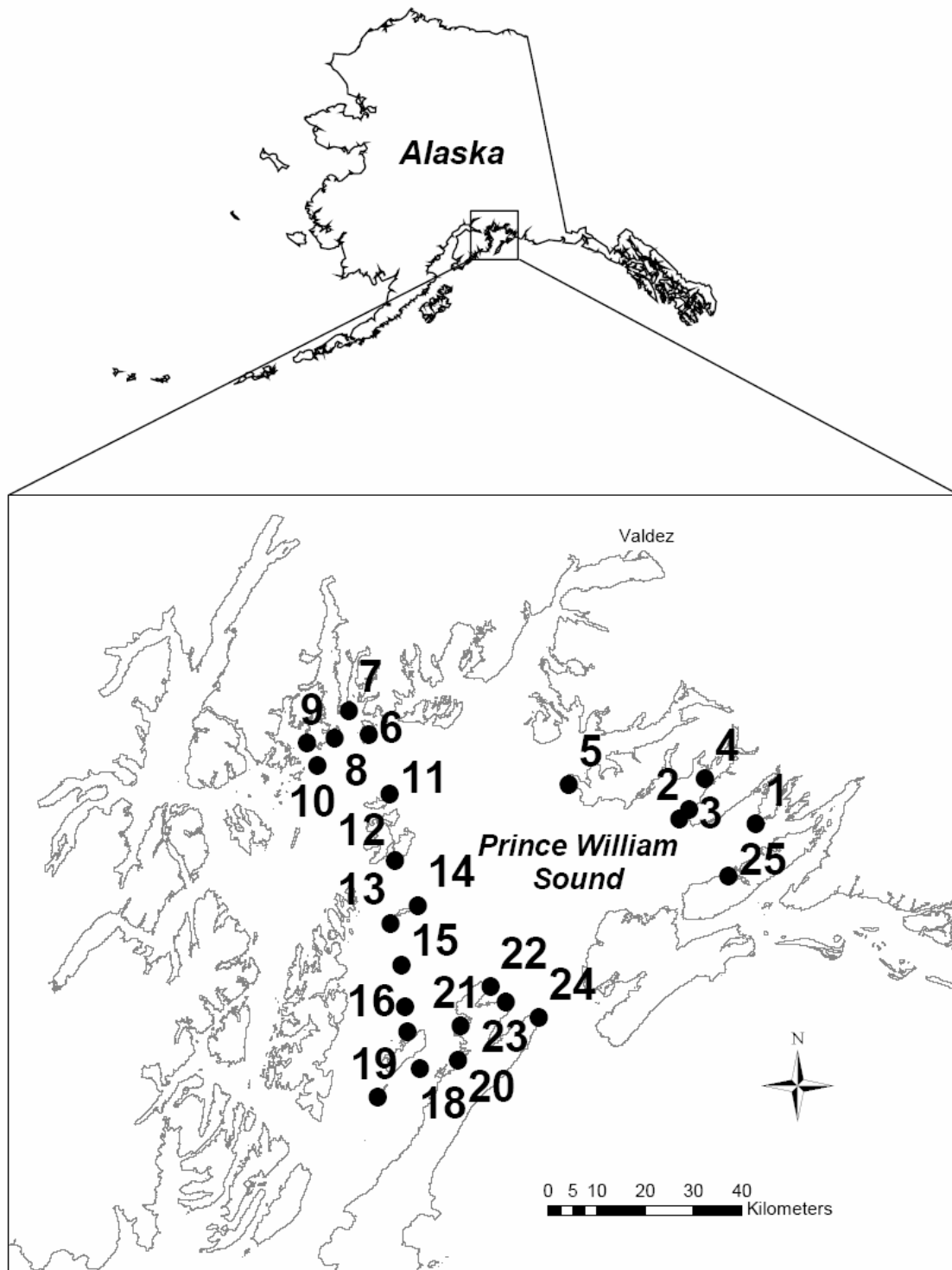
8 **Figure 2:** A schematic of the hierarchical statistical models for the harbor-seals study. (A): The
9 data model is the top level with the process model at the second level. A *Bayesian*
10 hierarchical statistical model would have a parameter model at a third level. (B): As in
11 (A), except the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

13

14 **Figure 3:** Posterior densities of slopes $\{\theta_i : i = 1, \dots, 25\}$, corresponding to the year (trend)
15 covariate, with the posterior density of θ_{1A} superimposed with the heavy solid line. These
16 posterior densities correspond to model in Figure 2(B). The solid circles on the x-axis
17 correspond to the peak densities for the model in Figure 2(A). The dashed lines link the
18 peaks for each site, and show how the Bayesian hierarchical model “shrinks” one-at-a-
19 time site estimates toward the global trend estimate.

