APPENDIX: GIBBS SAMPLER FOR BAYESIAN STATE-SPACE MODELS

Exponential model

The Gibbs sampler is described for ecological examples by Clark et al. (2003, 2004) and Wikle (2003). It involves alternately sampling from each of the conditional posterior densities that are obtained by factoring the joint posterior. Here we outline the basic elements of the Gibbs sampler for the exponential model. Models are implemented in the language *R*. Algebra is reviewed in Clark (2004).

States x_t—The conditional posterior for each of the unobserved log-abundances, x_t, is the product of three normal distributions (eqn 5), which is solved by "completing the square",

$$p\left(x_t \middle| x_{t-1}, x_{t+1}, y_t, b, \sigma^2, \tau^2\right) \propto N\left(x_t \middle| x_{t-1} + b, \sigma^2\right) N\left(x_{t+1} \middle| x_t + b, \sigma^2\right) N\left(y_t \middle| x_t, \tau^2\right)$$

$$= N\left(x_t \middle| Vv_t, V\right), \qquad \text{A.1}$$

The resulting normal distribution has mean Vv_t and variance V, given by

$$Vv_t = V\left[\frac{(x_{t-1} + x_{t+1})}{\sigma^2} + \frac{y_t}{\tau^2}\right]$$

$$V^{-1} = \frac{2}{\sigma^2} + \frac{1}{\tau^2} \qquad \text{A.2}$$

The variance is the harmonic mean of variances from the three arrows (A, B, and C) in Figure 1. The mean is the weighted average of contributions from the three arrows, the weights being the inverses of the three variances. Because the first census is conditioned on the 'prior' x_0, it is treated like the other censuses. Had we conditioned instead on the first census, it would conditionally depend only on arrows B and C, in which case we have,

$$Vv_1 = V\left[\frac{x_2}{\sigma^2} + \frac{y_1}{\tau^2}\right]$$

$$V^{-1} = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$$

For the last census, we have (from arrows A and C) the mean value

$$Vv_T = V\left[\frac{x_{T-1}}{\sigma^2} + \frac{y_T}{\tau^2}\right]$$

Process error σ^2—The conditional posterior for the process error is the product of the normal likelihood and the inverse gamma prior to give the conditional posterior

$$\sigma^2 \middle| \mathbf{x}, b \sim IG\left(\alpha_\sigma + \frac{T}{2}, \beta_\sigma + \frac{1}{2}\sum_{t=1}^{T}(x_t - x_{t-1} - b)^2\right) \qquad \text{A.3}$$

Observation error σ^2—If the same distribution applies to all censuses, then the conditional posterior is

$$\tau^2 \middle| \mathbf{x}, \mathbf{y} \sim IG\left(\alpha_\tau + \frac{T}{2}, \beta_\tau + \frac{1}{2}\sum_{t=1}^{T}(y_t - x_t)^2\right) \qquad \text{A.4}$$

Rate parameter b—For a normal prior, the conditional posterior for the parameter *b* is also normal. It is proportional to the product

$$p\left(b \middle| \mathbf{x}, \sigma^2\right) \propto \prod_{t=1}^{T} N\left(x_t \middle| x_{t-1} + b, \sigma^2\right) N\left(b \middle| b_0, V_b\right)$$

$$= N\left(b \middle| Vv, V\right) \qquad \text{A.5}$$

with mean and variance

$$Vv = V\left(\frac{\sum_{t=1}^{T}(x_t - x_{t-1})}{\sigma^2} + \frac{b_0}{V_b}\right)$$

$$V^{-1} = \frac{T}{\sigma^2} + \frac{1}{V_b} \qquad \text{A.6}$$

For a noninformative prior (large V_b), the posterior conditional mean tends to the mean change in log density per time increment.

Variable observation errors--Variable observation errors mean that each must be sampled from a different distribution. For the conditional posterior on x_t, we replace τ^2 in eqn A.4 with the value for that specific observation, τ_t^2. The conditional posterior for each variance is then

$$\tau_t^2 | x_t, y_t \sim IG\left(\alpha_t + \frac{1}{2}, \beta_t + \frac{1}{2}(y_t - x_t)^2\right). \qquad \text{A.7}$$

These conditionals replace eqn A.4.

