

# Incorporating Multiple Sources of Stochasticity into Dynamic Population Models

Catherine A. Calder<sup>1,\*</sup>, Michael Lavine<sup>1</sup>, Peter Müller<sup>1,2</sup>, James S. Clark<sup>3</sup>

April 4, 2002

<sup>1</sup>*Institute of Statistics and Decision Sciences, Duke University, Box 90251, Durham, North Carolina 27708 USA*

<sup>2</sup>*Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030-4009 USA*

<sup>3</sup>*Department of Biology & Nicholas School of the Environment, Duke University, Durham, North Carolina 27708 USA*

---

\*E-mail: [kate@stat.duke.edu](mailto:kate@stat.duke.edu)

*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; MCMC; Normal Dynamic Linear Models; Nonlinear Models; Observation Error; State-Space; Time-Series; Uncertainty*

## Introduction

Population biologists use time-series data to infer the factors that regulate natural populations (Bjørnstad et al. 1999, Stenseth 1995, Senseth et al. 1998) and to determine when populations may be at risk. 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, O.N. 2001). These analyses include process error, accounting for the fact the parameterized model does not accurately describe true process, but ignore the contribution to parameter uncertainty due to error in the observed data. It has been shown by deValpine and Hastings (2001) that standard population models can be extended to include both observation error and process error. Using maximum likelihood techniques, they show that inference procedures that ignore observation error can lead to inaccurate estimates of population parameters. Their approach provides a more flexible method than classical error-in-variables models (Carpender et al. 1994). We expand upon the work of deValpine and Hastings by describing how the Bayesian paradigm of statistics can be used to incorporate observation error in models of population dynamics. We hold that the coherence of the Bayesian method provides a straightforward way to account for multiple sources of uncertainty.

The state-space model framework provides a structure for extending time-series models to include both observation and process error. The data is assumed to depend on an unknown time-varying state variable which represents the “true” 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 non-linear structure and non-normal error distributions, including Markov chain Monte Carlo (MCMC) posterior simulation. The coherence of the Bayesian paradigm allows inference on the state-variables to be more straightforward than maximum likelihood methods and numerical quadrature techniques required for classical inference. In the Bayesian setting, the state variables can be treated as parameters and full posterior inference on them can be performed just 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. (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. Classical state-space models have been used frequently in the fisheries literature (e.g., see references in Millar and Meyer (2000a)). Millar and Meyer (2000a) and Meyer and Millar (2000b) introduce a Bayesian nonlinear state-space models to incorporate more realism into fish stock assessment models.

In this article, we show how all standard population models can be extended to

the state-space framework in order to include multiple sources of error. Rather than focusing on performing a detailed analysis of a particular time-series, we concentrate on presenting the techniques necessary to fit Bayesian state-space models. We hope to convince readers that the relative simplicity of the methods described here can lead to general models with broad application and to statements of uncertainty that take a more comprehensive accounting of variability.

## State-Space Modeling

The standard models used to describe changes in the size of populations over time have the form  $x_t = G(D_{t-1}) + \omega_t$ , where  $x_t$  is the size of the population at time  $t$  and  $D_t$  represents all of the observed population counts up until time  $t$ , i.e.  $D_t = \{x_1, x_2, \dots, x_{t-1}\}$ . 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 the data. The estimated values of these parameters provide insight on the behavior of a population regulation. For example, the estimated value of a parameter of a density-dependence model may indicate that the growth rate depends on population size. The term  $\omega_t$  represents the process error and is intrinsically related to 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 are typically observed with measurement error. These data can come from trappings, observed counts, photographs, and so forth, all of which may be inaccurate. Observed counts are rarely equal to the true size of the popula-

tion. The model for population dynamics described in the previous paragraph does not incorporate this type of stochasticity; it allows only process error. This model, however, can be extended to a state-space model which allows multiple sources of stochasticity. De Valpine and Hastings (2001) describe how inference can be performed for state-space models within the classical 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 which 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), in an expository article, present a less technical version of these issues.

