MCMC samplers for spatially-varying coefficient models in spOccupancy

Jeffrey W. Doser

2023

Contents

1   Introduction

This document provides statistical details on the MCMC algorithms used to fit spatially-varying coefficient (SVC) models in `spOccupancy`. In particular, we discuss the Gibbs samplers for the following two models presented in (Doser et al. 2024):

   1. A single-species SVC occupancy model using `svcPGOcc()`

2 Single-species spatially-varying coefficient occupancy model

2.1 Model description

Let $\boldsymbol{s}_j$ denote the spatial coordinates of site j, where $j = 1, \ldots, J$. We define $z(\boldsymbol{s}_j)$ as the true presence (1) or absence (0) of the target species at site j with spatial coordinates $\boldsymbol{s}_j$. We model $z(\boldsymbol{s}_j)$ as

$$z(\boldsymbol{s}_j) \sim \text{Bernoulli}(\psi(\boldsymbol{s}_j)), \tag{1}$$

where $\psi(\boldsymbol{s}_j)$ is the occupancy probability of the species at site j. We model $\psi(\boldsymbol{s}_j)$ according to

$$\text{logit}(\psi(\boldsymbol{s}_j)) = (\beta_1 + \delta_1 \text{w}_1(\boldsymbol{s}_j)) + \sum_{h=2}^{H} \text{x}_h(\boldsymbol{s}_j)\{\beta_h + \delta_h \text{w}_h(\boldsymbol{s}_j)\}, \tag{2}$$

where β_1 is an intercept, $\text{x}_h(\boldsymbol{s}_j)$ is the hth covariate with $h = 2, \ldots, H$, β_h is the non-spatial effect of covariate $\text{x}_h(\boldsymbol{s}_j)$, and $\text{w}_1(\boldsymbol{s}_j)$ and $\text{w}_h(\boldsymbol{s}_j)$ are spatially-varying effects for the intercept and covariates, respectively. We use indicator variables δ_h for $h = 1, \ldots, H$ to indicate those covariates whose effects vary spatially ($\delta_h = 1$) and those whose effects are assumed constant ($\delta_h = 0$). Note that the model reduces to a traditional single-species occupancy model when $\delta_h = 0$ for all h and a spatial occupancy model (Johnson et al. 2013; Doser et al. 2022) when $\delta_1 = 1$ and $\delta_h = 0$ for all $h > 1$. For later use, define $\tilde{H}$ as the total number of spatially-varying effects estimated in the model (i.e., $\tilde{H} = \sum_{h=1}^{H} \delta_h$), define $\tilde{\text{x}}(\boldsymbol{s}_j)$ as the $\tilde{H} \times 1$ vector of covariates at location j (including an intercept if applicable), and define $\tilde{\beta}_h(\boldsymbol{s}_j) = \beta_h + \text{w}_h(\boldsymbol{s}_j)$ as the spatially-varying coefficients for those effects with $\delta_h = 1$.

Let $\mathcal{L} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_J\}$ be the set of sampled spatial locations, and define $\mathbf{w}_h$ as a $J \times 1$ vector of the spatial random effects for covariate h for each of the $\tilde{H}$ effects with corresponding $\delta_h = 1$. The spatially-varying effects $\mathbf{w}_h$ serve as local adjustments of the covariate effects (or intercept) at each site j from the overall non-spatial effect β_h, resulting in the covariate having a unique effect (i.e., $\tilde{\beta}_h(\boldsymbol{s}_j)$) on species occupancy probability at each site j. Following Alan E. Gelfand et al. (2003), we envision each $h = 1, \ldots, \tilde{H}$ spatially-varying effect $\text{w}_h(\boldsymbol{s}_j)$ as a realization of a smooth latent surface $\{\text{w}_h(\boldsymbol{s}) \mid \boldsymbol{s} \in \mathcal{D}\}$, where $\mathcal{D}$ is the geographical domain of interest. Following standard approaches for modeling species distributions (e.g., Latimer et al. (2006)), we use Gaussian Processes (GPs) to model each of the $\tilde{H}$ smooth functions across the spatial domain. By definition, a GP model for the hth spatially-varying surface $\{\text{w}_h(\boldsymbol{s})\}$ implies that for any finite set of locations $\mathcal{L}$, the vector of random effects $\mathbf{w}_h$ follows a zero-mean multivariate Gaussian distribution with a $J \times J$ covariance matrix $\boldsymbol{C}_h(\boldsymbol{s}, \boldsymbol{s}', \boldsymbol{\theta}_h)$ that is a function of the distances between any pair of site coordinates $\boldsymbol{s}$ and $\boldsymbol{s}'$ and a set of parameters ($\boldsymbol{\theta}_h$) that govern the spatial process according to a parametric covariance function. In our subsequent simulations and case study, we use an exponential covariance function such that $\boldsymbol{\theta}_h = \{\sigma_h^2, \phi_h\}$, where σ_h^2 is a spatial variance parameter and ϕ_h is a spatial decay parameter. Large values of σ_h^2 indicate large variation in the magnitude of a covariate effect across space, while values of σ_h^2 close to 0 suggest little spatial variability in the magnitude of the effect. ϕ_h controls the distance-dependent decay of the spatial dependence in the covariate effect and is inversely related to the spatial range, such that when ϕ_h is small, the covariate effect has a larger range of spatial dependence and varies more smoothly across space compared to larger values of ϕ_h. When using an exponential correlation function, the effective spatial range, or the distance at which the spatial correlation between points drops to 0.05 (Banerjee, Carlin, and Gelfand 2014), corresponds to approximately $3 / \phi$ (i.e., since $3 \approx -\log(0.05)$).