20

21



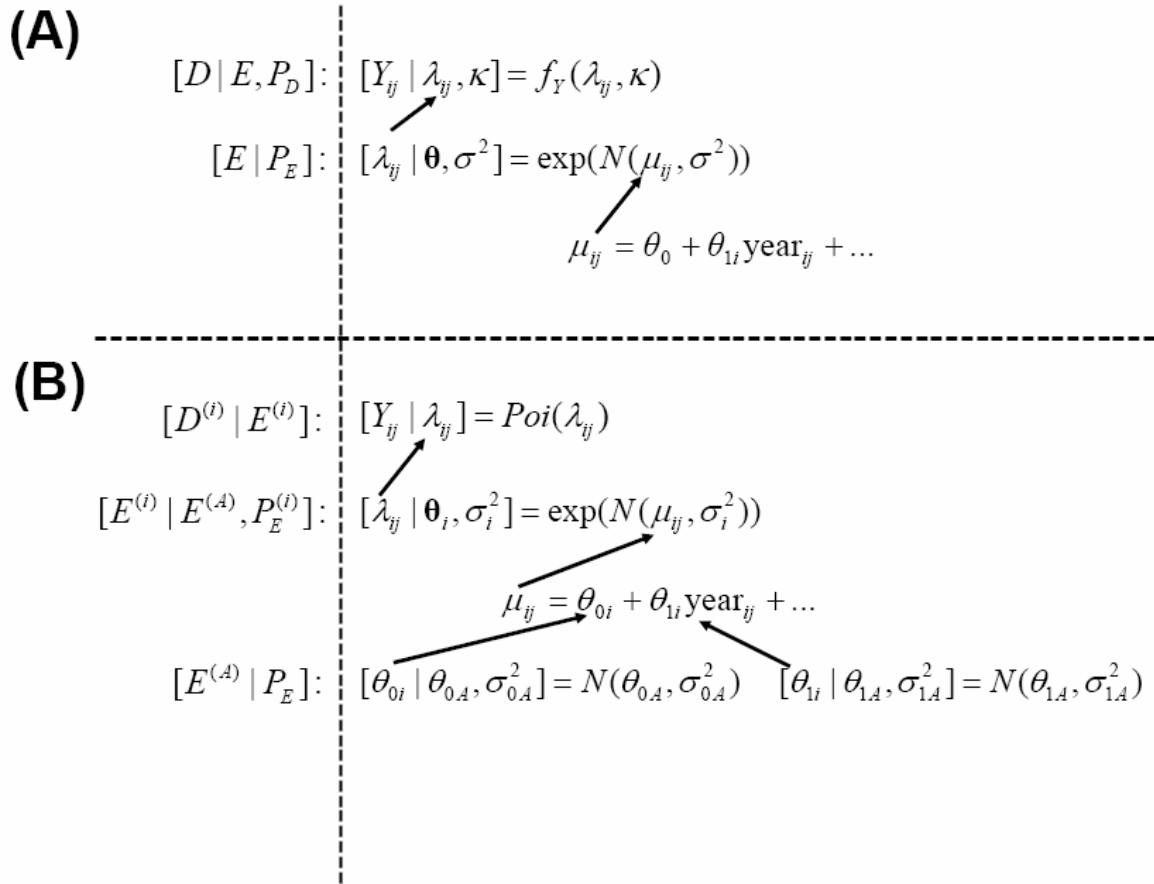
1

2

3

Figure 1

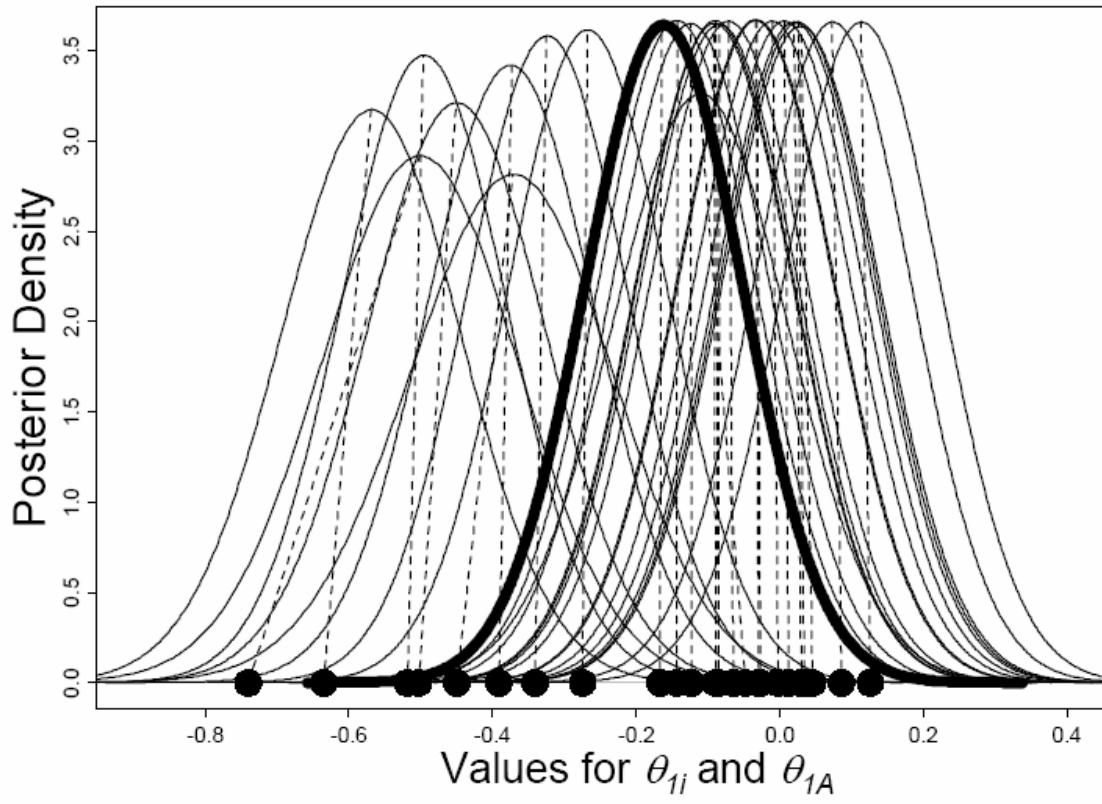
1
2
3
4
5



6
7
8
9
10
11

Figure 2

1
2
3
4



5
6
7
8
9
10
11

Figure 3