epEcology, 84(6), 2003, pp. 1395–1402

INCORPORATING MULTIPLE SOURCES OF STOCHASTICITY INTO DYNAMIC POPULATION MODELS

Catherine Calder,[1,4] Michael Lavine,[1] Peter Müller,[1,2] and James S. Clark[3]

[1]*Institute of Statistics and Decision Sciences, Duke University, Box 90251, Durham, North Carolina 27708 USA*
[2]*Department of Biostatistics, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030-4009 USA*
[3]*Department of Biology and Nicholas School of the Environment, Duke University, Durham, North Carolina 27708 USA*

Abstract. Many standard statistical models used to examine population dynamics ignore significant sources of stochasticity. Usually only process error is included, and uncertainty due to errors in data collection is omitted or not directly specified in the model. We show how standard time-series models for population dynamics can be extended to include both observational and process error and how to perform inference on parameters in these models in the Bayesian setting. Using simulated data, we show how ignoring observation error can be misleading. We argue that the standard Bayesian techniques used to perform inference, including freely available software, are generally applicable to a variety of time-series models.

Key words: Bayesian; Markov chain Monte Carlo (MCMC) simulation; nonlinear models; normal dynamic linear models; observation error; population dynamics; state-space; time-series; uncertainty.

Special Feature

Introduction

Population biologists use time-series data to infer the factors that regulate natural populations (Stenseth 1995, Stenseth et al. 1998, Bjørnstad et al. 1999) and to determine when populations may be at risk of extinction. Inference on the dynamics of natural populations is based on estimated model parameters and should incorporate the uncertainty in these parameters. Many analyses, however, ignore the multiple sources of stochasticity that commonly impact population time-series (Dennis and Taper 1994, Bjørnstad and Grenfell 2001). These analyses include process error, accounting for the fact that population dynamics is not a deterministic process and that the model for the process may be misspecified, but they often ignore uncertainty due to errors in the observations. In this way, the true abundances through time is an unobserved (''latent'') state variable on which the data provides imperfect information. DeValpine and Hastings (2002) extend standard population models to include both observation error and process error. Using maximum likelihood techniques, they show that ignoring observation error can lead to inaccurate estimates of population parameters. Their approach provides a more flexible method than classical error-in-variables models (Carpenter et al. 1994). We apply an alternative method for incorporating observation error in models of population dynamics, which is based on the Bayesian, rather than frequentist, paradigm of statistics. Frequentist methods, sometimes referred to in the statistical literature as classical methods, treat parameters as fixed, unknown constants. Alternatively, under the Bayesian paradigm, parameters are viewed as random variables; the data are used to update one's beliefs about the distribution of the model parameters. The coherence of the Bayesian method provides a straightforward way to account for observation error in addition to process error.

The state-space model framework provides a structure for extending time-series models to include both observation and process error. The data are assumed to arise from an unobserved state variable that represents the ''true'' dynamic process. This underlying variable evolves over time by a process model that explicitly models process error. The model for the relationship between the actual data and the state variable incorporates observation error. Bayesian state-space models with linear structure and normal error distributions allow entirely analytic results; we review the relevant expressions. We discuss extensions and alternative approaches for models with nonlinear structure and nonnormal error distributions, including Markov chain Monte Carlo (MCMC) posterior simulation. The coherence of the Bayesian paradigm allows for more straightforward inference than do maximum likelihood methods and numerical quadrature techniques required for classical inference. In the Bayesian setting, state variables can be treated as parameters, and full posterior inference can be performed as if they are time-invariant parameters. We describe how to derive full posterior distributions for the state variables/parameters and time invariant parameters. Throughout, we use models for density-dependent population dynamics as illustrative examples.

Bayesian statistics and Bayesian state-space models are not new to the ecology literature. Soudant et al.

Manuscript received 3 December 2001; revised 2 July 2002; accepted 18 July 2002; final version received 21 October 2002. Corresponding Editor: O. N. Bjørnstad. For reprints of this Special Feature, see footnote 1, p. 1349.

[4] E-mail: kate@stat.duke.edu

(1997) developed a Bayesian dynamic model for phytoplankton time series that allows for time-varying influence of the covariates. Cottingham and Schindler (2000) use a Bayesian dynamic linear model to model how phytoplankton respond to pulsed nutrient loading. State-space models have been used in the fisheries literature (e.g., see references in Millar and Meyer [2000*a*]). For example, Millar and Meyer (2000*a*) and Meyer and Millar (2000*b*) introduce a Bayesian nonlinear state-space model to incorporate more realism into fish stock assessment models.