Both frequentist and Bayesian estimation of the model defined by (1) and (2) requires taking the inverse and determinant of $\tilde{H}$ dense $J \times J$ covariance matrices (i.e., $\boldsymbol{C}_h(\boldsymbol{s}, \boldsymbol{s}', \boldsymbol{\theta}_h)$), which involves $O(J^3)$ computations for each of the $\tilde{H}$ spatially-varying coefficients (floating point operations or FLOPs), which quickly renders such an approach impractical for even moderately sized data sets (i.e., hundreds of spatial locations). Here we replace the GP prior for the spatially-varying coefficients with a Nearest Neighbor Gaussian Process (NNGP)

prior (Datta et al. 2016). The NNGP is a valid GP that is based on writing the full multivariate Gaussian distribution for $\mathbf{w}_h$ as a product of conditional densities, such that

$$p(\mathbf{w}_h) = p(\mathrm{w}_h(\boldsymbol{s}_1)) \cdot p(\mathrm{w}_h(\boldsymbol{s}_2) \mid \mathrm{w}_h(\boldsymbol{s}_1)) \cdots p(\mathrm{w}_h(\boldsymbol{s}_J) \mid \mathrm{w}_h(\boldsymbol{s}_{J-1}), \ldots, \mathrm{w}_h(\boldsymbol{s}_1)), \tag{3}$$

where $p(\cdot)$ denotes a probability density function. The NNGP prior achieves computational efficiency by replacing the conditioning sets on the right-hand side of (3) with a set of new conditioning sets, whose maximum size is determined by a pre-specified number of neighbors, m, where $m << J$. Datta et al. (2016) showed that $m = 15$ provides nearly identical inference to the full GP under a variety of scenarios. Let $N(\boldsymbol{s}_j)$ denote the set of at most m neighbors for location $\boldsymbol{s}_j$. Following Vecchia (1988), we set $N(\boldsymbol{s}_j)$ to be the set of at most m nearest neighbors of $\boldsymbol{s}_j$ from $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{j-1}\}$ with respect to Euclidean distance. Note that this requires the set of $\mathcal{L}$ locations to have some prespecified ordering, where here we order the coordinates along the horizontal axis. Through careful construction of the neighbor sets and set of spatial locations as a directed acyclic graph, Gaussian distribution theory reveals the NNGP prior yields a new joint density for $\mathbf{w}_h$, denoted $\tilde{p}(\mathbf{w}_h)$. Let $\mathbf{w}_h(N(\boldsymbol{s}_j))$ denote the at most m realizations of the hth NNGP at the locations in the neighbor set $N(\boldsymbol{s}_j)$. Let $C(\cdot, \boldsymbol{\theta}_h)$ denote the covariance function of the original GP from which the hth NNGP is derived. For any two sets A_1 and A_2, define $\mathrm{C}_{A_1, A_2}(\boldsymbol{\theta}_h)$ as the covariance matrix between the observations in A_1 and A_2 for the hth GP. For all $h = 1, \ldots, H$, our NNGP prior for $\mathbf{w}_h$ thus takes the form

$$\tilde{p}(\mathbf{w}_h) = \prod_{j=1}^{J} \mathrm{Normal}(\mathrm{w}_h(\boldsymbol{s}_j) \mid \mathbf{b}_h(\boldsymbol{s}_j)\mathbf{w}_h(N(\boldsymbol{s}_j)), \mathrm{f}_h(\boldsymbol{s}_j)), \tag{4}$$

where $\mathbf{b}_h(\boldsymbol{s}_j)$ is defined as

$$\mathbf{b}_h(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j, N(\boldsymbol{s}_j)}(\boldsymbol{\theta}_h)\mathbf{C}_{N(\boldsymbol{s}_j), N(\boldsymbol{s}_j)}^{-1}(\boldsymbol{\theta}_h), \tag{5}$$

and $\mathrm{f}_h(\boldsymbol{s}_j)$ is defined as

$$\mathrm{f}_h(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j, \boldsymbol{s}_j}(\boldsymbol{\theta}_h) - \mathbf{C}_{\boldsymbol{s}_j, N(\boldsymbol{s}_j)}(\boldsymbol{\theta}_h)\mathbf{C}_{N(\boldsymbol{s}_j), N(\boldsymbol{s}_j)}^{-1}(\boldsymbol{\theta}_h)\mathbf{C}_{N(\boldsymbol{s}_j), \boldsymbol{s}_j}(\boldsymbol{\theta}_h). \tag{6}$$

To account for imperfect detection in an occupancy modeling framework, $k = 1, \ldots, K(\boldsymbol{s}_j)$ sampling replicates are obtained at each site j to estimate whether a nondetection of the target species is truly an absence (MacKenzie et al. 2002; Tyre et al. 2003). We model the observed detection (1) or nondetection (0) of a study species at site j, denoted $y_k(\boldsymbol{s}_j)$, conditional on the true latent occupancy process $z(\boldsymbol{s}_j)$, following

$$y_k(\boldsymbol{s}_j) \mid z(\boldsymbol{s}_j) \sim \mathrm{Bernoulli}(p_k(\boldsymbol{s}_j)z(\boldsymbol{s}_j)), \tag{7}$$

where $p_k(\boldsymbol{s}_j)$ is the probability of detecting the species at site j during replicate k given the species is truly present at the site. We model detection probability as a function of site and/or observation-level covariates according to

$$\mathrm{logit}(p_k(\boldsymbol{s}_j)) = \mathbf{v}_k(\boldsymbol{s}_j)^{\top}\boldsymbol{\alpha}, \tag{8}$$

where $\boldsymbol{\alpha}$ is a vector of regression coefficients (including an intercept) that describe the effect of site and/or observation covariates $\mathbf{v}_k(\boldsymbol{s}_j)$ on detection. Note that following the standard occupancy model, we assume independence between the replicate surveys, conditional on the true occupancy status $z(\boldsymbol{s}_j)$ and covariates $\mathbf{v}_k(\boldsymbol{s}_j)$, the true occupancy status does not change over the $K(\boldsymbol{s}_j)$ replicate surveys, and there are no false positives (i.e., if $z(\boldsymbol{s}_j) = 0$, then $y_k(\boldsymbol{s}_j) = 0$ for all k). We assume effects of covariates on detection probability are constant over space, but in principle spatially-varying covariate effects could be added to the detection model using the same process described above.

