

USING SPATIOTEMPORAL STATISTICAL MODELS TO ESTIMATE ANIMAL ABUNDANCE AND INFER ECOLOGICAL DYNAMICS FROM SURVEY COUNTS

Paul B. Conn^{1*}, Devin S. Johnson¹, Jay M. Ver Hoef¹, Mevin B. Hooten^{2,3,4},
Josh M. London¹, and Peter L. Boveng¹

¹*National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA National Marine Fisheries Service, Seattle, Washington 98115 U.S.A.*

²*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO 80523 U.S.A.*

³*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523 U.S.A.*

⁴*Department of Statistics, Colorado State University, Fort Collins, CO 80523 U.S.A.*

Appendix A: Gibbs sampling for estimating spatiotemporal abundance from survey counts

For each spatiotemporal model fit to animal survey counts, we used Gibbs sampling (see e.g., [Gelman *et al.* 2004](#)) to obtain Markov chain Monte Carlo samples from the posterior distribution of model parameters given the data. Parameters were iteratively sampled from their full conditional distributions, some of which were available in closed form, and some of which had to be sampled using Metropolis-Hastings (MH) steps. For each spatiotemporal model structure considered, we provide a fuller mathematical treatment than in the main manuscript, provide details on prior distributions, and describe the full conditional distributions necessary for Gibbs sampler construction. Throughout this treatment, we use several standard statistical conventions. For instance, we use the bracket notation $[X]$ to denote the distribution (e.g., through a joint probability density function) of X , $[X|Y]$ to denote the conditional distribution of X given Y , and bold symbols to denote vectors or matrices of parameters. We use prime notation (e.g., \mathbf{X}') to denote matrix transpose, while matrix inverse is denoted by \mathbf{X}^{-1} . Unless noted differently, we use a Gaussian kernel centered at previous parameter values to generate parameter proposals for

MH updates, with a standard deviation tuned to yield an acceptance rate of 30 – 40% (Gelman *et al.* 2004). Software to implement these models on real data sets are provided in an online supplement, and are available online at <https://github.com/NMML/STabundance>.

1 Additive space-time (AST) model

The first type of spatiotemporal model that we implemented was a separable space-time model where spatial and temporal random effects are modeled independently. Ver Hoef & Jansen (2007) used such a model to successfully model harbor seal abundance from survey counts. This formulation is attractive due to the potentially small number of random effect parameters relative to a model with space-time interactions, but is restrictive in the sense that it does not permit residual spatial error to vary over time. Assuming that count data are available on spatially referenced sampling units, we write the joint posterior distribution of latent abundance ($\boldsymbol{\mu}$) and other model parameters given data as

$$[\boldsymbol{\mu}, \boldsymbol{\theta} | \mathbf{C}, \mathbf{o}] \propto [\mathbf{C} | \mathbf{o}, \boldsymbol{\mu}] [\boldsymbol{\mu} | \boldsymbol{\theta}] [\boldsymbol{\theta}], \quad (\text{A.1})$$

where $\boldsymbol{\theta}$ denote the collection of all parameters other than latent abundance, \mathbf{C} denote temporally and spatially referenced animal counts, and \mathbf{o} denote linear model offsets used to encapsulate the amount of area sampled in different sampling units. Note also that our convention throughout this treatment is to omit covariate data \mathbf{x} from distributional expressions (for instance, $\boldsymbol{\mu}$ may also be dependent on habitat covariates to help predict abundance, as described later).

The observation model $[\mathbf{C} | \mathbf{o}, \boldsymbol{\mu}]$ describes how count data arise from different levels of sampling effort and latent abundance parameters. For purposes of the AST model, we make the assumption that count data are Poisson distributed:

$$\begin{aligned} [\mathbf{C} | \mathbf{o}, \boldsymbol{\mu}] &\equiv \text{Poisson}(\boldsymbol{\lambda}), \text{ where} \\ \boldsymbol{\lambda} &= \exp(\mathbf{o} + \mathbf{H}\boldsymbol{\mu}). \end{aligned} \quad (\text{A.2})$$

Here, the matrix \mathbf{H} is constructed in such a manner to select elements of the full set of space- and

time-specific $\boldsymbol{\mu}$ values for which sampling effort is actually applied.

We adopt a log-linear formulation for the process model $[\boldsymbol{\mu}|\boldsymbol{\theta}, \mathbf{x}]$, and in this instance include additive spatial and temporal random effects:

$$\boldsymbol{\mu}_t = \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\eta} + \gamma_t + \boldsymbol{\epsilon}_t.$$

In particular, time-specific vectors of the log of abundance intensity ($\boldsymbol{\mu}_t$) are written in a regression-like framework that includes a time-specific design matrix \mathbf{X}_t , regression parameters $\boldsymbol{\beta}$, spatial random effects $\boldsymbol{\eta}$, temporal random effects, γ_t , and residual Gaussian-distributed error.