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

$$\begin{aligned}
 \text{Observation Equation:} \quad & y_t = F(x_t) + \nu_t, & \nu_t &\sim N[0, V] \\
 \text{Evolution Equation:} \quad & x_t = G(x_{t-1}) + \omega_t, & \omega_t &\sim N[0, W] \\
 & x_0 = N[m_0, C_0]
 \end{aligned} \tag{1}$$

where

- $y_t$  is the value of the time-series at time  $t$ .
- $x_t$  is an unknown underlying state variable 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 true size of the population and what is observed. If the data are believed to be unbiased,  $F(\cdot)$  can be taken to be the identity (Coulson et al. 2001).
- The variable  $\nu_t$  represents the measurement or observation error. It is modeled using a normal distribution with a mean of 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 also straightforward (West and Harrison 1997).
- The function  $G(\cdot)$  is the same as above. It describes the relationship between the size of the population at time  $t$  to the size of the population in the past.
- The variable  $\omega_t$  represents the model or 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 independent.

Figure 1 shows the conditional independence structure of a NDLM. 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, 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 evolution equation respectively. In the case when either the observation equation or the evolution equation is linear with an intercept term, the underlying state parameter can be written as  $\{1, x_t\}$  and the following discussion on posterior analysis for NDLMs is applicable.

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

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

Given the linear structure of the evolution equation, the fact that  $p(x_{t-1}|D_{t-1})$  is normal allows  $p(x_t|D_{t-1})$ , the second term on the right side of equation 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 evolution equation. The distribution of  $x_t$  given  $D_{t-1}$  can be viewed as the prior distribution for  $p(x_t|D_t)$ . Once the next data point,  $y_t$ , is processed, this prior can be updated to the posterior distribution  $p(x_t|D_t)$ .

In this manner, the distributions  $p(x_t|D_t)$  can be computed sequentially given  $p(x_0)$  which we assume in an NDLM to be normal with mean equal to  $m_0$  and variance

equal to  $C_0$ . This procedure, known as the Forward Filtering algorithm 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  $D_t$ . The desired posterior distributions, however, are  $p(x_t|D_T)$  for  $t = 1, 2, \dots, T$ , 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 known as the Forward Filtering Backward Smoothing (FFBS) algorithm. 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 (equation 1).

### Forward Filtering Backward Smoothing Algorithm

**1) Forward Filtering Algorithm:** The posterior and one-step forecast distribution in the NLDM can be calculated as follows:

a) Posterior at time  $t - 1$ :

For some mean  $m_{t-1}$  and variance  $C_{t-1}$ ,

$$x_{t-1}|D_{t-1} \sim N[m_{t-1}, C_{t-1}].$$

b) Prior at time  $t$ :

$$x_t | D_{t-1} \sim N[a_t, R_t],$$

where

$$a_t = Gm_{t-1} \text{ and } R_t = GC_{t-1}G' + W.$$

c) One-step forecast:

$$y_t | D_{t-1} \sim N[f_t, Q_t].$$

where

$$f_t = F_t' a_t \text{ and } Q_t = F_t' R_t F_t + V.$$

d) Posterior at time  $t$ ,

$$x_t | D_t \sim N[m_t, C_t],$$

with

$$m_t = a_t + A_t e_t \text{ and } C_t = R_t - A_t A_t' Q_t,$$

where

$$A_t = R_t F_t Q_t^{-1} \text{ and } e_t = y_t - f_t.$$

**2) Backward Smoothing Algorithm:** Given that  $p(x_t | D_t) \sim N[m_t, C_t]$ , for all  $k$

such that  $1 \leq k \leq t$ , the filtering marginal distributions are

$$x_{t-k} | D_t \sim N[a_t(-k), R_t(-k)]$$

where given that  $B_t = C_t G' R_{t+1}^{-1}$

$$a_t(-k) = m_{t-k} + B_{t-k}[a_t(-k+1) - a_{t-k+1}]$$

and

$$R_t(-k) = C_{t-k} + B_{t-k}[R_t(-k+1) - R_{t-k+1}]B'_{t-k}$$

with starting values

$$a_t(0) = m_t \text{ and } R_t(0) = C_t,$$

and where

$$a_{t-k}(1) = a_{t-k+1} \text{ and } R_{t-k}(1) = R_{t-k+1}.$$

Figure 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 equation 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  $N(5, 3)$ . The dashed line in the first graph of Figure 2 shows the prior distribution of  $x_1$  given no data, i.e.  $p(x_1) = N(5, 4)$ . 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,  $p(x_1|y_1) = N(3.4, 0.8)$ . 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 the second graph of represents the prior distribution of  $x_2$  given only the first data point,  $p(x_2|y_1) = N(3.4, 1.8)$ . Its moments are calculated by processing the posterior distribution of  $x_1$  given  $y_1$  through the evolution 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,  $p(x_1|y_1, y_2) = N(0.8, 0.643)$ . If we had more data, we would follow the same procedure adding an additional step in both the FF and BS algorithms for every 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. After the chain is allowed to run for a sufficient amount of time to approach its stationary distribution, samples are taken which can be used to approximate the posterior distribution.