Uneven sample intervals—From eqn 9, we have the conditional density for the state at time *t*

$$p\left(x_t \middle| x_{t-\delta_t}, x_{t+\delta_{t+1}}, y_t, b, \sigma^2, \tau^2\right) \propto N\left(x_t \middle| x_{t-\delta_t} + \delta_t b, \delta_t^2 \sigma^2\right) N\left(x_{t+\delta_{t+1}} \middle| x_t + \delta_{t+1} b, \delta_{t+1}^2 \sigma^2\right) N\left(y_t \middle| x_t, \tau_t^2\right)$$

$$= N\left(x_t \middle| V_t v_t, V_t\right)$$

This replaces eqn A.1. Completing the square, we obtain mean and variance

$$V_t v_t = V_t \left(\frac{x_{t-\delta_t} + \delta_t b}{\delta_t^2 \sigma^2} + \frac{x_{t+\delta_{t+1}} - \delta_{t+1} b}{\delta_{t+1}^2 \sigma^2} + \frac{y_t}{\tau_t^2}\right)$$

$$V_t^{-1} = \frac{1}{\tau_t^2} + \frac{1}{\delta_t^2 \sigma^2} + \frac{1}{\delta_{t+1}^2 \sigma^2}$$

The process error is sampled from

$$\sigma^2 | \mathbf{x}, b \sim IG\left(\alpha_\sigma + \frac{T}{2}, \beta_\sigma + \frac{1}{2}\sum_{t=1}^{T}\left(\frac{x_t - f(x_{t-1})}{\delta_t}\right)^2\right) \qquad \text{A.8}$$

For the growth rate parameter, the conditional posterior is

$$b\left|\mathbf{x},\mathbf{y},\sigma^2 \propto \prod_{t=1}^{T} N\left(x_t \left| x_{t-\delta_t} + \delta_t b, \delta_t^2 \sigma^2 \right.\right) N\left(b \left| b_0, V_0\right.\right)\right.$$

$$= N\left(b \left| Vv, V\right.\right)$$

having mean Vv and variance V

$$Vv = V\left(\frac{1}{\sigma^2}\sum_{t=2}^{T}\left(x_t - x_{t-\delta_t}\right)/\delta_t + \frac{b_0}{V_0}\right)$$

$$V^{-1} = \frac{T}{\sigma^2} + \frac{1}{V_0}$$

Missing values-- For missing values, the conditional posterior involves only arrows A and B in Figure 3,

$$Vv_t = V\left[\frac{x_{t-1} + x_{t+1}}{\sigma^2}\right] = \frac{x_{t-1} + x_{t+1}}{2}$$

$$V = \sigma^2/2$$

So we have two ways of drawing estimates of x_t, depending on whether or not there is an observation available at census t.

Nonlinear models

For the nonlinear model with unequal sample intervals (eqn 10), the conditional posterior for x_t is proportional to

$$p\left(x_t \left| x_{j \neq t}, \mathbf{b}, \sigma^2, \tau^2\right.\right) \propto \overset{A}{N\left(x_t \left| f\left(x_{t-\delta_t}; \mathbf{b}\right), \delta_t^2 \sigma^2\right.\right)} \overset{B}{N\left(x_{t+\delta_t} \left| f\left(x_t; \mathbf{b}\right), \delta_{t+1}^2 \sigma^2\right.\right)} \overset{C}{N\left(y_t \left| x_t, \tau^2\right.\right)}$$

The product of A and C can be integrated, being quadratic in x_t, with mean and variance

$$V_t v_t = V_t\left(\frac{f\left(x_{t-\delta_t}\right)}{\delta_t^2 \sigma^2} + \frac{y_t}{\tau^2}\right)$$

$$V_t^{-1} = \frac{1}{\delta_t^2 \sigma^2} + \frac{1}{\tau^2}$$

The target distribution can now be written as the product of two normal distributions, one of which (B) is nonlinear in x_t,

$$p\left(x_t \middle| x_{j \neq t}, \mathbf{b}, \sigma^2, \tau^2\right) = \overbrace{N\left(x_t \middle| Vv_t, V\right)}^{A \times C} \overbrace{N\left(x_{t+\delta_t} \middle| f\left(x_t; \mathbf{b}\right), \delta_{t+1}^2 \sigma^2\right)}^{B}$$

To sample from this target distribution we must embed within the Gibbs sampler either a rejection (Carlin et al. 1992) or a Metropolis (Calder et al. 2003) step. Methods for both are described in Clark (2004). The conditionals for the variances are

$$\sigma^2 \middle| \mathbf{x}, \mathbf{y}, \mathbf{b}, x_0 \sim IG\left(\alpha_\sigma + \frac{T}{2}, \beta_\sigma + \frac{1}{2}\sum_{t=1}^{T}\left(\frac{x_t - f\left(x_{t-\delta_t}\right)}{\delta_t}\right)^2\right)$$

$$\tau^2 \middle| \mathbf{x}, \mathbf{y} \sim IG\left(\alpha_\tau + \frac{T}{2}, \beta_\tau + \frac{1}{2}\sum_{t=1}^{T}\left(y_t - x_t\right)^2\right)$$