To impose spatial structure on $\boldsymbol{\eta}$, we employed a process-convolution formulation (see e.g., [Higdon 1998](#), [Calder *et al.* 2002](#)), whereby

$$\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\alpha}. \tag{A.3}$$

Here, the $1 \times m$ dimensional vector $\boldsymbol{\alpha}$ holds spatial random effects on a reduced parameter space. The reduced parameter space is achieved by distributing a total of m knots evenly across the landscape, and associating a Gaussian kernel with each such knot. Under this formulation, the random effects $\boldsymbol{\alpha}$ are equivalent to weights on each of the Gaussian kernels. The $S \times m$ dimensional matrix \mathbf{K} maps the random effects (or weights) to S -dimensional space. The entries in \mathbf{K} are proportional to distance-specific kernel densities used to interpolate between the m knots and the S spatiotemporal locations being modeled. With rich datasets, it is possible to computationally intensive algorithms to optimize values for the elements of \mathbf{K} (and to induce anisotropy), but we will typically not be so lucky in ecological applications. Thus, in subsequent examples we adopt the commonly used practice of setting the standard deviations for each kernel (σ) equal to the distance between knot locations, and setting the i th row and j th column of \mathbf{K} to $N(d_{ij}; 0, \sigma^2)$, where N denotes a standard Gaussian probability density function and d_{ij} gives the distance from the centroid of sampling unit i to the location of knot j . The elements of \mathbf{K} are then renormalized so that rows sum to 1.0. As is typical in such applications, we model the random effects $\boldsymbol{\alpha}$, as iid Gaussian with mean zero and precision $1/\tau_\eta$.

We account for temporal autocorrelation by assigning an RW2(τ_γ) prior distribution (see

e.g., [Rue & Held 2005](#), section 3.4) to temporal random effects. Implemented using Gaussian Markov random field (GMRF) machinery (again see [Rue & Held 2005](#), for a thorough treatment), this prior imposes temporal dependence among the γ and a greater degree of smoothness than a simpler RW1 structure (a potentially useful feature for analysis of typically sparse ecological datasets).

We must of course assign prior distributions to the lowest level parameters in θ in order to conduct Bayesian inference. Our strategy is to use the following choices:

$$\begin{array}{ll} \beta & \text{Regression parameters} \quad \text{MVN}(\mathbf{0}, (\tau_\beta X'X)^{-1}) \\ \tau_\eta, \tau_\epsilon, \tau_\gamma & \text{Gaussian precisions} \quad \text{Gamma}(1.0, 0.01). \end{array}$$

Here, we use MVN to denote the multivariate normal distribution, and set $\tau_\beta = 0.01$ in all applications. We chose 1.0 and 0.01 for the shape and rate parameters on the conjugate Gamma prior for precision parameters because these values make the prior reasonably flat near the origin, which seems to impart a greater degree of stability in estimation than many other common choices.

1.1 Bayesian analysis of the AST model

We used Gibbs sampling to cycle through updates of different groups of parameters and latent variables. This process involves sampling from the so-called full conditional distributions of these parameters ([Gelman *et al.* 2004](#)), which are specified as follows:

1. Updating μ for surveyed sample units

For each sample unit s surveyed at a given time t , $\mu_{s,t}$ is updated with a MH step, with full conditional

$$[\mu_{s,t}|\cdot] \propto \text{Normal}(\mu_{s,t}; \mathbf{X}_{s,t}\beta + \eta_s + \gamma_t, \tau_\epsilon^{-0.5}) \times \text{Poisson}(C_{s,t}, \exp(o_{s,t} + \mu_{s,t})).$$

2. Updating β

The full conditional distribution for β is available in closed form; we use a Gibbs step to sample

from

$$[\boldsymbol{\beta}|\cdot] \equiv \text{Normal}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\mu} - \boldsymbol{\eta} - \boldsymbol{\gamma}), (\mathbf{X}'\mathbf{X})^{-1}(\tau_\epsilon + \tau_\beta)^{-1}).$$

Note that only surveyed sample units are included in this update (e.g., the design matrix \mathbf{X} and random effect vectors are constructed using sample units at times and locations where surveys are conducted).

3. Updating τ_ϵ

The full conditional distribution for the precision of exchangeable errors is also available in closed form as

$$[\tau_\epsilon|\cdot] \equiv \text{Gamma}(0.5n + 1.0, 0.5\boldsymbol{\Delta}'\boldsymbol{\Delta} + 0.01),$$

and can be simulated from directly. Here, $\boldsymbol{\Delta} = \boldsymbol{\mu} - (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\gamma})$, and once again we only use surveyed sample units to construct the elements of each matrix and vector.

4. Updating $\boldsymbol{\alpha}$

The full conditional distribution for $\boldsymbol{\alpha}$ is available in closed form as

$$[\boldsymbol{\alpha}|\cdot] \equiv \text{MVN}(\mathbf{m}_\alpha, \mathbf{v}_\alpha),$$

(where MVN denotes the multivariate normal probability density function) and can also be simulated from directly. The variance-covariance matrix \mathbf{v}_α of the full conditional can be written as

$$\mathbf{v}_\alpha = (\mathbf{K}'\mathbf{K}\tau_\epsilon + \mathcal{D}(\tau_\eta))^{-1},$$

where $\mathcal{D}(\tau_\eta)$ here gives an $m \times m$ diagonal matrix with diagonal entries all equal to τ_η . The mean of the full conditional is given by

$$\mathbf{m}_\alpha = \mathbf{v}_\alpha\tau_\epsilon\mathbf{K}'(\boldsymbol{\mu} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma})$$

As in previous steps, we only use surveyed sample units to construct the elements of each matrix and vector included in this full conditional.