In this article, we show how standard population models can be extended to the state space framework in order to include multiple sources of error. Rather than focusing on a particular model, we concentrate on general techniques necessary to fit Bayesian state-space models. The relative simplicity of the methods described here have broad application and lead to statements of uncertainty that take a more comprehensive accounting of variability.

State-space Modeling

The standard models used to describe changes in population size have the form $y_t = G(D_{t-1}) + w_t$, where y_t is a function of the size of the population at time t and D_t represents all of the observed population counts up until time t, i.e., $D_t = \{y_1, y_2, \ldots, y_{t-1}\}$. The function $G(\cdot)$ provides a deterministic relationship between the size of the population at time t and its size in the past. $G(\cdot)$ will typically also be a function of unknown parameters that can be estimated from data. The term w_t represents the process error and depends intrinsically on the function $G(\cdot)$; it accounts for the variability in the size of the population that cannot be captured by $G(\cdot)$.

Population census data come from trapping, counts, photographs, and so forth and are typically observed with measurement error. Counts are rarely equal to the true size of the population. The error term w_t in the model for population dynamics described in the previous paragraph subsumes both process error and measurement error. It is often useful to extend this model to a state-space model that explicitly separates multiple sources of stochasticity. DeValpine and Hastings (2002) describe how inference can be performed for state-space models within the frequentist framework. We focus on Bayesian state-space models. The Bayesian state-space model is based on the Kalman filter (Kalman 1960), which is a popular technique used in engineering and statistical quality control. While not inherently a Bayesian technique, the Kalman filter provides a method for forecasting that is consistent with the theory of Bayesian inference. Harrison and Stevens (1976) discuss the principles of Bayesian forecasting and its relationship to the Kalman filter. Meinhold and Singpurwalla (1983) present a less technical version of these issues.

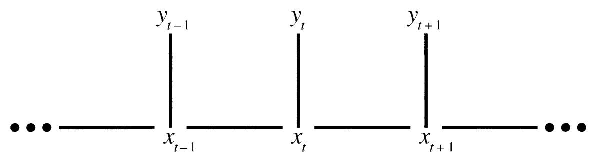

FIG. 1. This diagram represents the conditional independence structure of a state-space model. Each of the x's, given the values of the surrounding nodes, is conditionally independent of the rest of the graph.

The standard framework for a Bayesian state-space model is as follows:

Observation equation

$$y_t = F(\mathrm{x}_t) + \nu_t, \qquad \nu_t \sim \mathcal{N}[0, V]$$

Process equation

$$x_t = G(x_{t-1}) + w_t, \qquad w_t \sim \mathcal{N}[0, W]$$
$$x_0 = N[m_0, C_0]. \tag{1}$$

In Eq. 1, y_t is a function of the value of the time series at time t; x_t is an unknown underlying state variable, e.g., the log of population density at time t, that is propagated through time by the function $G(\cdot)$. It represents the true size of the population at time t. The function $F(\cdot)$ models the deterministic relationship between the underlying state variable and the observations, y_t. If the data are believed to be unbiased estimates of y_t, $F(\cdot)$ can be taken to be the identity (Coulson et al. 2001). The variable ν_t represents the measurement or observation error. It is modeled using a normal distribution with mean 0 and a known variance, V. This normality assumption is not necessary; it is possible under the Bayesian framework to perform inference assuming any error distribution. Generalization to unknown V is straightforward (West and Harrison 1997). Bjørnstad et al. (1999) suggest a Poisson counting process for the observation error. The variable w_t represents the process error. It is also usually assumed to come from a normal distribution with mean 0 and variance W. In the Bayesian framework, it is necessary to specify the prior distribution of the initial latent state variable x_0. We assume that it comes from a normal distribution with known mean, m_0, and variance, C_0.

Standard state-space models further assume that the error terms are conditionally independent. Fig. 1 shows the structure of a state-space model. Each of the x's given the values of the surrounding nodes is conditionally independent of the rest of the graph.

Normal dynamic linear models

When the functions $F(\cdot)$ and $G(\cdot)$ are linear and the error distributions are normal, the Bayesian state-space model is termed a normal dynamic linear model (NDLM). The functions $F(\cdot)$ and $G(\cdot)$ are replaced by the constants f and g that premultiply x_t in the observation equation and x_{t-1} in the process equation, re-