There are two regression parameters. Using conjugate normal priors, we can sample directly. For the intercept, we have the normal mean and variance

$$Vv = V\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}\frac{x_t - x_{t-\delta_t} - \delta_t b_1 e^{x_{t-\delta_t}}}{\delta_t} + \frac{B_0}{V_0}\right)$$

$$V^{-1} = \frac{T}{\sigma^2} + \frac{1}{V_0}$$

For the slope, we have mean and variance

$$Vv = V\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}\frac{e^{x_{t-\delta_t}}\left(x_t - x_{t-\delta_t} - \delta_t b_0\right)}{\delta_t} + \frac{B_1}{V_1}\right)$$

$$V^{-1} = \frac{1}{\sigma^2}\sum\nolimits_{t=1}^{T} e^{2x_{t-1}} + \frac{1}{V_1}$$

Lagged effects

From the model in eqn 11, assuming log-normal process error, we have the likelihood

$N(\Delta|\mathbf{Xb},\sigma^2\mathbf{I})N(\mathbf{y}|\mathbf{x},\tau^2\mathbf{I})$,

where **I** is the rank T identity matrix. The conditional posterior for state x_t is the product of normal densities for each of (x_{t-k}, …, x_t, …, x_{t+k}). These can be sampled by straightforward extension of the approach discussed above. The parameter vector **b** is sampled directly from the multivariate normal

$N(\mathbf{Vv},\mathbf{V})$

where

$$\mathbf{Vv} = \mathbf{V}\left(\sigma^{-2}\mathbf{X}^{\mathbf{T}}\boldsymbol{\Delta} + \mathbf{V}_b^{-1}\mathbf{b}_0\right)$$

$$\mathbf{V}^{-1} = \sigma^{-2}\mathbf{X}^{\mathbf{T}}\mathbf{X} + \mathbf{V}_b^{-1}$$

Variances are sampled as before (eqn A.8, with $\delta = 1$).

Model selection

Our application of Gelfand and Ghosh's (1998) predictive loss entails determination of expected values and second moments from MCMC followed by determination of a goodness of fit G_m and penalty term P_m. At each iteration g of the Gibbs sampler we predict states from currently imputed parameter values for model m, $E\left[x_t^{(g)}\right] = f_m\left(x_{t-1}^{(g)}; b^{(g)}\right)$. The mean for each predicted state is taken over the converged sequence and used to calculate

$$G_m = \sum_{t=1}^{T}\left(y_t - E[x_t]\right)^2$$

The second moment is calculated as the predicted sum of first moment squared and the variance, $E\left[\left(x_t^{(g)}\right)^2\right]=\left[f_m\left(x_{t-1}^{(g)};b^{(g)}\right)\right]^2+\left(\sigma^2\right)^{(g)}+\left(\tau^2\right)^{(g)}$. The penalty term is determined as

$$P_m=\sum_{t=1}^{T}\operatorname{var}[x_t]$$

where the predictive variance is taken as the difference between second moment and the squared first moment. We choose the model with the lowest value of $D_m = G_m + P_m$.

TSIR model

The posterior conditionals required for Gibbs sampling are as follows:

Infected states I_t—State I_t conditionally depends on past and future states and on the observed individuals at time t,

$$p\left(I_t \middle| I_{t-1}, I_{t+1}, I_t^{(o)}, S_{t-1}, \beta_w, \rho, \ldots\right) \propto$$
$$Bin\left(I_t \middle| S_{t-1}, \varphi\left(I_{t-1}, S_{t-1}, \beta_w\right)\right) Bin\left(I_{t+1} \middle| S_t, \varphi\left(I_t, S_t, \beta_{w+1}\right)\right) Bin\left(I_t^{(o)} \middle| I_t, \rho\right)$$

This is sampled with a Metropolis step.

Transmission rates β_w— The β_w each conditionally depend only on the transitions that correspond to the 2-wk period w,

$$p\left(\beta_w \middle| I_{t\in w}, S_{(t-1)\in w}, \ldots\right) \propto \prod_{t\in w} Bin\left(I_t \middle| S_{t-1}, \varphi\left(I_{t-1}, S_{t-1}, \beta_w\right)\right) Unif\left(\beta_w \middle| 0,1000\right)$$

We use a Metropolis step, with the normal proposal density.

The Gibbs sampler begins by drawing 26 values for β_w followed by a single value for ρ, in both cases, based on the currently imputed values for all other parameters. Infected individuals are sampled chronologically, with susceptibles being updated according to(12). We condition on an initial susceptible value of S_0 = 120,000. The last infected value is not constrained by a subsequent observation and, thus, is sampled from $Bin\left(I_t \left| S_{t-1}, \varphi\left(I_{t-1}, S_{t-1}, \beta_w\right)\right.\right) Bin\left(I_t^{(o)} \left| I_t, \rho\right.\right)$.