5. Updating τ_η

The precision τ_η is also available in closed form and can be simulated directly via the full conditional

$$[\tau_\eta|\cdot] \equiv \text{Gamma}(0.5m + 1.0, 0.5\boldsymbol{\alpha}'\boldsymbol{\alpha} + 0.01).$$

6. Updating $\boldsymbol{\gamma}$

The full conditional for temporal random effects is also available in closed form as

$$[\boldsymbol{\gamma}|\cdot] \equiv \text{MVN}(\mathbf{m}_\gamma, \mathbf{v}_\gamma)$$

and can be simulated directly. The variance-covariance matrix \mathbf{v}_γ can be written as

$$\mathbf{v}_\gamma = (\tau_\epsilon \mathbf{Z}'\mathbf{Z} + \tau_\gamma \mathbf{Q})^{-1},$$

where \mathbf{Z} is an $(n \times T)$ design matrix linking each observed count with an associated time step (thus the elements of \mathbf{Z} , z_{ij} , are 1.0 if observation i was obtained at time step j and zero otherwise), and where \mathbf{Q} is the structure matrix for an RW2 intrinsic GMRF (see [Rue & Held 2005](#), section 3.4). The mean of the full conditional is given by

$$\mathbf{m}_\gamma = \mathbf{v}_\gamma \tau_\epsilon \mathbf{Z}'(\boldsymbol{\mu} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\eta}).$$

7. Updating τ_γ

The precision τ_γ is also available in closed form and can be simulated directly via the full conditional

$$[\tau_\gamma|\cdot] \equiv \text{Gamma}(0.5T + 1.0, 0.5\boldsymbol{\gamma}'\mathbf{Q}'\boldsymbol{\gamma} + 0.01).$$

Generating posterior predictions of abundance

To predict animal abundance in locations and times that were not sampled, we first sampled $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu} \sim \text{Normal}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\gamma}, \tau_e^{-1}),$$

where the design matrix \mathbf{X} and vectors $\boldsymbol{\mu}$, $\boldsymbol{\eta}$, and $\boldsymbol{\gamma}$ are in this case understood to be composed of unique times and sampling units which are not surveyed. Predictions of abundance across the grid can then be made using

$$N_{st} \sim \text{Poisson}(A_s \exp(\mu_{s,t})),$$

where A_s gives the area of sample unit s relative to the mean sample unit area. Predictions of animal at each time step can simply be calculated as $N = \sum_s N_{s,t}$.

2 Spatiotemporal process convolution (STPC) model

The STPC model shares many similarities with the AST model, but allows for spatiotemporal interactions. We start with the same general structure and observation model as for the AST model, as given by eqns. [A.8](#) & [A.2](#). However, the log of abundance intensity is in this case written as a function of a spatio-temporal effect, $\kappa_{s,t}$:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\kappa} + \boldsymbol{\epsilon}_t.$$

As before, we will write the $\boldsymbol{\kappa}$ as a function of spatiotemporal parameters on a reduced dimensional space associated with m knots placed evenly across the landscape, writing

$$\boldsymbol{\kappa} = \mathbf{L}\boldsymbol{\alpha}.$$

However, in this case, there are a total of mT α_{kt} values (one for each knot and time step), and \mathbf{L} is an $(ST \times mT)$ matrix used to map spatio-temporal grid cells to respective α_{kt} values. The entries of \mathbf{L} include values from the \mathbf{K} matrix from the previous section, rearranged to pull off

appropriate values of α , and is constructed as follows:

$$\mathbf{L} = \begin{bmatrix} K_{11} & \mathbf{0}_{T-1} & K_{12} & \mathbf{0}_{T-1} & \dots & K_{1m} & \mathbf{0}_{T-2} & 0 \\ K_{21} & \mathbf{0}_{T-1} & K_{22} & \mathbf{0}_{T-1} & \dots & K_{2m} & \mathbf{0}_{T-2} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ K_{S1} & \mathbf{0}_{T-1} & K_{S2} & \mathbf{0}_{T-1} & \dots & K_{Sm} & \mathbf{0}_{T-2} & 0 \\ 0 & K_{11} & \mathbf{0}_{T-1} & K_{12} & \mathbf{0}_{T-1} & \dots & K_{1m} & \mathbf{0}_{T-2} \\ 0 & K_{21} & \mathbf{0}_{T-1} & K_{22} & \mathbf{0}_{T-1} & \dots & K_{2m} & \mathbf{0}_{T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & K_{S1} & \mathbf{0}_{T-1} & K_{S2} & \mathbf{0}_{T-1} & \dots & K_{Sm} & \mathbf{0}_{T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{T-1} & K_{S1} & \mathbf{0}_{T-1} & K_{S2} & \mathbf{0}_{T-1} & \dots & \mathbf{0}_{T-1} & K_{Sm} \end{bmatrix}.$$

Here, $\mathbf{0}_x$ denotes an all zero row vector of length x . We allow the α_{kt} parameters to change smoothly over time by giving α_{kt} values at each knot k an RW2 Gaussian Markov random field prior distribution (Rue & Held 2005). However, the RW2 prior is degenerate by an order of two; we restore propriety by following the suggestion of Johnson *et al.* (2013) and including separate intercept and slope terms for the α_{kt} parameters at each knot, giving these intercept and slope terms proper priors. In particular, we model

$$\alpha = \mathbf{X}_0^{rw2} \beta_0^{rw2} + \mathbf{X}_1^{rw2} \beta_1^{rw2} + \varpi,$$

where \mathbf{X}_0^{rw2} is a design matrix for the intercept term associated with each knot, \mathbf{X}_1^{rw2} is a design matrix associated with the linear slope effect for each knot, β_0^{rw2} give intercepts for each knot, β_1^{rw2} give slope parameters for each knot, and $\varpi \sim \mathcal{N}(\mathbf{0}, \tau_\alpha \mathbf{Q}^*)$ denote random effects with

precision parameter τ_γ and a $(mT \times mT)$ block diagonal structure matrix:

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{Q} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q} \end{bmatrix}$$

(\mathbf{Q} being the “standard” RW2 structure matrix for a single time series). For each knot k , the ϖ are constrained to have a mean and slope of 0. We specify prior distributions for regression and Gaussian precision parameters in the same manner as for the AST model (cf., eqn. A.9), but impose informative priors on the β_0^{rw2} and β_1^{rw2} parameters. In particular, for all of our applications, we let each element of β_0^{rw2} have a $\mathcal{N}(0, \tau_{\beta_0}^{-1})$ prior distribution, and each element of β_1^{rw2} have a $\mathcal{N}(0, \tau_{\beta_1}^{-1})$ prior. We selected $\tau_{\beta_0} = 1$ and $\tau_{\beta_1} = 10$ so as to permit considerable unmodeled spatiotemporal variation in the log of abundance intensity, while providing enough structure to limit the chance of extrapolating “past the range of observed data” for times and locations where the density of sampling efforts is low.