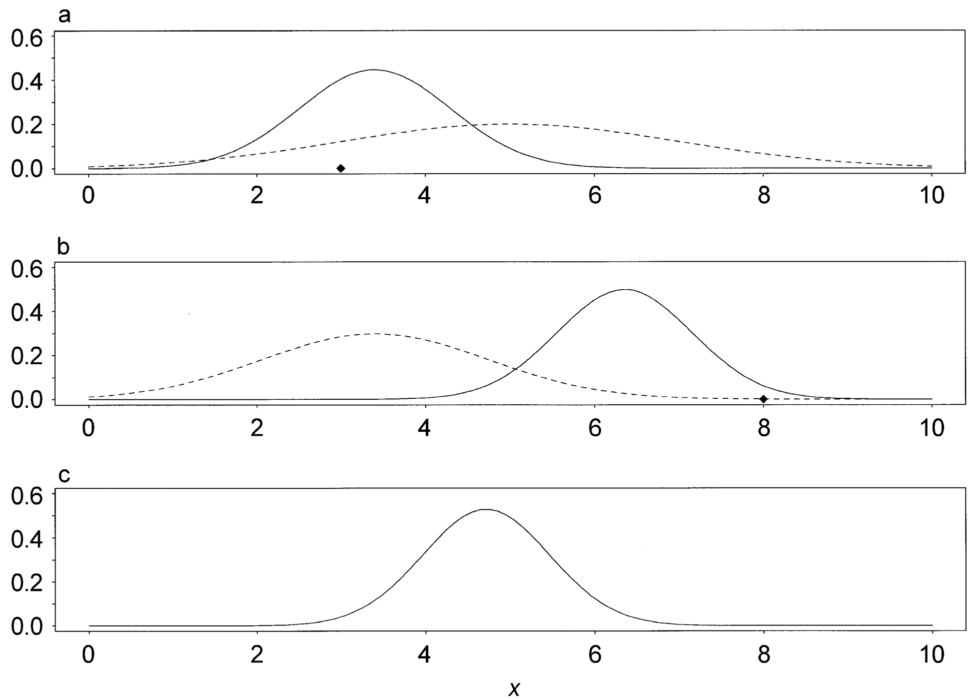

FIG. 2. This figure demonstrates how the forward filtering backward smoothing (FFBS) algorithm updates posterior distributions as it processes data sequentially. The details of the model and data are given in *State-space modeling: Normal dynamic linear models.* The two data points, y_1 and y_2, are represented by the black diamonds in (a) and (b), respectively. The lines represent the following distributions: (a) dashed line, $p(x_1)$; solid line, $p(x_1|y_1)$; (b) dashed line, $p(x_2|y_1)$; solid line, $p(x_2|y, y_2)$; (c) solid line, $p(x_1|y_1, y_2)$.

spectively. In the case when either the observation equation or the process equation is linear with an intercept term, the underlying state parameter is replaced by the vector $\{1, x_t\}$.

The posterior distribution of the x_t's in an NDLM can be found analytically by taking advantage of the model's conditional independence structure. As before, $D_t = \{y_1, y_2, \ldots, y_t\}$. The desired posterior distributions are $p(x_t|D_T)$ for $t = 1, 2, \ldots, T$, but for the moment we consider finding $p(x_t|D_t)$ for all t. Using Bayes' Theorem,

$$p(x_t|D_t) \propto p(y_t|x_t, D_{t-1})p(x_t|D_{t-1}) \tag{2}$$

where $p(y_t|x_t, D_{t-1})$ is the likelihood of y_t given all of the past data and $p(x_t|D_{t-1})$ is prior density of x_t conditional on the past data values.

Given the linear structure of the process equation, the fact that $p(x_{t-1}|D_{t-1})$ is normal allows $p(x_t|D_{t-1})$, the second factor on the right side of Eq. 2, to be found in closed form. This is done simply by updating the moments of $p(x_{t-1}|D_{t-1})$ according to the process equation. Once the next data point, y_t, is processed, the prior of x_t given D_{t-1} is updated to the posterior distribution $p(x_t|D_t)$.

In this manner, the distributions $p(x_t|D_t)$ are computed sequentially. We begin with $p(x_0)$, which we assume to be normal with mean m_0 and variance C_0. This procedure, known as the forward filtering algorithm, due to Carter and Kohn (1994) and Frühwirth-Schnatter (1994), is Theorem 4.1 in West and Harrison (1997).

The forward filtering algorithm defines a procedure for sequentially determining the posterior distribution of each of the x_t's given D_t. But we desire the posterior distributions $p(x_t|D_T)$ for $t = 1, 2, \ldots, T$, i.e., the posterior distributions of the latent variables given all of the data. These distributions can be found by recursively updating the moments of $p(x_T|D_T)$. The formulas for this backward smoothing algorithm are given in Theorem 4.4 of West and Harrison (1997). Together, these two algorithms are know as the forward filtering backward smoothing (FFBS) algorithm and are due to Carter and Kohn (1994) and Frühwirth-Schnatter (1994). It is a two-step algorithm. First, the forward filtering algorithm is run on the data. Then, using the final posterior distribution derived by the first algorithm, $p(x_T|D_T)$, the backward smoothing algorithm is used to recursively find $p(x_t|D_T)$ for all t. Both of these algorithms simply update the moments of normal distributions using the structure of the state-space model (Eq. 1). For details on the algorithm, see the appendix.