The Gibbs sampler is a version of MCMC simulation that we will 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:

- sample  $A \sim p(A|B, C)$
- sample  $B \sim p(B|A, C)$
- sample  $C \sim p(C|A, B)$
- repeat steps 1-3 many times

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 1 still applies. This reduces the conditional distributions; we need only condition on neighboring states. The Gibbs sampler for the nonlinear dynamics models works as follows:

1. Choose initial values for each of the latent state variables and denote them as

$$x_0^{(0)}, x_1^{(0)}, \dots, x_T^{(0)}.$$

2. For each  $1 \leq t \leq T$ , sequentially draw a sample from

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

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

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

for each  $t$ .

3. Repeat step 2, I times.

The Gibbs sampler above requires simulation from the distribution which is proportional  $p(y_t | x_t) p(x_t | x_{t-1}^{(i)}) p(x_{t+1}^{(i-1)} | x_t)$ . This can be done using a Metropolis-Hastings (M-H) step (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  $i > \text{BURNIN}$ , where BURNIN is the number of iterations until the Markov chain converges, will be samples from  $p(x_t | D_T)$ .

## Linear Approximation

For the state-space models used here, the observation equation is linear; it simply adds observation error to established models for population dynamics. The nonlinearity is most likely to occur in the evolution equation; i.e., the function  $G(\cdot)$  is nonlinear. By taking the Taylor series expansion of  $G(\cdot)$  about the prior mean for  $x_t$ ,  $m_{t-1}$ , it is possible to linearize the evolution equation and eliminate the need for MCMC.

**Linearized Evolution Equation** (using a 1st order Taylor series approximation):

$$\begin{aligned} x_t &\approx G^*(x_{t-1}) + \omega_t \\ &= G(m_{t-1}) + \left. \frac{dG}{dx_{t-1}} \right|_{m_{t-1}} (x_{t-1} - m_{t-1}) + \omega_t \end{aligned} \quad (6)$$

The approximate posterior distributions of the  $x_t$ 's can be found in closed form using  $G^*(\cdot)$  in the FFBS algorithm discussed above.

## 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 any model parameters and the latent state variable 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.

We can also allow both the observation and evolution error variances,  $V$  and  $W$ , to be unknown. Their posterior distributions can 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 between the following steps.

1. Sample from  $p(x_t|x_{t-1}, x_{t+1}, a, V, W, D_T)$  for  $t = 1, 2, \dots, T$  using the M-H algorithm or directly using FFBS.
2. Sample from  $p(a|x_1, x_2, \dots, x_T, V, W, D_T)$  using the M-H algorithm or directly, if possible.
3. Sample from  $p(V|x_1, x_2, \dots, x_T, a, W, D_T)$  directly.
4. Sample from  $p(W|x_1, x_2, \dots, x_T, a, V, D_T)$  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. The form of each of the necessary conditional distributions for the Gibbs sampler are included.

## Simulation Study

We use a simulation study, similar to the one included in deValpine and Hastings (2001), that demonstrates the importance of including observational error and additionally demonstrates 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 (Dennis and Taper 1994). This model serves as the evolution equation in both cases. The observation error model, however, incorporates observation error while the no observation error model considers only process error.

### Observation Error Model

$$\log(y_t) = x_t + \nu_t, \quad \nu_t \sim N[0, V]$$

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

$$x_0 \sim N[m_0, C_0]$$

$$V \sim IG[\alpha_V, \beta_V]$$

$$W \sim IG[\alpha_W, \beta_W]$$

$$p(a, b) \sim N_2[\mu, \Sigma]$$

IG represents the inverse gamma distribution and  $N_2$  represents the bivariate normal distribution.

### No Observation Error Model

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

$$x_0 \sim N[m_0, C_0]$$

$$W \sim IG[\alpha_W, \beta_W]$$

$$p(a, b) \sim N_2[\mu, \Sigma]$$

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 model parameters  $a$  and  $b$  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 conditionals needed are listed below.