We assign independent Gaussian priors to all non-spatial regression coefficients ($\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$), independent inverse-Gamma priors to the spatial variance parameters (σ_h^2) for each spatially-varying effect, and independent uniform priors for the spatial decay parameters (ϕ_h).

3

2.2 Gibbs sampler

2.2.1 Update occurrence auxiliary variables $(\omega_\beta(\boldsymbol{s}_j))$

We first sample the occurrence auxiliary variable $(\omega_\beta(\boldsymbol{s}_j)$ individually for each site j. Our full conditional is

$$\omega_\beta(\boldsymbol{s}_j) \mid \cdot \sim \mathrm{PG}(1, \mathbf{x}(\boldsymbol{s}_j)^\top \boldsymbol{\beta} + \tilde{\mathbf{x}}(\boldsymbol{s}_j)^\top \mathbf{w}(\boldsymbol{s}_j)). \tag{9}$$

2.2.2 Update detection auxiliary variables $(\omega_{k,\alpha}(\boldsymbol{s}_j))$

We next update the latent Pólya-Gamma auxiliary variable for the detection process, $\omega_{k,\alpha}(\boldsymbol{s}_j)$, for each replicate k at each site j. Note that we only need to sample $\omega_{k,\alpha}(\boldsymbol{s}_j)$ when $\sum_{k=1}^{K_j} y_k(\boldsymbol{s}_j) = 0$. Following Polson, Scott, and Windle (2013), we have

$$\omega_{k,\alpha}(\boldsymbol{s}_j) \mid \cdot \sim \mathrm{PG}(1, \mathbf{v}(\boldsymbol{s}_j)^\top \boldsymbol{\alpha}). \tag{10}$$

2.2.3 Update non-spatial occurrence regression coefficients $(\boldsymbol{\beta})$

Define $\tilde{\mathbf{w}}(\boldsymbol{s}_j) = \tilde{\mathbf{x}}(\boldsymbol{s}_j)^\top \mathbf{w}(\boldsymbol{s}_j)$, and similarly let $\tilde{\mathbf{w}}$ be the $J \times 1$ vector of $\tilde{\mathbf{w}}(\boldsymbol{s}_j)$ for all $j = 1, \ldots, J$. Let $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_{0,\beta}, \boldsymbol{\Sigma}_\beta)$ denote the prior distribution for $\boldsymbol{\beta}$, with $\boldsymbol{\Sigma}_\beta$ being a diagonal matrix. The Pólya-Gamma scheme induces a Gibbs update for the occurrence regression coefficients, which are updated at each iteration according to

$$\boldsymbol{\beta} \mid \cdot \sim \mathrm{Normal}\Big([\boldsymbol{\Sigma}_\beta^{-1} + \mathbf{X}^\top \boldsymbol{S}_\beta \mathbf{X}]^{-1}[\mathbf{X}^\top(\boldsymbol{z} - 0.5\mathbf{1}_J - \boldsymbol{S}_\beta \tilde{\mathbf{w}}) + \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\mu}_{0,\beta}], [\boldsymbol{\Sigma}_\beta^{-1} + \mathbf{X}^\top \boldsymbol{S}_\beta \mathbf{X}]^{-1}\Big), \tag{11}$$

where $\boldsymbol{S}_\beta$ is a diagonal $J \times J$ matrix with diagonal entries equal to the latent PG variable values $(\omega_{1,\beta}, \ldots, \omega_{J,\beta})$ and $\boldsymbol{z}$ is the $J \times 1$ vector of the latent occurrence values.

2.2.4 Update detection regression coefficients $(\boldsymbol{\alpha})$

Similarly, let $\boldsymbol{\alpha} \sim N(\boldsymbol{\mu}_{0,\alpha}, \boldsymbol{\Sigma}_\alpha)$ denote the prior distribution for $\boldsymbol{\alpha}$, with $\boldsymbol{\Sigma}_\alpha$ being a diagonal matrix. We sample the detection regression coefficients $\boldsymbol{\alpha}$ from

$$\boldsymbol{\alpha} \mid \cdot \sim \mathrm{Normal}\Big([\boldsymbol{\Sigma}_\alpha^{-1} + \tilde{\boldsymbol{V}}^\top \boldsymbol{S}_\alpha \tilde{\boldsymbol{V}}]^{-1}[\tilde{\boldsymbol{V}}^\top(\tilde{\boldsymbol{y}} - 0.5\mathbf{1}_{J^*}) + \boldsymbol{\Sigma}_\alpha^{-1} \boldsymbol{\mu}_{0,\alpha}], [\boldsymbol{\Sigma}_\alpha^{-1} + \tilde{\boldsymbol{V}}^\top \boldsymbol{S}_\alpha \tilde{\boldsymbol{V}}]^{-1}\Big). \tag{12}$$

The detection regression coefficients $\boldsymbol{\alpha}$ are only informed by the locations where $z(\boldsymbol{s}_j) = 1$, since we assume no false positive detections in the standard occupancy model. We define J^* as the total number of sites at the current iteration of the MCMC with $z(\boldsymbol{s}_j) = 1$. $\boldsymbol{S}_\alpha$ is a diagonal matrix with diagonal entries equal to the latent Pólya-Gamma variable values $(\omega_{1,1,\alpha}, \ldots, \omega_{J^*,K_{J^*},\alpha})$. The matrix $\tilde{\boldsymbol{V}}$ is the matrix of detection covariates associated with the sites where $z(\boldsymbol{s}_j) = 1$. Similarly, $\tilde{\boldsymbol{y}}$ is a vector of stacked detection-nondetection data values at the entries associated with $z(\boldsymbol{s}_j) = 1$.