2.1 Bayesian analysis of the STPC model

Bayesian analysis of the STPC model largely followed the approach used for AST model, with the exception that instead of separate spatial ($\boldsymbol{\eta}$) and temporal ($\boldsymbol{\gamma}$) effects, there is a single spatiotemporal effect, $\boldsymbol{\kappa}$. In fact, steps 1-3 of the STPC Gibbs sampler are exactly the same as for the AST model, as is the approach for generating posterior predictions of abundance, and simply require placing $\boldsymbol{\eta} + \boldsymbol{\gamma}$ with $\boldsymbol{\kappa}$. The remaining steps are as follows:

4. Updating β_0^{rw2} and β_1^{rw2} :

To update knot-specific intercept and slope parameters, we first let $\mathbf{X}_{rw2} = [\mathbf{X}_0^{rw2} \ \mathbf{X}_1^{rw2}]$, $\boldsymbol{\beta}_{rw2} = [\beta_0^{rw2} \ \beta_1^{rw2}]'$, and $\Sigma^{-1} = \mathcal{D}(\mathbf{1}\tau_{\beta_0} \ \mathbf{1}\tau_{\beta_1})$, where \mathcal{D} denotes a diagonal matrix, and $\mathbf{1}$ is a column vector of length m . The full conditional distribution for $\boldsymbol{\beta}_{rw2}$ is then available in closed form as

$$[\boldsymbol{\beta}_{rw2}|\cdot] \equiv \text{MVN}(\mathbf{m}_b, \mathbf{v}_b), \text{ where}$$

$$\mathbf{v}_b^{-1} = (\mathbf{L}\mathbf{X}_{rw2})'(\mathbf{L}\mathbf{X}_{rw2})\tau_\epsilon + \Sigma^{-1}, \text{ and}$$

$$\mathbf{m}_b = \mathbf{v}_b\tau_\epsilon(\mathbf{L}\mathbf{X}_{rw2})'(\boldsymbol{\mu} - \mathbf{X}\boldsymbol{\beta} - \mathbf{L}\boldsymbol{\varpi}).$$

6. Updating $\boldsymbol{\varpi}$

The full conditional distribution for $\boldsymbol{\varpi}$ is also available in closed form as

$$[\boldsymbol{\varpi}|\cdot] \equiv \text{MVN}(\mathbf{m}_\varpi, \mathbf{v}_\varpi), \tag{A.4}$$

and can also be simulated from directly. The inverse of the variance-covariance matrix v_α for the full conditional can be written as

$$\mathbf{v}_\alpha^{-1} = (\mathbf{L}'\mathbf{L}\tau_\epsilon + \tau_\alpha\mathbf{Q}^*)^{-1},$$

where \mathbf{Q}^* gives the $mT \times mT$ block diagonal matrix defined in the previous section. The mean of the full conditional is given by

$$\mathbf{m}_\alpha = \mathbf{v}_\alpha\tau_\epsilon\mathbf{L}'(\boldsymbol{\mu} - \mathbf{X}\boldsymbol{\beta} - \mathbf{L}\mathbf{X}_{rw2}\boldsymbol{\beta}_{rw2}).$$

For each knot, we impose a zero mean and zero slope constraint on the α_k . using “conditioning by Kriging” (cf., [Rue & Held 2005](#), , Eq. 2.30). In particular, we first simulate $\boldsymbol{\varpi}$ using the unconstrained distribution given in Eq. [A.4](#). We then calculate a constrained sample $\check{\boldsymbol{\varpi}}$ as

$$\check{\boldsymbol{\varpi}} = \boldsymbol{\varpi} - \mathbf{v}_\alpha^{-1}\mathbf{A}'(\mathbf{A}\mathbf{v}_\alpha^{-1}\mathbf{A}')^{-1}\mathbf{A}\boldsymbol{\varpi},$$

where $\mathbf{A} = (\mathbf{X}'_{rw2}\mathbf{X}_{rw2})^{-1}\mathbf{X}'_{rw2}$.

6. Updating τ_α

The precision τ_η is also available in closed form and can be simulated directly via the full conditional

$$[\tau_\eta|\cdot] \equiv \text{Gamma}(0.5mT + 1.0, 0.5\boldsymbol{\alpha}'\boldsymbol{\alpha} + 0.01).$$

3 Open population resource selection (OPRS) model

As with the AST and STPC models, the OPRS formulation makes use of a Poisson observation model, such that

$$C_{s,t} \sim \text{Poisson}(\lambda_{s,t}).$$

However, we alter the formulation for $\lambda_{s,t}$ to allow for dynamical resource selection. In particular, we set

$$\lambda_{s,t} = \exp(o_{s,t} + Z_{s,t}), \text{ where}$$

$$\mathbf{Z}_t \sim \text{MVN}(\boldsymbol{\mu}_t, \Sigma_Z) \text{ and} \tag{A.5}$$

$$\boldsymbol{\mu}_t = \begin{cases} \mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\epsilon}_1 & \text{if } t = 1 \\ \mathbf{M}_t\boldsymbol{\mu}_{t-1} + \boldsymbol{\gamma}_t + \boldsymbol{\epsilon}_t & \text{if } t > 1. \end{cases} \tag{A.6}$$