Fig. 2 demonstrates how the FFBS algorithm updates posterior distributions as it processes data sequentially. For simplicity, we assume that the data arise from a NDLM as in Eq. 1 with the function $F(\cdot)$ and $G(\cdot)$ taken to be the identity and $V = W = 1$.

Assume we have two data points, $y_1 = 3$ and $y_2 = 8$, and that our prior on x_0 is $\mathcal{N}(5, 3)$. The dashed line in Fig. 2a shows the prior distribution of x_1 given no data, i.e., $p(x_1) = \mathcal{N}(5, 4)$. The moments of this distribution are computed using the formulas in step b of the forward filtering (FF) algorithm. This distribution can be interpreted as the prior distribution for x_1. The point on this graph represents the first data point, $y_1 = 3$, and the solid line represents the posterior distribution of x_1 given the first data point, $p(x_1|y_1) = \mathcal{N}(3.4, 0.8)$. The moments of this distribution are calculated using step d of the FF algorithm. Since y_1 is less than the prior mean for x_1, and $F(\cdot)$ is the identity, y_1 pulls the distribution of x_1 to the left.

The dashed line in Fig. 2b represents the prior distribution of x_2 given only the first data point, $p(x_2|y_1) = \mathcal{N}(3.4, 1.8)$. Its moments are calculated by processing the posterior distribution of x_1 given y_1 through the process equation (using step b of the FF algorithm). Since $G(\cdot)$ is the identity, this process stretches out the distribution represented by the solid line in the first graph but keeps it centered at 3.4. This distribution can be thought of as the prior distribution of x_2 given only y_1 and can be updated to the posterior distribution of x_2 given both y_1 and y_2, $p(x_2|y_1, y_2)$ using step d of the FF algorithm. Since $y_2 = 8$, this posterior distribution is to the right of the prior distribution.

We have computed the posterior distribution for x_2 given all the data in this simple example, but we must use the backward smoothing (BS) algorithm to derive the posterior distribution of x_1 given all of the data. Using the BS algorithm, $p(x_1|y_1, y_2) = \mathcal{N}(4.7, 0.571)$. If we had more data, we would follow the same procedure adding an additional step in both the FF and BS algorithms for each data point.

Nonlinear dynamic models

Inference on the latent state variables in nonlinear dynamic models is slightly more complicated. The posterior distributions of the x_t's cannot be found in closed form as they are above. Instead, we rely on numerical techniques to explore their posteriors. The standard method for performing this inference is the Markov chain Monte Carlo (MCMC) algorithm that approximates the desired posterior distributions (Gilks et al. 1996). The algorithm constructs a Markov chain with the posterior distribution as its stationary distribution. The chain is run for sufficient time allowing it to approach its stationary distribution. Following convergence, samples approximate the posterior distribution.

The Gibbs sampler is a version of MCMC simulation that we use to approximate the posterior distributions of the x_t's. The algorithm works by iteratively sampling from each of the full conditional distributions of each of the states given the current value of the other variables. For example, if we want to sample from the joint distribution $p(A, B, C)$, we could construct a Gibbs sampler as follows:

1) sample $A^{(i)} \sim p(A|B^{(i-1)}, C^{(i-1)})$
2) sample $B^{(i)} \sim p(B|A^{(i)}, C^{(i-1)})$
3) sample $C^{(i)} \sim p(C|A^{(i)}, B^{(i)})$
4) repeat steps 1–3 many times, incrementing i after each iteration.

When sampling from these distributions, we condition on the currently imputed values, i.e., the most recent parameter values.

In the nonlinear dynamic model, each of the states, given the next state and the previous state, is independent of the other states. The conditional independence structure shown in Fig. 1 still applies. Thus, we need only condition on neighboring states. The Gibbs sampler for the nonlinear dynamics model works as follows:

1) Choose initial values for each of the latent state variables and denote them as $x_0^{(0)}, x_1^{(0)}, \ldots\ x_T^{(0)}$.
2) For each $1 \leq t \leq T$, sequentially draw a sample from

$$x_t^{(i)} \sim p(x_t \,|\, x_0^{(i)}, x_1^{(i)}, \ldots, x_{t-1}^{(i)}, x_{t+1}^{(i-1)}, \ldots, x_T^{(i-1)}, D_t) \quad (3)$$

$$\propto p(x_t \,|\, x_{t-1}^{(i)}, x_{t+1}^{(i-1)}, y_t) \quad (4)$$

$$\propto p(y_t \,|\, x_t) p(x_t \,|\, x_{t-1}^{(i)}) p(x_{t+1}^{(i-1)} \,|\, x_t) \quad (5)$$

for each t.
3) Repeat step 2, I times.