- $p(x_0|-) \propto \exp(-\frac{1}{2C_0}(x_0 - m_0)^2 - \frac{1}{2W}(x_1 - x_0 - a - be^{x_0})^2)$
- $p(x_t|-) \propto \exp(-\frac{1}{2V}(\log(y_t) - x_t)^2 - \frac{1}{2W}(x_t - x_{t-1} - a - be^{x_{t-1}})^2 - \frac{1}{2V}(x_{t+1} - x_t - a - be^{x_t})^2)$  for  $t = 1, 2, \dots, T - 1$
- $p(x_T|-) \propto \exp(-\frac{1}{2V}(\log(y_T) - x_T)^2 - \frac{1}{2W}(x_T - x_{T-1} - a - be^{x_{T-1}})^2)$

- $p((a, b)|-) \propto N_2[m, S]$

where  $S = (\frac{1}{W}X'X + \Sigma^{-1})^{-1}$ ,  $m = S(\frac{1}{W}X'Y + \Sigma^{-1}m)$

and  $X$  and  $Y$  are defined to be

$$X = \begin{bmatrix} 1 & e^{x_0} \\ 1 & e^{x_1} \\ \vdots & \vdots \\ 1 & e^{x_T} \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix} - \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{T-1} \end{bmatrix}$$

- $p(V|-) \propto IG[\alpha_V + T/2, \beta_V + \frac{1}{2} \sum_{t=1}^T (\log(y_t - x_t)^2)]$
- $p(W|-) \propto IG[\alpha_W + T/2, \beta_W + \frac{1}{2} \sum_{t=1}^T (x_t - x_{t-1} - a - be^{x_{t-1}})^2]$

We can sample the posterior distributions for  $a$ ,  $b$ ,  $V$ , and  $W$  directly under the no observational error model without the need for MCMC using the standard results from linear regression discussed above. 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 repeat this procedure for 50 random data sets and summarize our results with boxplots of posterior distributions for each of the model parameters. Because we know that the observational error model is the correct model for the data, differences in boxplots demonstrate misleading inference that results from ignoring observational error. The boxplots for the parameter  $a$  are shown in Figure 3. The posterior distribution under the no observational model are much larger than the posterior distributions under the correct model signifying that we underestimate the precision with which

we can estimate  $a$  by ignoring observational error. The results are switched for the parameter  $b$ , shown in Figure 4. In this case, the posterior distributions under the no observation error model are much smaller. This means that when performing posterior analysis on the parameter  $b$  while ignoring observational error leads to an overestimate of the precision with which the parameter is being estimated. In some of the trials, the boxplots for  $b$  under the no observation error model do not even overlap the true value.

In addition to comparing posterior parameters under the two models, we also computed Bayes factors which assess the amount of scientific evidence provided by the data have in favor of one model as opposed to another 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:

$$B_{12} = \frac{p(\text{data}|H_1)}{p(\text{data}|H_2)}. \quad (7)$$

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 Bayes factors in various settings as well as provide guidelines for interpreting them. Also provided in this article is a recipe for approximating Bayes factors using output from an MCMC simulation which is due to Newton and Raftery (1994). The numerator and denominator

of the Bayes factor, equation 7, can be approximated as

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

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  $\hat{p}(data|H_j)$  converges almost surely to  $p(data|H_j)$  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.

## Discussion

The results of the simulation study are not general; for other models the results of ignoring observation error will be different. We simply show that it can make a difference.

Ignoring observational error can result in misleading inference, especially in estimates of posterior uncertainty of model parameters. When performing a statistical analysis on data believed to be observed with observation error, we recommend considering a state-space model. If the desired inference is only a point estimate of model parameters or if for some other reason capturing the uncertainty in model parameters is not important, then perhaps the state-space model is not necessary. In these cases, the time-series model that only includes process error with most likely will be simpler

and less time-consuming to fit so there is no reason to complicate the analysis. Also, if the mechanism for the observation error is poorly understood, then perhaps in certain cases it would be advisable to ignore observation error rather than misspecifying it. However, if a more detailed understanding of the posterior or predictive uncertainty is required, then one should explicitly include observation error using a state-space model.

Bayesian state-space models provide a framework for incorporating observational error into dynamic models of population size. These 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 also be done using the Metropolis-Hastings algorithm. There is also 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 for run 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 is available in Pole et al. (1994) and the software is can be found at <ftp://stat.duke.edu/pub/bats>.