2.2.5 Update spatially-varying coefficients $(\mathbf{w}(\boldsymbol{s}_j))$

Next we update the spatially-varying coefficients $\mathbf{w}(\boldsymbol{s}_j)$ sequentially for each of the $j = 1, \ldots, J$ spatial locations. Let $N(\boldsymbol{s}_j)$ denote the set of m nearest neighbors of $\boldsymbol{s}_j$ among $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{J-1}$. For all model implementations in `spOccupancy`, we order the spatial locations by the horizontal axis. Let $\mathbf{w}_h(N(\boldsymbol{s}_j))$ denote the m realizations of the hth NNGP at the locations in $N(\boldsymbol{s}_j)$, where h iterates across the total $\tilde{H}$ spatially-varying coefficients. Let $C(\cdot, \boldsymbol{\theta}_h)$ denote the covariance function of the original Gaussian Process (GP) from which the hth NNGP is derived. For any two sets A_1 and A_2, define $\mathrm{C}_{A_1,A_2}(\boldsymbol{\theta}_h)$ as the covariance matrix between the observations in A_1 and A_2 for the hth GP. For $j \geq 1$, we have

$$\mathbf{b}_h(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j, N(\boldsymbol{s}_j)}(\boldsymbol{\theta}_h)\mathbf{C}_{N(\boldsymbol{s}_j), N(\boldsymbol{s}_j)}^{-1}(\boldsymbol{\theta}_h), \tag{13}$$

where $\mathbf{b}_h(\boldsymbol{s}_1) = \mathbf{0}$ for all $h = 1, \ldots, \tilde{H}$. Further, we have

$$f_h(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j, \boldsymbol{s}_j}(\boldsymbol{\theta}_h) - \mathbf{C}_{\boldsymbol{s}_j, N(\boldsymbol{s}_j)}(\boldsymbol{\theta}_h) \mathbf{C}_{N(\boldsymbol{s}_j), N(\boldsymbol{s}_j)}^{-1}(\boldsymbol{\theta}_h) \mathbf{C}_{N(\boldsymbol{s}_j), \boldsymbol{s}_j}(\boldsymbol{\theta}_h). \tag{14}$$

For any two locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, if $\boldsymbol{s}_1 \in N(\boldsymbol{s}_2)$ and is the lth member of $N(\boldsymbol{s}_2)$, then define $b_h(\boldsymbol{s}_2, \boldsymbol{s}_1)$ as the lth entry of $\mathbf{b}_h(\boldsymbol{s}_2)$. Let $U(\boldsymbol{s}_1) = \{\boldsymbol{s}_2 \in S \mid \boldsymbol{s}_1 \in N(\boldsymbol{s}_2)\}$ be the collection of locations $\boldsymbol{s}_2$ for which $\boldsymbol{s}_1$ is a neighbor, where S is the set of all J spatial locations. For every $\boldsymbol{s}_2 \in U(\boldsymbol{s}_1)$, define $a_h(\boldsymbol{s}_2, \boldsymbol{s}_1) = \text{w}_h(\boldsymbol{s}_2) - \sum_{\boldsymbol{s} \in N(\boldsymbol{s}_2), \boldsymbol{s} \neq \boldsymbol{s}_2} \text{w}_h(\boldsymbol{s}) b_h(\boldsymbol{s}_2, \boldsymbol{s})$. Extending this to matrix notation, let $\boldsymbol{B}(\boldsymbol{s}_j)$ be a $\tilde{H} \times m\tilde{H}$ block matrix, with $\tilde{H} \times \tilde{H}$ diagonal blocks containing the elements of $\boldsymbol{b}_h(\boldsymbol{s}_j)$ for each of the $h = 1, \ldots \tilde{H}$ spatial NNGPs for each of the specific m neighbors. Let $\boldsymbol{F}(\boldsymbol{s}_j)$ be a $\tilde{H} \times \tilde{H}$ diagonal matrix with diagonal elements of $f_h(\boldsymbol{s}_j)$. Let $\boldsymbol{a}(\boldsymbol{s}, \boldsymbol{s}_j)$ contain the values $a_h(\boldsymbol{s}, \boldsymbol{s}_j)$ for each of the $h = 1, \ldots, \tilde{H}$ spatially-varying coefficients. Using this notation, the full conditional for $\mathbf{w}(\boldsymbol{s}_j)$ is

$$\begin{aligned}
\mathbf{w}(\boldsymbol{s}_j) &\mid \cdot N_{\tilde{H}}(\boldsymbol{\mu}_j \boldsymbol{\Sigma}_j, \boldsymbol{\Sigma}_j) \text{ where,} \\
\boldsymbol{\mu}_j &= \boldsymbol{F}(\boldsymbol{s}_j)^{-1} \boldsymbol{B}(\boldsymbol{s}_j) \mathbf{w}(N(\boldsymbol{s}_j)) + \sum_{\boldsymbol{s} \in U(\boldsymbol{s}_j)} \boldsymbol{B}(\boldsymbol{s}, \boldsymbol{s}_j)^\top \boldsymbol{F}(\boldsymbol{s}_j)^{-1} \boldsymbol{a}(\boldsymbol{s}, \boldsymbol{s}_j) + \\
&\quad \tilde{\mathbf{x}}(\boldsymbol{s}_j) \omega_\beta(\boldsymbol{s})_j ((z(\boldsymbol{s}_j) - 0.5)/\omega_\beta(\boldsymbol{s}_j) - \mathbf{x}(\boldsymbol{s}_j)^\top \boldsymbol{\beta}) \text{ and} \\
\boldsymbol{\Sigma}_j &= (\boldsymbol{F}(\boldsymbol{s}_j)^{-1} + \sum_{\boldsymbol{s} \in U(\boldsymbol{s}_j)} \boldsymbol{B}(\boldsymbol{s}, \boldsymbol{s}_j)^\top \boldsymbol{F}(\boldsymbol{s}_j)^{-1} \boldsymbol{B}(\boldsymbol{s}, \boldsymbol{s}_j) + \tilde{\mathbf{x}}(\boldsymbol{s}_j)^\top \boldsymbol{S}_{j,\beta} \tilde{\mathbf{x}}(\boldsymbol{s})_j)^{-1},
\end{aligned} \tag{15}$$