Here, $\Sigma_z = \tau_z^{-1}\mathcal{I}$ is a diagonal covariance matrix, \mathbf{X}_1 is a design matrix specific to the time step at which initial surveys are conducted, and other parameters are defined in the same manner as for the AST model. After the first time step, changes in abundance between time steps are governed by a resource selection-based transition matrix \mathbf{M}_t , and also subject to $\boldsymbol{\gamma}_t$, a temporally autocorrelated random effect (as in previous models $\boldsymbol{\epsilon}_t$ is Gaussian white noise). The astute reader may notice that we have included Gaussian error in two places (i.e., in both eqns A.5 and A.6). We also model $\boldsymbol{\gamma}_t$ using an AR1 process instead of an RW2 process. These modifications were made to increase the efficiency of MCMC computation (see below).

To construct an ecologically plausible transition matrix \mathbf{M}_t , we appeal to discrete-space resource selection theory employing weighted distributions (cf., Patil & Rao 1978, Lele & Keim 2006, Johnson *et al.* 2008). In such applications, the probability of transitioning from a particular

location a at time $t - 1$ to location b at time t is given as

$$\psi_t^{ab} = \frac{w_{b,t}\varphi_{a,b}}{\sum_s w_{s,t}\varphi_{a,s}}.$$

This formulation involves two components: a measure of habitat preference for location s at time t ($w_{s,t}$), and a redistribution kernel consisting of the elements $\varphi_{a,s}$. The redistribution kernel describes animal movement in absence of a gradient in habitat quality. We use a truncated, symmetric normal distribution kernel to summarize $\varphi_{a,s}$ in all applications, thus approximating simple Brownian (diffusive) movement:

$$\varphi_{a,b} \propto \text{Normal}(d(a,b), \tau_d^{-1}),$$

Here, $d(a,b)$ simply gives the Euclidean distance between the centroid of sample units a and b , and τ_d is a precision parameter to be estimated. We specify further structure on $w_{s,t}$ using a log-linear formulation:

$$\log(\mathbf{w}_t) = \mathbf{X}_t\boldsymbol{\beta},$$

where the design matrix \mathbf{X}_t is allowed to vary over time to reflect changing environmental and habitat covariates. An intercept is omitted from \mathbf{X}_t as it is not identifiable.

The matrix \mathbf{M}_t is constructed using values of ψ_t^{ab} in a computationally efficient framework. In particular,

$$\mathbf{M}_t = (\mathbf{W}_t^{-1}\boldsymbol{\varphi}_t\mathbf{I}_{w_t})', \tag{A.7}$$

where each row a of the $S \times S$ matrix $\boldsymbol{\varphi}_t$ holds entries corresponding to ψ_t^{ab} ($b \in \{1, 2, \dots, S\}$), and \mathbf{I}_{w_t} is a diagonal matrix with elements w_t along the main diagonal. The \mathbf{W}_t matrix provides normalization (i.e., so all rows of \mathbf{M}_t sum to one), and is calculated as $\mathbf{W}_t = \boldsymbol{\varphi}_t\mathbf{I}_{w_t}\mathbf{1}$, where $\mathbf{1}$ is an $(S \times 1)$ vector with each element equal to 1.0.

The previous formulation of the OPRS model induces a joint posterior distribution

specified by

$$[\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\theta} | \mathbf{C}, \mathbf{o}] \propto [\mathbf{C} | \mathbf{o}, \mathbf{Z}] [\mathbf{Z} | \boldsymbol{\mu}, \boldsymbol{\theta}] [\boldsymbol{\mu} | \boldsymbol{\theta}] [\boldsymbol{\theta}], \quad (\text{A.8})$$

where we use $\boldsymbol{\theta}$ to denote the set of parameters consisting of $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_d, \tau_\eta, \tau_\gamma, \tau_z\}$.

To analyze the OPRS model, we need to assign prior distributions to several model parameters. We selected the following priors for all analyses:

β_i	Regression parameters	Normal(0, 100)
$\tau_\epsilon, \tau_\eta, \tau_\gamma, \tau_d$	Gaussian precisions	Gamma(1.0, 0.01)
σ	AR1 parameter	Uniform(0.0, 1.0)

3.1 Bayesian analysis of the OPRS model

1. Updating \mathbf{Z}

The full conditional distribution for \mathbf{z} differs depending on whether surveys occur in site s at time t . Letting \mathcal{S}_t denote the set of sampling units surveyed at time t , and \mathcal{U}_t denote the set of sampling units not surveyed at time t :

$$[\mathbf{Z} | \cdot] \equiv \begin{cases} \text{Poisson}(C_{s,t}; \exp(o_{s,t} + Z_{s,t})) \times \text{Normal}(Z_{s,t}; 0, \tau_z^{-1}) & \text{if } s \in \mathcal{S}_t \\ \text{Normal}(Z_{s,t}; 0, \tau_z^{-1}) & \text{if } s \in \mathcal{U}_t. \end{cases}$$

Thus, for unsurveyed cells, $Z_{s,t}$ can be simulated directly from a normal distribution, while another approach is needed if $s \in \mathcal{S}$. Our approach in this study is to use separate Langevin-Hastings steps (see e.g., [Givens & Hoeting 2005](#), section 7.1.4) for each time step t to simulate values of \mathbf{Z}_t for which $s \in \mathcal{S}_t$. The Langevin algorithm exploits the first derivative of the log of the full conditional for efficient proposal generation. The algorithm was tuned to achieve target acceptance rates of 20-30%, a desirable range for higher dimensional block updates ([Gelman *et al.* 2004](#)).