We propose fitting Bayesian state-space models as opposed to classical state-space

when necessary. The coherence of Bayesian paradigm allows inference for Bayesian state-space models to be straightforward; both the time-invariant parameters and the underlying variables can be treated in exactly the same way. This is true no matter how the computations required for inference are performed whether it be an MCMC analysis programmed specifically to analyze a particular model or using a software package such as BUGS or BATS.

## Literature Cited

- Bjørnstad, O.N., Fromentin, J.-M., Stenseth, N.C., Gjøsaeter, J. (1999) Cycles and trends in cod populations. *Ecology*, **96** 5066-5071.
- Bjørnstad, O.N., Grenfell, B.T. (2001) Noisy Clockwork: Time Series Analysis of Population Fluctuations in Animals. *Science*, **293** 638-643.
- Carlin, Bradley P., Polson, Nicholas G. and Stoffer, David S. (1992) A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, **87** 493-500.
- Carpenter, S.R., Cottingham, K.L., Stow, C.A. (1994) Fitting Predator-Prey Models to Time Series with Observation Errors. *Ecology*, **75** 1254-1264.
- Cottingham, Kathryn, Schindler, Daniel E. (2000) Effects of grazer community structure on phytoplankton response to nutrient pulses. *Ecology*, **81**(1) 183-200.
- Coulson, T., Catchpole, E.A., Albon, S.D., Morgan, B.J.T., Pemberton, J.M., Clutton-Brock, T.H., Crawley, M.J., Grenfell, B.T. (2001) Age, Sex, Density, Winter Weather, and Population Crashes in Soay Sheep. *Science*, **292** 1528-1531.
- de Valpine, P., Hastings, A. (2001) Fitting Population Models with Process Noise and Observation Error. *Ecology*, in press.
- Dennis B., Taper, M. L. (1994) Density dependence in time series observations of natural populations - estimation and testing. *Ecological Monographs*, **64** 205-224.
- Gelfand, A.E., Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal

- densities. *Journal of the American Statistical Association*, **85** 129-133.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (1995) *Bayesian Data Analysis*.  
New York: Chapman & Hall/CRC.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*. New York: Chapman & Hall.
- Harrison, P.J., Stevens, D.F. (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., Raftery, A.E. (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**(430) 773-795.
- Meinhold, R.J., Singpurwalla, Nozer D. (1983) Understanding the Kalman Filter. *The American Statistician*, **37**(2) 123-127.
- Meyer, Renate and Millar, Russell B. (1999) BUGS in Bayesian stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, **56** 1078-1086.
- Millar, Russell B., Meyer, Renate (2000a) State-space modeling of non-linear fisheries biomass dynamics using the Gibbs sampler. *Applied Statistics*, **49** 327-342.
- Millar, Russell B., and Meyer, Renate (2000b): 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., Raftery, A.E. (1994) Approximate Bayesian Inference by the Weighted

- Likelihood Bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B*, **56** 3-48.
- Pole, A., West, M., Harrison, P.J. (1994) Applied Bayesian Forecasting and Time Series Analysis. New York: Chapman & Hall.
- Robert, C.P. and Casella, G. (1999) *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- Soudant, D., Beliaeff, B., Thomas, G. (1997) Dynamic linear Bayesian models in phytoplankton ecology. *Ecological Modelling*, **99** 161-169.
- Spiegelhalter, D.J., Thomas, A., Best, N., Gilks, W.R. (1996) BUGS 0.5, Bayesian inference using Gibbs sampling. Manual version ii. Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, England.
- Stenseth, N.C. (1995) Snowshoe hare populations - squeezed from below and above. *Science*, **269** 1061-1062.
- Stenseth, N.C., Falck, W., Chan, K., Bjørnstad, O.N., O'Donoghue, M., Tong, H., Boonstra, R., Boutin, S., Krebs, C., Yoccoz, N.G. (1998) From patterns to processes: Phase and density dependencies in the Canadian lynx cycle *Proceedings of the National Academy of Sciences*, **95** 15430-15435.
- Tanner, M.A. (1996) *Tools for Statistical Inference*. New York: Springer-Verlag.
- Tierney, L. (1994) Markov chains for exploring posterior distributions. *The Annals of Statistics*, **22** 1701-1728.
- Turchin, P., Taylor, A.D. (1992) Complex dynamics in ecological time series. *Ecology*,

**73** 289-305.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models (Second Edition)*. New York: Springer-Verlag.

$H_{obs}$  - model with observation error  
 $H_{no\ obs}$  - model without observation error

Bayes Factor	Evidence against $H_{no\ obs}$	Number of Trials
less than 1	none	10
1 to 3	not worth more than a bare mention	7
3 to 20	positive	17
20 to 150	strong	10
greater than 150	very strong	6

Table 1: Summary of Bayes factors for 50 trials. The Bayes factor is computed by taking the ratio of the  $p(data|H_{obs})$  to  $p(data|H_{no\ obs})$ . See text for details on the calculations. The interpretations of the Bayes factors is taken from Kass and Raftery (1995).