The Gibbs sampler above requires simulation from the distribution that is proportional to

$$p(y_t|x_t)p(x_t|x_{t-1}^{(i)})p(x_{t+1}^{(i-1)}|x_t).$$

This can be done using a Metropolis-Hastings (M-H) step, or in some cases a special case of M-H known as a Metropolis step (Gelfand and Smith 1990, Tierney 1994). See, for example, Gilks et al. (1996) for a general review of the algorithm. Carlin, Polson, and Stoffer (1992) develop an efficient M-H algorithm for sampling the latent state parameters in a nonlinear dynamic model. The samples $x_t^{(i)}$ for $i > B_x$, where B_x is the number of iterations until the Markov chain converges, will be samples from $p(x_t|D_T)$.

Inference for parametric state-space models

The posterior distribution of parameters of the model for population dynamics captured in $G(\cdot)$ are usually the focus of a time-series analysis. An extended Gibbs sampler can be used to find the posterior distributions of model parameters and the latent state variables simultaneously. After assigning initial values to all parameters, we can sample from the full conditional distribution of the model parameters given the currently imputed values of the x_t's. We then sample from the full conditional distributions of the x_t's conditioning on the currently imputed values of the model parameters.

The observation and process error variances, V and W, are often unknown. Their posterior distributions can

SPECIAL FEATURE

be recovered by sampling from their full conditional distributions within the Gibbs sampler. If a is a parameter of the function $G(\cdot)$, the Gibbs sampler iterates over the following steps:

1) Sample from $p(x_t|x_{t-1}, x_{t+1}, a, V, W, y_t)$ for $t = 1, 2, \ldots, T$ using the M-H algorithm or directly using FFBS.
2) Sample from $p(a|x_1, x_2, \ldots, x_T, V, W, D_T)$ using the M-H algorithm or, if possible, directly.
3) Sample from $p(V|x_1, x_2, \ldots, x_T, a, W, D_T)$ directly.
4) Sample from $p(W|x_1, x_2, \ldots, x_T, a, V, D_T)$ directly.

In each of the steps, the values of the variables to the right of the conditional bar are assumed to be the currently imputed values of each of the parameters. It is also necessary to specify prior distributions for a, V, and W if they are assumed to be unknown. These prior distributions must be included when determining each of the full conditional distributions. In the next section, we show the results of a simulation study for a state-space model.

Simulation Study

We use a simulation study, similar to the one included in deValpine and Hastings (2002), to demonstrate the importance of including observational error and additionally to demonstrate how this particular model, when framed as a Bayesian dynamic model, can be fit. We focus on two models, the observation-error model and the no-observation-error model. Both models are based on a common model for density dependence, the Ricker model (e.g., Dennis and Taper 1994). The observation-error model explicitly incorporates observation error into the Ricker model, while the no-observation-error model lumps process error and observation error together.

The observation-error model can be written as:

$$\log(y_t) = x_t + v_t \qquad v_t \sim \mathcal{N}[0, V]$$

$$x_t = x_{t-1} + a + be^{x_{t-1}} + w_t \qquad w_t \sim \mathcal{N}[0, W]$$

$$x_0 \sim \mathcal{N}[m_0, C_0] \qquad V \sim \mathrm{IG}[\alpha_V, \beta_V]$$

$$W \sim \mathrm{IG}[\alpha_W, \beta_W] \qquad p(a, b) \sim \mathcal{N}_2[\mu, \Sigma].$$

IG represents the inverse gamma distribution, and $\mathcal{N}_2$ represents the bivariate normal distribution.

The no-observation-error model can be written as:

$$y_t^* = y_{t-1}^* + a + be^{y_{t-1}^*} + w_t \qquad w_t \sim \mathcal{N}[0, W]$$

$$y_0^* \sim \mathcal{N}[m_0, C_0] \qquad W \sim \mathrm{IG}[\alpha_W, \beta_W]$$

$$p(a, b) \sim \mathcal{N}_2[\mu, \Sigma]$$

and y_t^* is defined to be $\log(y_t)$.

First, we simulate data from the observation-error model with the parameter a fixed at 0.1 and the parameter b fixed at -0.01. This is the ''correct'' model for the data. We then find the posterior distributions of the model parameter b under both models and compare them.