2. Updating $\boldsymbol{\mu}$

In order to update $\boldsymbol{\mu}$, we make use of a forward filtering, backward sampling algorithm

formulated for state space time series models (Durbin & Koopman 2002). This algorithm makes use of Kalman filtering/smoothing methodology, but works with “disturbances” (the Gaussian errors) rather than the actual state and observation vectors to speed up computation. This approach requires Gaussian errors on “state” and “observation” vectors. To make use of this algorithm, we first make several adjustments. First, we write an expanded state vector as

$$\Theta_t = \begin{bmatrix} \boldsymbol{\mu}_t \\ \gamma_t \end{bmatrix},$$

and define an expanded transition matrix as

$$\mathcal{M}_t = \begin{bmatrix} \mathbf{M}_t & \mathbf{1} \\ \mathbf{0} & \sigma \end{bmatrix},$$

where $\mathbf{1}$ is a length S column vector of ones, and $\mathbf{0}$ is a length S row vector of zeros. This change of notation allows us to write system dynamics in terms of familiar Gaussian state and observation processes amenable to processing by the Kalman filter:

$$\Theta_t = \mathcal{M}_t \Theta_{t-1} + \mathbf{E}_t, \tag{A.9}$$

$$\mathbf{Z}_t = \mathbf{H}_t \Theta_t + \varepsilon_t. \tag{A.10}$$

Here, the $S \times (S + 1)$ dimensional matrix \mathbf{H}_t maps $\Theta \rightarrow \mathbf{Z}$, and $\mathbf{E}_t = [\boldsymbol{\epsilon}_t \delta_t]$, where $\delta_t \sim \text{Normal}(0, \tau_\gamma^{-1})$. Of course, we do not actually observe \mathbf{Z}_t either, but can still use the Kalman filtering machinery to provide efficient MCMC updates. To do so, we write the observation error covariance matrix (i.e., for ε_t) as $\mathbf{R}_t = \tau_z^{-1} \mathcal{D}(S)$, where $\mathcal{D}(S)$ is an $(S \times S)$ identity matrix, and write the process error covariance matrix (i.e., for $\boldsymbol{\epsilon}_t$) as

$$\mathbf{Q}_t = \begin{bmatrix} \tau_\epsilon^{-1} \mathcal{D}(S) & \mathbf{0} \\ \mathbf{0} & \tau_\gamma^{-1} \end{bmatrix}.$$

We then use the following steps to update $\boldsymbol{\mu}_t$:

1. Draw random vectors of the Gaussian disturbances, ϵ^+ and ε^+ (the ‘+’ denoting that these are a simulated draw).
2. Initialize the state vector Θ_1^+ using a $\text{Normal}(\mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\eta}, \tau_\epsilon^{-1})$ distribution for the first S elements, and a $\text{Normal}(0, \tau_\gamma^{-1})$ distribution for the $(S + 1)\text{th}$ element, use the vectors of ϵ^+ and ε^+ together with the state and observation equations (i.e., eqns. [A.9](#) & [A.10](#)) to define simulated state and observation vectors Θ^+ and \mathbf{Z}^+ .
3. Compute $\mathbf{Z}_t^* = \mathbf{Z}_t - \mathbf{Z}_t^+$, which we run through the Kalman filter using the following forward equations:

$$\begin{aligned}
\mathbf{P}_t &= \mathcal{M}_t \mathbf{P}_{t-1} \mathbf{L}'_{t-1} + \mathbf{Q} \\
\mathbf{L}_t &= \mathcal{M}_{t+1} - \mathbf{G}_t \mathbf{H}_t \\
\mathbf{G}_t &= \mathcal{M}_{t+1} \mathbf{P}_t \mathbf{H}'_t \mathbf{F}^{-1} \\
\mathbf{F}_t &= \mathbf{H}_t \mathbf{P}_t \mathbf{H}'_t + \mathbf{R}_t \\
\boldsymbol{\nu}_t &= \mathbf{Z}_t^* - \mathbf{H}_t \mathbf{A}_t \\
\mathbf{A}_t &= \mathcal{M}_t \mathbf{A}_{t-1} + \mathbf{G}_{t-1} \boldsymbol{\nu}_{t-1}.
\end{aligned}$$

Here, $\mathbf{A}_t = \mathbf{0}$ indicates that \mathbf{A} is initialized at $t = 1$ to be an all zero vector of length $S + 1$.

4. Compute vectors of smoothed disturbances at time t , \mathbf{D}_t by setting $\mathbf{D}_T = \mathbf{0}$, and moving backward by backward recursion:

$$\mathbf{D}_{t-1} = \mathbf{H}'_t \mathbf{F}_t^{-1} \boldsymbol{\nu}_t + \mathbf{L}'_t \mathbf{D}_t.$$

5. Next, we generate smoothed mean residuals for the state vector, $\hat{\Theta}_t^*$ using eqn 8. of [Durbin & Koopman \(2002\)](#) as

$$\begin{aligned}
\hat{\Theta}_1^* &= \mathbf{A}_1 + \mathbf{P}_1 (\mathbf{H}'_1 \mathbf{F}_1^{-1} \boldsymbol{\nu}_1 + \mathbf{L}'_1 \mathbf{D}_1), \text{ and} \\
\hat{\Theta}_t^* &= \mathcal{M}_t \hat{\Theta}_{t-1}^* + \mathbf{Q} \mathbf{D}_{t-1} \text{ for } t > 1.
\end{aligned}$$