where $\mathbf{w}(N(\boldsymbol{s}_j))$ is a stacked $mq \times 1$ vector of the m realizations of each of the h NNGPs at the locations in $N(\boldsymbol{s}_j)$ and $\boldsymbol{S}_{j,\beta}$ is a $\tilde{H} \times \tilde{H}$ diagonal matrix with each diagonal entry equal to the Pólya-Gamma auxiliary variable at site j.

2.2.6 Update spatial decay parameters (ϕ)

We use a Metropolis within Gibbs step to sample $\boldsymbol{\phi}$, the spatial decay parameters for each spatially-varying coefficient. We assign uniform priors for all spatial decay parameters The full conditional posterior density for ϕ_h for each $h = 1, \ldots, \tilde{H}$ is proportional to

$$\begin{aligned}
p(\phi_h \mid \cdot) &\propto p_h(\phi_h) p(\mathbf{w}_h \mid \phi_h) \\
&\propto p(\phi_h) \times \prod_{j=1}^{J} N\left(\text{w}_h(\boldsymbol{s}_j) \mid \mathbf{b}_h(\boldsymbol{s}_j) \mathbf{w}_h(N(\boldsymbol{s}_j)), f_h(\boldsymbol{s}_j).\right)
\end{aligned} \tag{16}$$

The same update is used for the spatial smoothness parameter ν_h if using a Matérn correlation function.

2.2.7 Update spatial variances ($\boldsymbol{\sigma^2}$)

The default prior for the spatial variance parameter σ_h^2 for each spatially-varying coefficient in `spOccupancy` is an inverse-Gamma distribution (i.e., $\sigma_h^2 \sim IG(a_\sigma, b_\sigma)$). When using an inverse-Gamma prior, we sample from the following full conditional:

$$\sigma_h^2 \mid \cdot \sim IG\left(a_\sigma + J/2, \ b_\sigma + \frac{1}{2} \sum_{j=1}^{J} \left(\text{w}_h(\boldsymbol{s}_j) - \mathbf{b}_h(\boldsymbol{s}_j) \mathbf{w}_h(N(\boldsymbol{s}_j))\right)^2 / f_h^*(\boldsymbol{s}_j),\right. \tag{17}$$

where $f_h^*(\boldsymbol{s}_j) = f_h(\boldsymbol{s}_j)/\sigma_h^2$. Note that to generate σ_h^2 for the lth iteration of the Gibbs sampler, we calculate $f_h^*(\boldsymbol{s}_j)$ using the value of σ_h^2 in the $(h-1)$th iteration.

If using a uniform prior on the spatial variances, the update is analogous to the update for the Metropolis within Gibbs step of the spatial decay parameters in Equation 16.

2.2.8 Update latent occurrence values ($z(\boldsymbol{s}_j)$)

Finally, we sample the latent occurrence states. We set $z(\boldsymbol{s}_j) = 1$ for all sites where there is at least one detection of our species of interest, and so we only need to sample $z(\boldsymbol{s}_j)$ at sites where there are no detections. Thus, for all locations with no detections, we sample $z(\boldsymbol{s}_j)$ according to

$$z(\boldsymbol{s}_j) \mid \cdot \sim \text{Bernoulli}\left(\frac{\psi(\boldsymbol{s}_j)\prod_{k=1}^{K_j}(1 - p_k(\boldsymbol{s}_j))}{1 - \psi(\boldsymbol{s}_j) + \psi(\boldsymbol{s}_j)\prod_{k=1}^{K_j}(1 - p_k(\boldsymbol{s}_j))}\right). \tag{18}$$

3 Multi-species spatially-varying coefficient occupancy model

3.1 Model description

Now consider the case where there are multiple species of interest, N, that are observed during data collection. We seek to jointly model the occupancy of the N species in a single model that accommodates residual correlations between species and allows for sharing of information across species via hierarchical effects. Such multi-species approaches often have improved precision and accuracy of estimates compared to single-species models (Clark et al. 2014; Zipkin et al. 2010). Given the interest in modeling large-scale patterns across space, a subset of spatial locations in $\mathcal{L}$ may fall well outside the potential range of a given species. In our subsequent modeling development, we consider the case where species-specific range maps are available as an auxiliary data source to more readily accommodate large differences in the potential locations where each species can occur.

Using similar notation to the single-species models, we model the true presence-absence state of species i at site $\boldsymbol{s}_j$ following

$$z_i(\boldsymbol{s}_j) \sim \text{Bernoulli}(\psi_i(\boldsymbol{s}_j) \cdot z_i^*(\boldsymbol{s}_j)), \tag{19}$$

where $\psi_i(\boldsymbol{s}_j)$ is the occupancy probability of species i at site j, and $z_i^*(\boldsymbol{s}_j)$ is a binary auxiliary data source indicating whether site j is within the range of species i. Such data can be obtained from a variety of sources, including international databases (e.g., BirdLife International, IUCN) or expert opinion. We suggest buffering the auxiliary data range map by a suitable distance to account for potential inaccuracies in the auxiliary data. Inclusion of such auxiliary range data can drastically reduce the computational burden of the model if certain species can only exist at a subset of the spatial locations in $\mathcal{L}$ (Socolar et al. 2022). If auxiliary range data are not available, $z_i^*(\boldsymbol{s}_j)$ can be removed from (19) (or equivalently, $z_i^*(\boldsymbol{s}_j) = 1$ for all j).