For the observation-error model, we use a Gibbs sampler to iteratively sample from the full conditional distributions of each of the parameters in the model. Each of the x_t's are sampled using a M-H step and the parameters a, b, V, and W are sampled directly given the current values of the x_t's using standard linear model theory. The full conditional distributions are

$$p(x_0|-) \propto \exp\left[-\frac{1}{2C_0}(x_0 - m_0)^2 - \frac{1}{2W}(x_1 - x_0 - a - be^{x_0})^2\right]$$

$$p(x_t|-) \propto \exp\left\{-\frac{1}{2V}[\log(y_t) - x_t]^2 - \frac{1}{2W}(x_t - x_{t-1} - a - be^{x_{t-1}})^2 - \frac{1}{2W}(x_{t+1} - x_t - a - be^{x_t})^2\right\}$$

$$\text{for } t = 1, 2, \ldots, T - 1$$

$$p(x_T|-) \propto \exp\left\{-\frac{1}{2V}[\log(y_T) - x_T]^2 - \frac{1}{2W}(x_T - x_{T-1} - a - be^{x_{T-1}})^2\right\}$$

$$p[(a, b)|-] \propto \mathcal{N}_2[m, S]$$

where

$$S = \left(\frac{1}{W}X'X + \Sigma^{-1}\right)^{-1} \qquad m = S\left(\frac{1}{W}X'Z + \Sigma^{-1}\mu\right)$$

and X and Z are defined to be

$$X = \begin{bmatrix} 1 & e^{x_0} \\ 1 & e^{x_1} \\ \vdots & \vdots \\ 1 & e^{x_{T-1}} \end{bmatrix} \qquad Z = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix} - \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{T-1} \end{bmatrix}$$

$$p(V|-) \propto \mathrm{IG}\left\{\alpha_V + \frac{T}{2}, \beta_V + \frac{1}{2}\sum_{t=1}^{T}[\log(y_t) - x_t]^2\right\}$$

$$p(W|-) \propto \mathrm{IG}\left[\alpha_W + \frac{T}{2}, \beta_W + \frac{1}{2}\sum_{t=1}^{T}(x_t - x_{t-1} - a - be^{x_{t-1}})^2\right]$$

We can sample the posterior distributions for a, b, and W directly under the no-observation-error model without the need for MCMC using standard results from linear regression. We generate samples from these posteriors of the same size as the samples generated from the MCMC algorithm used to analyze the observational error model.

We use three different values of V. For each value of V, we simulate 10 random data sets and summarize the resulting posterior distributions for b under both