6. A sample from the absolute state distribution at each time step is then calculated as

$$\tilde{\Theta}_t = \hat{\Theta}_t^* + \Theta_t^+.$$

3. Updating β_i

The full conditional distribution for each regression parameter, β_i , is given by

$$[\beta_i|\cdot] \equiv \text{Normal}(\beta_i; 0, 100) \times \text{MVN}(\boldsymbol{\mu}_1; \mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\eta}, \mathbf{Q}_1) \times \prod_{t=2}^T \text{MVN}(\boldsymbol{\mu}_t; \mathbf{M}_t\boldsymbol{\mu}_{t-1}, \mathbf{Q}_t),$$

where dependence upon $\boldsymbol{\beta}$ occurs both in the initial state distribution and in dictating the elements of \mathcal{M} through the resource selection formulation. As there is no closed form solution, we obtain samples using MH updates.

4. Updating τ_ϵ

The full conditional distribution for the precision of exchangeable errors is available in closed form as

$$[\tau_\epsilon|\cdot] \equiv \text{Gamma}(0.5n + 1.0, 0.5\boldsymbol{\Delta}'\boldsymbol{\Delta} + 0.01),$$

and can be simulated from directly. Here, $\boldsymbol{\Delta}$ is a vector of length ST , where the first S elements are residuals associated with the initial state distribution (i.e., $\boldsymbol{\mu}_1 - \mathbf{X}_1\boldsymbol{\beta} - \boldsymbol{\eta}$), and subsequent elements are residuals associated with the resource selection process (i.e., $\boldsymbol{\mu}_t - \mathbf{M}_t\boldsymbol{\mu}_{t-1} \forall t \geq 2$).

5. Updating $\boldsymbol{\alpha}$, τ_η

Reduced dimension spatial random effects for the initial state distribution ($\boldsymbol{\alpha}$) and their accompanying precision (τ_η) are updated using the same procedure as for the AST model. In this case, we use all survey units at time $t = 1$ to form relevant model matrices.

6. Updating σ

The full conditional distribution for σ (the autocorrelation parameter for AR1 temporal random effects) is available in closed form and can be used to simulate values of σ directly:

$$[\sigma|\cdot] \equiv \text{MVN}((\boldsymbol{\gamma}'_-\boldsymbol{\gamma}_-)^{-1}\boldsymbol{\gamma}'_-\boldsymbol{\gamma}_+, \tau_\gamma^{-1}(\boldsymbol{\gamma}'_-\boldsymbol{\gamma}_-)^{-1}),$$

where

$$\boldsymbol{\gamma}_- = \{\gamma_2, \gamma_3, \dots, \gamma_{t-1}\}', \text{ and}$$

$$\boldsymbol{\gamma}_+ = \{\gamma_3, \gamma_4, \dots, \gamma_T\}'.$$

Note that the initial state distribution has $\gamma_1 = 0$, so there are only $T - 1$ effective γ parameters with which to base σ updates. We enforce the constraint that $0 \leq \sigma \leq 1$ by rejecting any updates of σ that occur outside of this range.

7. Updating τ_γ The full conditional distribution for the precision of temporal random effects is available in closed form as

$$[\tau_\gamma | \cdot] \equiv \text{Gamma}(0.5(T - 1) + 1.0, 0.5\boldsymbol{\Delta}'_\gamma \boldsymbol{\Delta}_\gamma + 0.01),$$

and can be simulated from directly. Here, $\boldsymbol{\Delta}_\gamma$ is a vector of length $T - 1$, where the first element is γ_2 , and subsequent elements are $\gamma_t - \sigma\gamma_{t-1}$.

8. Updating τ_d

The parameter τ_d controls the redistribution kernel, and thus the elements of \mathcal{M}_t . We update τ_d via a MH step, where the full conditional distribution is given by

$$[\tau_d | \cdot] \equiv \text{Gamma}(\tau_d; 1.0, 0.01) \times \prod_{t=2}^T \text{MVN}(\boldsymbol{\mu}_t, \mathcal{M}_t \boldsymbol{\Theta}_t, \mathbf{Q}_t).$$

4 Closed population ideal free (CPIF) model

For the closed population ideal free model, we base inference on the joint posterior

$$[N, N_{s,t}, \boldsymbol{\omega}, \boldsymbol{\theta} | \mathbf{C}, \mathbf{o}, \mathbf{p}] \propto [\mathbf{C} | N_{s,t}, \mathbf{o}, \boldsymbol{\omega}] [\boldsymbol{\omega} | \boldsymbol{\theta}, \mathbf{o}] [N_{s,t} | N, \boldsymbol{\omega}] [N] [\boldsymbol{\theta}]. \quad (\text{A.11})$$

As stated in the main article, we now model the counts \mathbf{C} as arising from a binomial (as opposed to a Poisson) sampling process, where

$$C_{s,t} \sim \text{Binomial}(N_{s,t}; p_{s,t}),$$

where the fraction of sampling unit s surveyed at time t ($p_{s,t}$) is known with certainty. Abundance in sampling unit s at time t , $N_{s,t}$ is modeled using a multinomial distribution, where

$$N_{s,t} \sim \text{Multinomial}(N; \boldsymbol{\pi}).$$

Multinomial cell probabilities are modeled with a multinomial logit link: namely,

$$\pi_{i,t} = \frac{\exp(\omega_{i,t})}{\sum_s \exp(\omega_{s,t})}.$$

We express habitat preference values ($\omega_{s,t}$) using a log-linear formulation that includes spatio-temporal autocorrelation

$$\boldsymbol{\omega}_t = \mathbf{o} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\kappa}_t + \boldsymbol{\epsilon}_t.$$

We use the same prior distributions for parameters as in previous sections, and use the same formulation for $\boldsymbol{\kappa}$ as for the STPC model. For absolute abundance, we use the scale prior

$$[N] \propto N^{-1}$$

as suggested by [Link \(2013\)](#).