At sites within each species set of potentially suitable areas, we model species-specific occupancy probability as

$$\text{logit}(\psi_i(\boldsymbol{s}_j)) = (\beta_{i,1} + \delta_1 \text{w}_{i,1}^*(\boldsymbol{s}_j)) + \sum_{h=2}^{H} \text{x}_h(\boldsymbol{s}_j)\{\beta_{i,h} + \delta_j \text{w}_{i,h}^*(\boldsymbol{s}_j)\}, \tag{20}$$

with $\text{w}_{i,h}^*(\boldsymbol{s}_j)$ being the spatially-varying component of the intercept ($h = 1$) or coefficient ($h > 1$) for species i at site $\boldsymbol{s}_j$ and all parameters as defined before, but now spatial and non-spatial effects are unique to each species. Note we assume the same variables have spatially-varying effects across species (i.e., $\delta_{i,h} = \delta_h$ for all i), although this could be modified to allow δ_h to vary by species by either setting values *a priori* or estimating the indicator variables within the model itself. We model the non-spatial component of the hth regression coefficient for each species i hierarchically from a common community-level distribution to share information across species (Dorazio and Royle 2005; Alan E. Gelfand et al. 2005). More specifically, we have

$$\beta_{i,h} \sim \text{Normal}(\mu_{\beta_h}, \tau_{\beta_h}^2), \tag{21}$$

where μ_{β_h} is the average non-spatial effect across all species in the community, and $\tau_{\beta_h}^2$ is the variability in the non-spatial effect across all species.

We seek to jointly model the species-specific spatial effects, $\mathbf{w}_{i,t}^*$, to account for correlation in species-specific responses to covariate effects, as well as residual correlation between species after accounting for their relationships with any covariates included in the model. For a small number of species (e.g., 5), a linear model of coregionalization (LMC) framework (Alan E. Gelfand et al. 2004) is a viable solution, but such an approach quickly becomes computationally intractable as the number of species in the community increases. Instead, we use a spatial factor model (Hogan and Tchernis 2004), a dimension reduction technique that accounts for correlations in species-specific responses, while drastically reducing computational run time compared to a LMC that requires estimation of a full $N \times N$ cross-covariance matrix for each effect that is assumed to vary spatially. Here, we decompose $\mathrm{w}_{i,h}^*(\boldsymbol{s}_j)$ into a linear combination of q latent factors and their associated species-specific coefficients (i.e., factor loadings). Thus for each SVC in the model, we have

$$\mathrm{w}_{i,h}^*(\boldsymbol{s}_j) = \boldsymbol{\lambda}_{i,h}^\top \mathbf{w}_h(\boldsymbol{s}_j), \tag{22}$$

where $\boldsymbol{\lambda}_{i,h}^\top$ is the ith row of factor loadings from the $N \times q$ loadings matrix $\boldsymbol{\Lambda}_h$, and $\mathbf{w}_h(\boldsymbol{s}_j)$ is a $q \times 1$ vector of independent spatial factors at site j. As in the single-species SVC occupancy model, we model each of the $r = 1, \ldots, q$ spatial factors for each of the $\tilde{H}$ spatially-varying effects with an NNGP prior following

$$\tilde{p}(\mathbf{w}_{r,h}) = \prod_{j=1}^{J} \mathrm{Normal}(\mathrm{w}_{r,h}(\boldsymbol{s}_j) \mid \mathbf{b}_{r,h}(\boldsymbol{s}_j)\mathbf{w}_{r,h}(N(\boldsymbol{s}_j)), \mathrm{f}_{r,h}(\boldsymbol{s}_j)), \tag{23}$$

with $\mathbf{b}_{r,h}(\boldsymbol{s}_j)$ and $\mathrm{f}_{r,h}(\boldsymbol{s}_j)$ defined in (5) and (6), respectively.

For each SVC, we can derive an inter-species covariance matrix $\boldsymbol{\Sigma}_h = \boldsymbol{\Lambda}_h \boldsymbol{\Lambda}_h^\top$, which has rank $q << N$, and, thus, is singular. However, the inter-species covariance matrices can still be used to detect species clustering (Shirota, Gelfand, and Banerjee 2019; Doser, Finley, and Banerjee 2023). For a spatially-varying intercept, the inter-species covariance matrix provides information on the residual co-occurrence patterns between each pair of species across space after accounting for any covariates included in the model, which can be used to generate hypotheses regarding the abiotic and/or biotic drivers of residual species co-occurrence patterns (e.g., Tobler et al. (2019)). For a spatially-varying covariate effect, a positive correlation between species in $\boldsymbol{\Sigma}_h$ indicates similar responses to an environmental covariate across space, which can be used as a model-based ordination technique to identify groups of species that respond similarly to changes in environmental variables.

Analogous to the single-species case, we model the observed detection-nondetection of each species i at site j during replicate survey k, $y_{i,k}(\boldsymbol{s}_j)$ conditional on the true presence-absence of each species, $z_i(\boldsymbol{s}_j)$, following (7) and (8), with all parameters now varying by species. We model the species-specific detection regression coefficients ($\boldsymbol{\alpha}_i$) hierarchically, analogous to the non-spatial occupancy regression coefficients in (21).

We assume Gaussian priors for all mean parameters and inverse-Gamma priors for variance parameters. Additional restrictions on the factors loadings matrix $\boldsymbol{\Lambda}_h$ for each spatially-varying coefficient h are required to ensure identifiability (Taylor-Rodriguez et al. 2019). We fix all elements in the upper triangle to 0 and set the diagonal elements to 1. We additionally fix the spatial variance parameters σ_h^2 of each latent spatial process to 1. We assign standard Gaussian priors for the lower triangular elements in $\boldsymbol{\Lambda}$ and assign each spatial range parameter $\phi_{r,h}$ an independent uniform prior.