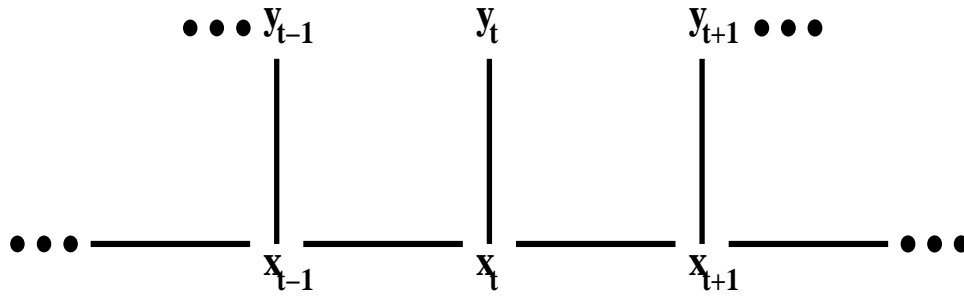


Figure 1: This figure shows the conditional independence structure of a dynamic linear model. Each of the  $x$ 's given the values of the surrounding nodes is conditionally independent of the rest of the graph.

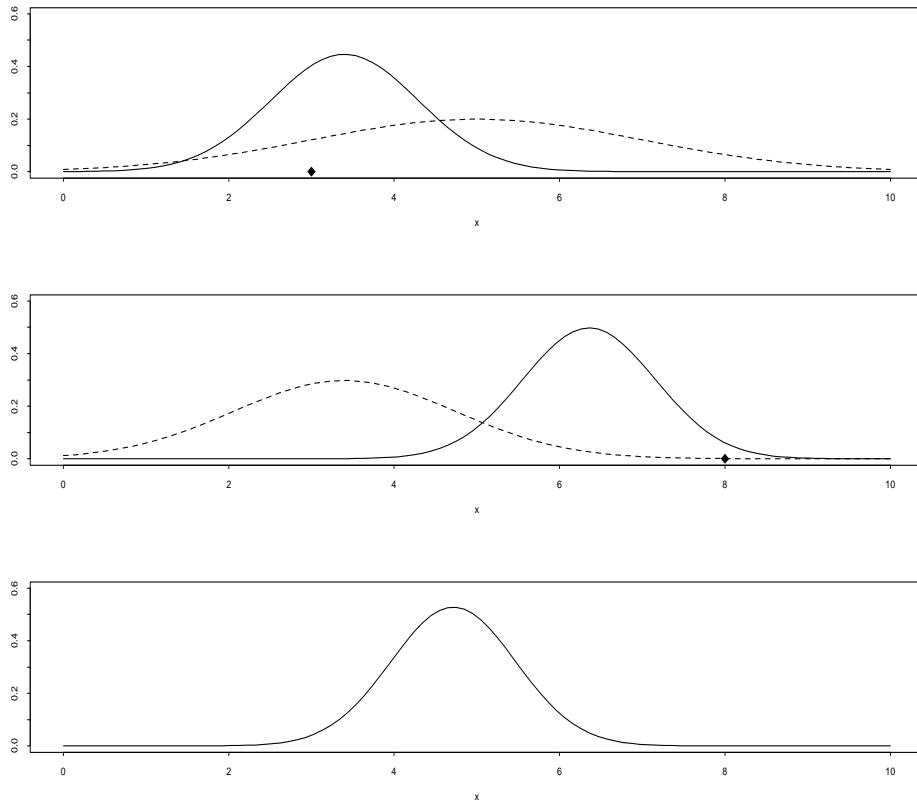


Figure 2: This figure demonstrates how the FFBS algorithm updates posterior distributions as it processes data sequentially. The details of the model and data are given in the text. The two data points,  $y_1$  and  $y_2$ , are represented by the black diamonds in graphs 1 and 2 respectively. The lines in the graphs represent the following distributions: Graph 1, dashed line -  $p(x_1)$ ; Graph 1, solid line -  $p(x_1|y_1)$ ; Graph 2, dashed line -  $p(x_2|y_1)$ ; Graph 2, solid line -  $p(x_2|y_1, y_2)$ ; Graph 3, solid line -  $p(x_1|y_1, y_2)$