4.1 Bayesian analysis of the CPIF model

Our strategy for simulating from eqn. [A.11](#) is to condition on cells where data are observed to estimate the latent parameters $\boldsymbol{\omega}$ and regression parameters, and then to simulate values of $N_{s,t}$ and N using posterior prediction. To do this, we employ the following collapsed Gibbs sampler:

1. Updating $\boldsymbol{\omega}$

We employ a different strategy for updating $\omega_{s,t}$ depending upon whether sample unit s was surveyed at time t or not. For each time t , we block update all surveyed units ω_t with a Langevin-Hastings step (see e.g., [Givens & Hoeting 2005](#), section 7.1.4). Letting \mathcal{S}_t denote the set of surveyed sample units at time t (note also that ω_t has dimension $[\mathcal{S}_t \times 1]$), the full conditional for time t is given as

$$[\omega_t|\cdot] \propto \text{Normal}(\omega_t; \mathbf{o}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\kappa}_t, \tau_\epsilon^{-0.5}\mathcal{D}_t) \times \prod_{s \in \mathcal{S}} \xi_{s,t}^{C_{s,t}}.$$

Here, \mathcal{D}_t is an $(\mathcal{S}_t \times \mathcal{S}_t)$ identity matrix. We use ξ in place of π since we are conditioning on surveyed cells only, but the calculation is similar:

$$\xi_{s,t} = \frac{\exp(\omega_{s,t})}{\sum_{s \in \mathcal{S}} \exp(\omega_{s,t})}.$$

For each $(s, t) \in \mathcal{U}$ (\mathcal{U} denoting the set of times and locations when surveying does not occur), we simply simulate

$$\omega_{s,t} \sim \text{Normal}(o_{s,t} + \mathbf{X}_{s,t}\boldsymbol{\beta} + \boldsymbol{\kappa}_{s,t}, \tau_\epsilon^{-1}).$$

2-4. Updating $\boldsymbol{\beta}$, τ_ϵ , $\boldsymbol{\alpha}$, and τ_α

Updates of these parameters are conducted in the same manner as for the STPC model, simply replacing $\mu_{s,t}$ with $\omega_{s,t}$. Recall that $\boldsymbol{\alpha}$ are temporally evolving weights associated with each of m knot locations. Spatio-temporal effects are modeled in the same manner as for the STPC model.

5. Updating total abundance

The total count of animals for surveys that occurred at time t , $C_t = \sum_{s \in \mathcal{S}} C_{s,t}$, is distributed as

$$C_t \sim \text{Binomial}(N, v_t),$$

where

$$v_t = \sum_{s \in \mathcal{S}} \pi_{s,t} p_{s,t},$$

and where the full set of multinomial cell probabilities $\pi_{s,t}$ is

$$\pi_{s,t} = \frac{\exp(\omega_{s,t})}{\sum_s \exp(\omega_{s,t})}.$$

Using this approach, the full conditional distribution for N is given (up to a proportionality constant) as

$$[N|\cdot] \propto N^{-1} \frac{(N!)^T}{\prod_t (N - C_t)!} \prod_t (1 - v_t)^{N - C_t}. \quad (\text{A.12})$$

We used the MH algorithm to sample from eqn. A.12. Posterior predictions of abundance across the landscape could then be generated using

$$\mathbf{N}_t \sim \text{Multinomial}(N, \boldsymbol{\pi}_t).$$

Literature Cited

- Calder, C. A., Holloman, C. & Higdon, D. (2002). Exploring space-time structure in ozone concentration using a dynamic process convolution model. *Case Studies in Bayesian Statistics 6* (eds. C. Gatsonis, R. E. Kass, A. Carriquiry, A. Gelman, D. Higdon, D. K. Pauler & I. Verdinelli), pp. 165–176. Springer-Verlag.
- Durbin, J. & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, **89**, 603–615.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004). *Bayesian Data Analysis, 2nd Edition*. Chapman and Hall, Boca Raton.
- Givens, G. H. & Hoeting, J. A. (2005). *Computational Statistics*. Wiley, Hoboken, New Jersey.
- Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, **5**, 173–190.
- Johnson, D. S., Ream, R. R., Towell, R. G., Williams, M. T. & Leon Guerrero, J. D. (2013).

- Bayesian clustering of animal abundance trends for inference and dimension reduction. *Journal of Agricultural, Biological, and Environmental Statistics*, **18**, 299–313.
- Johnson, D. S., Thomas, D. L., Ver Hoef, J. M. & Christ, A. (2008). A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, **64**, 968–976.
- Lele, S. R. & Keim, J. L. (2006). Weighted distributions and estimation of resource selection probability functions. *Ecology*, **87**, 3021–3028.
- Link, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial N. *Ecology*, **94**, 2173–2179.
- Patil, G. P. & Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, **34**, 179–189.
- Rue, H. & Held, L. (2005). *Gaussian Markov Random Fields*. Chapman & Hall/CR, Boca Raton, Florida, USA.
- Ver Hoef, J. & Jansen, J. (2007). Space-time zero-inflated count models of harbor seals. *Environmetrics*, **18**, 697–712.