3.2 Gibbs sampler

Here we describe the Gibbs sampler for fitting the spatially-varying coefficient multi-species occupancy model using `svcMsPGOcc()`.

3.2.1 Update community-level occurrence coefficients ($\boldsymbol{\mu}_\beta$)

We first sample all community-level parameters followed by species level parameters. First we sample the community-level occurrence coefficients. Let $\boldsymbol{\mu}_\beta$ denote the vector of all community-level occurrence means, and similarly let $\boldsymbol{T}_\beta$ denote the variance matrix of all community-level occurrence variance parameters. Note that $\boldsymbol{T}_\beta$ is a diagonal matrix. Let $\boldsymbol{\mu}_\beta \sim N(\boldsymbol{\mu}_{0,\beta}, \boldsymbol{\Sigma}_\beta)$ denote our prior distribution, where $\boldsymbol{\Sigma}_\beta$ is a

diagonal matrix. Note this is equivalent to assigning an independent normal prior for each coefficient. Our full conditional for the community-level regression coefficients $\boldsymbol{\mu_\beta}$ is then

$$\boldsymbol{\mu_\beta} \mid \cdot \sim N([\boldsymbol{\Sigma}_\beta^{-1} + N\boldsymbol{T}_\beta^{-1}]^{-1}\Big[\sum_{i=1}^N (\boldsymbol{T}_\beta^{-1}\boldsymbol{\beta}_i) + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_{0,\beta}\Big], [\boldsymbol{\Sigma}_\beta^{-1} + N\boldsymbol{T}_\beta^{-1}]^{-1}). \tag{24}$$

3.2.2 Update community-level detection coefficients ($\boldsymbol{\mu_\alpha}$)

Next, we sample the community-level detection coefficients. Let $\boldsymbol{\mu_\alpha}$ denote the vector of all community-level detection means, and similarly let $\boldsymbol{T}_\alpha$ denote the diagonal variance matrix of all community-level detection variance parameters. Let $\boldsymbol{\mu_\alpha} \sim N(\boldsymbol{\mu}_{0,\alpha}, \boldsymbol{\Sigma}_\alpha)$ denote the prior distribution, where $\boldsymbol{\Sigma}_\alpha$ is a diagonal matrix. Our full conditional then takes the form

$$\boldsymbol{\mu_\alpha} \mid \cdot \sim N([\boldsymbol{\Sigma}_\alpha^{-1} + N\boldsymbol{T}_\alpha^{-1}]^{-1}\Big[\sum_{i=1}^N (\boldsymbol{T}_\alpha^{-1}\boldsymbol{\alpha}_i) + \boldsymbol{\Sigma}_\alpha^{-1}\boldsymbol{\mu}_{0,\alpha}\Big], [\boldsymbol{\Sigma}_\alpha^{-1} + N\boldsymbol{T}_\alpha^{-1}]^{-1}). \tag{25}$$

3.2.3 Update community-level occurrence variances ($\boldsymbol{\tau_\beta^2}$)

Let $\tau_{h,\beta}^2$ denote the community-level variance for the hth occurrence parameter ($h = 1, \ldots, p_\psi$). We assign an inverse gamma prior to $\tau_{h,\beta}^2$ with shape parameter $a_{\tau_{h,\beta}}$ and scale parameter $b_{\tau_{h,\beta}}$. Our full conditional is then

$$\tau_{h,\beta}^2 \mid \cdot \sim \text{IG}(a_{\tau_{h,\beta}} + \frac{N}{2}, b_{\tau_{h,\beta}} + \frac{\sum_{i=1}^N (\beta_{i,h} - \mu_{\beta_h})^2}{2}). \tag{26}$$

3.2.4 Update community-level detection variances ($\boldsymbol{\tau_\alpha^2}$)

Let $\tau_{h,\alpha}^2$ denote the community-level variance for the hth detection parameter ($h = 1, \ldots, p_H$). We assign an inverse gamma prior to $\tau_{h,\alpha}^2$ with shape parameter $a_{\tau_{h,\alpha}}$ and scale parameter $b_{\tau_{h,\alpha}}$. Our full conditional is then

$$\tau_{h,\alpha}^2 \mid \cdot \sim \text{IG}(a_{\tau_{h,\alpha}} + \frac{N}{2}, b_{\tau_{h,\alpha}} + \frac{\sum_{i=1}^N (\alpha_{i,h} - \mu_{\alpha_h})^2}{2}). \tag{27}$$

3.2.5 Update species-specific occurrence auxiliary variables ($\omega_{i,\beta}(\boldsymbol{s}_j)$)

We next sample the occurrence auxiliary variable ($\omega_{i,\beta}(\boldsymbol{s}_j)$) individually for each species i and site j. If $z_i^*(\boldsymbol{s}_j) = 1$ (i.e., site $\boldsymbol{s}_j$ is within the species range as defined by any auxiliary data source supplied to the model), our full conditional is

$$\omega_{i,\beta}(\boldsymbol{s}_j) \mid \cdot \sim \text{PG}(1, \boldsymbol{x}(\boldsymbol{s}_j)^\top \boldsymbol{\beta}_i + \tilde{\text{x}}(\boldsymbol{s}_j)^\top \text{w}_i^*(\boldsymbol{s}_j)). \tag{28}$$

For sites and species where $z_i^*(\boldsymbol{s}_j) = 0$, we do not need to generate an auxiliary variable.