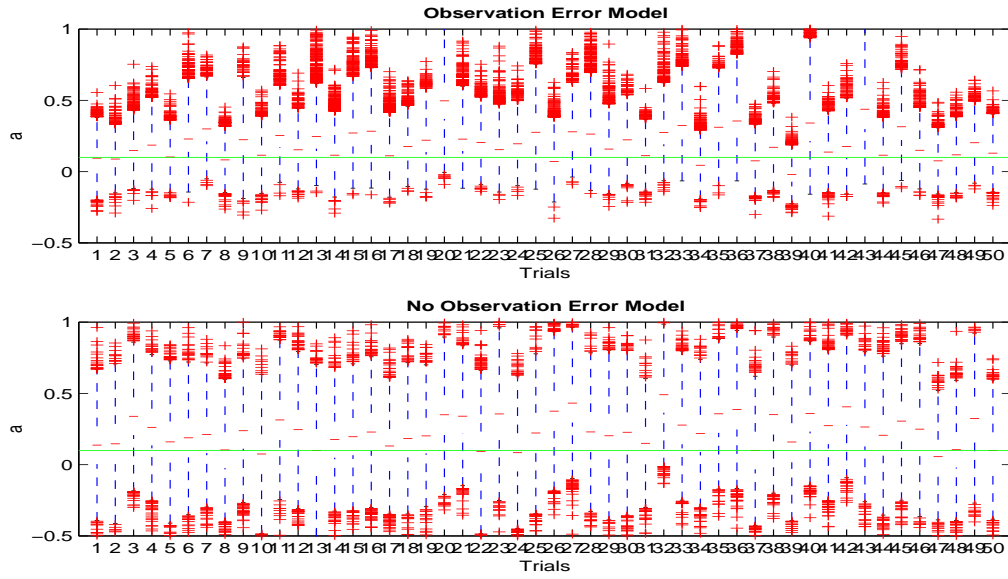


Figure 3: Boxplots for the posterior distribution of  $a$ . The true value of  $a$  is represented by the horizontal line at  $a = 0.1$ .

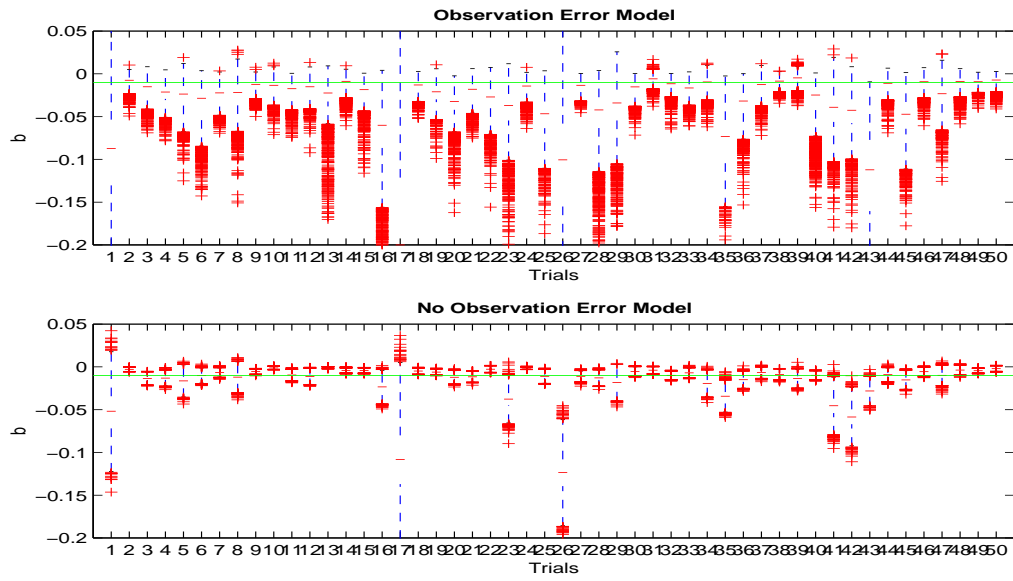


Figure 4: Boxplots for the posterior distribution of  $b$ . The true value of  $b$  is represented by the horizontal line at  $b = -0.01$ .