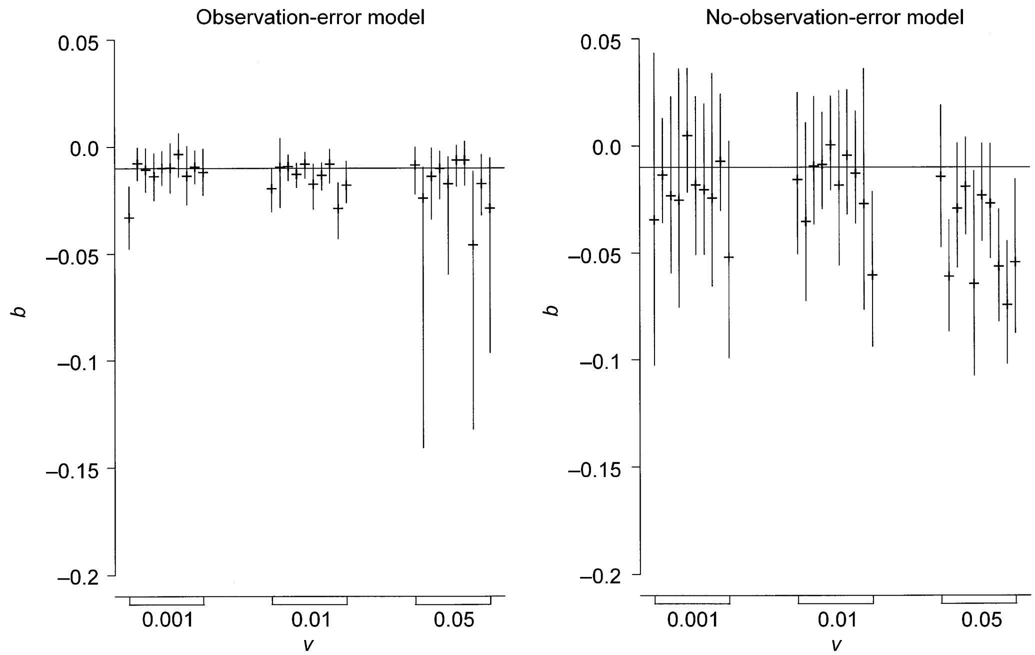

FIG. 3. The vertical lines represent 95% credible intervals of the posterior distribution of the parameter b under the two different models for 10 data sets simulated from the observation-error model. (A 95% Bayesian credible interval is the range from the 0.025 quantile to the 0.975 quantile of a posterior distributions. It can be approximated using samples from a posterior distribution.) The posterior medians are denoted by the dash through the vertical lines. This simulation study was repeated for $V = 0.001$, 0.01, and 0.05, and the corresponding posterior summaries are grouped according to the value of V used to simulate the data set. The true value of b is represented by the horizontal line at $b = -0.01$. For each of the trials, the value of W used to simulate the data was 0.01 and the length of the simulated data sets, T, was 60.

models (Fig. 3). Because we know that the observation-error model is the correct model for the data, differences in the posterior distributions demonstrate misleading inference that results from combining the two types of error. In this case, the posterior distributions under the observation-error model are mostly tighter than under the no-observation-error model. In other words, in the no-observation-error model, the two types of error are inappropriately combined, leading to overly broad credible intervals. It also appears from Fig. 3 that the posterior median becomes a considerably more biased estimate of b as the true value of V used to simulate the data increases. These results are not meant to be generalized; if we use a different model or different parameter values, we expect different results. The simulation study presented here demonstrates that for one particular example it makes a difference whether or not observation error is modeled explicitly.

TABLE 1. Summary of Bayes factors for 50 trials.

Bayes factor	Evidence against $H_{\text{no obs}}$	Number of trials
<1	none	0
1–3	not worth more than a bare mention	0
3–20	positive	0
0–150	strong	1
>150	very strong	49

Notes: The true values of the parameters in the Ricker model are $V = 0.01$, $W = 0.01$, $a = 0.1$, $b = -0.01$. The Bayes factor is computed by taking the ratio of the $p\,(\text{data}|H_{\text{obs}})$ to $p\,(\text{data}|H_{\text{no obs}})$. The interpretations of the Bayes factors are taken from Kass and Raftery (1995). H_{obs}, model with observation error; $H_{\text{no obs}}$, model without observation error.

In addition to comparing posterior parameters under the two models, we also computed Bayes factors, which assess the weight of the evidence in favor of one model as opposed to the other model. (See Kass and Raftery [1995] for an overview of the technique.) If H_1 represents the hypothesis that the data are generated by model 1 and H_2 represents the hypothesis that the data are generated by model 2, the Bayes factor B_{12} is defined as follows:

$$B_{12'} = \frac{p(\text{data}\,|\,H_1)}{p(\text{data}\,|\,H_2)}. \tag{6}$$

The Bayes factor represents the ratio of the posterior odds of hypothesis H_1 to its prior odds. Kass and Raftery (1995) illustrate how to calculate and interpret Bayes factors in various settings. A recipe for approximating Bayes factors using output from an MCMC simulation is due to Newton and Raftery (1994). The numerator and denominator of the Bayes factor, Eq. 6, can be approximated as

$$\hat{p}(\text{data}\,|\,H.) = \left\{\frac{1}{m}\sum_{i=1}^{m} p[\text{data}\,|\,\theta^{(i)}, H.]^{-1}\right\}^{-1}.$$

This is the harmonic mean of the likelihood of the data, given the samples of the parameters taken in the MCMC algorithm. It can be shown that $p(\text{data}|H_j)$ converges almost surely to $p(\text{data}|H_j)$ as $m \rightarrow \infty$. Table 1 summarizes the Bayes factors computed in this way for 50 trials; the two models are fit and the Bayes factor comparing them is computed for each trial. Also provided in the table are interpretations of the Bayes factors. Table 1 shows that all but one of the 50 trials provided strong positive evidence against the null hypothesis, i.e., against the hypothesis that the data were generated from the no-observation-error model.

Discussion

State-space models provide a framework for incorporating observational error into dynamic models of population size. Bayesian state-space models and procedures used to perform inference on them are general enough to be used on more complicated models than the one discussed here, including those having higher embedding dimension and thresholds. We need only to iteratively sample from the full conditional distributions of each model parameter, which can be done directly using the Metropolis-Hastings algorithm.

When performing a statistical analysis on data containing observation error, we recommend considering a state-space model because it explicitly separates process error from observation error and therefore affords a more detailed understanding of the posterior distribution. Our simulation study illustrates that ignoring observational error can result in misleading inference, especially in estimates of posterior uncertainty of model parameters. The results of this specific simulation study are not general. For others models, the results of ignoring observation error will be different; we simply show that explicitly modeling observation error can make a difference.

There is freely available software that can be used to fit Bayesian state-space models. BUGS (Spiegelhalter et al. 1996), Bayesian inference using Gibbs sampling, takes advantage of the acyclical graphical structure of a Bayesian model to determine the full conditional distributions required to construct a Gibbs sampler. See Meyer and Millar (1999) for an illustration of the BUGS software in the context of fitting a nonlinear Bayesian state-space model to analyze fisheries stock assessments. Also freely available is the BATS software, which fits a variety of Bayesian dynamic models as described in West and Harrison (1997). Instructions for this package are available in Pole et al. (1994). General purpose statistical software packages can also be used to fit state-space models.

We propose fitting Bayesian state-space models as opposed to classical state-space when it is necessary to explicitly include observation error in a time-series model. The coherence of Bayesian paradigm allows straightforward inference; both the time-invariant parameters and the underlying variables can be treated in the same way. This is true regardless of how the analysis is performed.

Literature Cited

Bjørnstad, O. N., J.-M. Fromentin, N. C. Stenseth, and J. Gjøsæter. 1999. Cycles and trends in cod populations. Ecology **96**:5066–5071.

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

Carlin, B. P., N. G. Polson, and D. S. Stoffer. 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. Journal of the American Statistical Association **87**:493–500.

Carpenter, S. R., K. L. Cottingham, and C. A. Stow. 1994. Fitting predator–prey models to time series with observation errors. Ecology **75**:1254–1264.

Carter, C. K., and R. Kohn. 1994. On Gibbs sampling for state space models. Biometrika **81**(3):541–553.

Cottingham, K., and D. E. Schindler. 2000. Effects of grazer community structure on phytoplankton response to nutrient pulses. Ecology **81**:183–200.

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

Dennis, B., and M. L. Taper. 1994. Density dependence in time series observations of natural populations—estimation and testing. Ecological Monographs **64**:205–224.

deValpine, P., and A. Hastings. 2002. Fitting population models with process noise and observation error. Ecological Monographs **72**:57–76.

Frühwirth-Schnatter, S. 1994. Data augmentation and dynamic linear models. Journal of Time Series Analysis **15**: 183–202.

Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association **85**:129–133.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Markov chain Monte Carlo in practice. Chapman and Hall, New York, New York, USA.

Harrison, P. J., and D. F. Stevens. 1976. Bayesian forecasting (with discussion). Journal of the Royal Statistical Society, Series B **38**:205–247.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering, Series B **82**:34–45.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association **90**(430):773–795.

Meinhold, R. J., and N. D. Singpurwalla. 1983. Understanding the Kalman filter. American Statistician **37**(2):123–127.

Meyer, R., and R. B. Millar. 1999. BUGS in Bayesian stock assessment. Canadian Journal of Fisheries and Aquatic Sciences **56**:1078–1086.

Millar, R. B., and R. Meyer. 2000*a*. State-space modeling of non-linear fisheries biomass dynamics using the Gibbs sampler. Applied Statistics **49**:327–342.

Millar, R. B., and R. Meyer. 2000*b*. Bayesian state-space modeling of age structured data: fitting a model is just the beginning. Canadian Journal of Fisheries and Aquatic Sciences **57**:43–50.

Newton, M. A., and A. E. Raftery. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). Journal of the Royal Statistical Society, Series B **56**:3–48.

Pole, A., M. West, and P. J. Harrison. 1994. Applied Bayesian forecasting and time series analysis. Chapman and Hall, New York, New York, USA.

Soudant, D., B. Beliaeff, and G. Thomas. 1997. Dynamic linear Bayesian models in phytoplankton ecology. Ecological Modelling **9**:161–169.

Spiegelhalter, D. J., A. Thomas, N. Best, and W. R. Gilks. 1996. BUGS 0.5, Bayesian inference using Gibbs sampling. Manual version ii. Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, UK.

Stenseth, N. C. 1995. Snowshoe hare populations—squeezed from below and above. Science **269**:1061–1062.

Stenseth, N. C., W. Falck, K. Chan, O. N. Bjørnstad, M. O'Donoghue, H. Tong, R. Boonstra, S. Boutin, C. Krebs, and N. G. Yoccoz. 1998. From patterns to processes: phase and density dependencies in the Canadian lynx cycle. Proceedings of the National Academy of Sciences (USA) **95**: 15430–15435.

Tierney, L. 1994. Markov chains for exploring posterior distributions. Annals of Statistics **22**:1701–1728.

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

APPENDIX

The forward filtering backward smoothing algorithm is available in ESA's Electronic Data Archive: *Ecological Archives* E084-033-A1.