3.2.6 Update species-specific detection auxiliary variables ($\omega_{i,k,\alpha}(\boldsymbol{s}_j)$)

We next update the latent Pólya-Gamma auxiliary variable for the detection process, $\omega_{i,k,\alpha}(\boldsymbol{s}_j)$, for each replicate k at each site j for each species i. Note that we only need to sample $\omega_{i,k,\alpha}(\boldsymbol{s}_j)$ when $z_i(\boldsymbol{s}_j) = 1$, which can change across different MCMC iterations. For such cases when $z_i(\boldsymbol{s}_j) = 0$, set $\omega_{i,k,\alpha}(\boldsymbol{s}_j) = 0$. Following Polson, Scott, and Windle (2013), for all $\boldsymbol{s}_j$ with $z_i(\boldsymbol{s}_j) = 1$ at the current iteration, we have

$$\omega_{i,k,\alpha}(\boldsymbol{s}_j) \mid \cdot \sim \text{PG}(1, \boldsymbol{v}(\boldsymbol{s}_j)^\top \boldsymbol{\alpha}_i). \tag{29}$$

3.2.7 Update species-level occurrence regression coefficients ($\boldsymbol{\beta}_i$)

We update the species-level occurrence regression coefficients ($\boldsymbol{\beta}_i$), including the intercept, from the following multivariate normal full conditional

$$\boldsymbol{\beta}_i \mid \cdot \sim \mathrm{Normal}(\mathbf{L}^*_{\beta,i}\boldsymbol{\mu}^*_{\beta,i}, \mathbf{L}^*_{\beta,i}), \tag{30}$$

where

$$\boldsymbol{\mu}^*_{\beta,i} = \boldsymbol{X}^\top \boldsymbol{Z}^*_i(\boldsymbol{z}_i - 0.5\mathbf{1}_J - \boldsymbol{S}_\beta \sum_{h=1}^{\tilde{H}} \tilde{\boldsymbol{X}}_{h,J}\boldsymbol{Z}^*_i\mathbf{W}^\top_h \boldsymbol{\lambda}_{i,h}) + \boldsymbol{T}^{-1}_\beta \boldsymbol{\mu_\beta}, \tag{31}$$

and

$$\mathbf{L}^*_{\beta,i} = [\boldsymbol{T}^{-1}_\beta + \boldsymbol{X}^\top \boldsymbol{Z}^*_i \boldsymbol{S}_\beta \boldsymbol{Z}^*_i \boldsymbol{X}]^{-1}, \tag{32}$$

where $\boldsymbol{S}_\beta$ is a diagonal $J \times J$ matrix with diagonal entries equal to the latent Pólya-Gamma variable values for species i, $\boldsymbol{z}_i$ is the $J \times 1$ vector of latent occurrence values for species i, $\mathbf{1}_J$ is a $J \times 1$ vector of 1s, $\boldsymbol{Z}^*_i$ is a $J \times J$ diagonal matrix with diagonal elements equal to the auxiliary range binary data points for species i, $\tilde{\boldsymbol{X}}_{h,J}$ is a $J \times J$ diagonal matrix with diagonal elements equal to the hth covariate value at site j whose effect is assumed to vary spatially, $\mathbf{W}_h$ is the $q \times J$ matrix of spatial factors for the hth spatially-varying coefficient, and $\boldsymbol{\lambda}_{i,h}$ is the ith row of factor loadings for the tth spatially-varying coefficient.

3.2.8 Update species-level detection regression coefficients ($\boldsymbol{\alpha}_i$)

Next, we sample the species-specific detection regression coefficients for species i ($\boldsymbol{\alpha}_i$). Define $K_{\mathrm{TOT}} = \sum_{j=1}^J K_j$. Our full conditional update takes the form

$$\boldsymbol{\alpha}_i \mid \cdot \sim \mathrm{Normal}(\mathbf{L}^*_{\alpha,i}\boldsymbol{\mu}^*_{\alpha,i}, \mathbf{L}^*_{\alpha,i}), \tag{33}$$

$$\boldsymbol{\mu}^*_{\alpha,i} = \boldsymbol{V}^\top \boldsymbol{Z}_i(\boldsymbol{y}_i - 0.5\mathbf{1}_{K_{\mathrm{TOT}}}) + \boldsymbol{T}^{-1}_\alpha \boldsymbol{\mu_\alpha}, \tag{34}$$

and

$$\mathbf{L}^*_{\alpha,i} = [\boldsymbol{T}^{-1}_\alpha + \boldsymbol{V}^\top \boldsymbol{Z}_i \boldsymbol{S}_\alpha \boldsymbol{Z}_i \boldsymbol{V}]^{-1}, \tag{35}$$

where $\boldsymbol{S}_\alpha$ is a diagonal $K_{\mathrm{TOT}} \times K_{\mathrm{TOT}}$ matrix with diagnoal entries equal to the latent Pólya-Gamma variable values for species i, $\boldsymbol{y}_i$ is the $K_{\mathrm{TOT}} \times 1$ vector of detection-nondetection values for species i, $\mathbf{1}_{K_{\mathrm{TOT}}}$ is a $K_{\mathrm{TOT}} \times 1$ vector of 1s, and $\boldsymbol{Z}_i$ is a $K_{\mathrm{TOT}} \times K_{\mathrm{TOT}}$ diagonal matrix with diagonal elements equal to the current value of the latent occurrence state $\boldsymbol{z}_i$ at the site corresponding to the given element of $\boldsymbol{y}_i$.

3.2.9 Update latent spatial factors for each SVC ($\mathbf{w}_h(\boldsymbol{s}_j)$)

Let $N(\boldsymbol{s}_j)$ denote the set of m nearest neighbors of $\boldsymbol{s}_j$ among $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{j-1}$. Let $\mathbf{w}_{r,h}(N(\boldsymbol{s}_j))$ denote the m realizations of the rth NNGP factor for the hth spatially-varying coefficient at the locations in $N(\boldsymbol{s}_j)$. Let $C(\cdot, \phi_{r,h})$ denote the correlation function of the original Gaussian Process (GP) from which the NNGP is derived. For any two sets A_1 and A_2, define $\mathrm{C}_{A_1,A_2}(\phi_{r,h})$ as the correlation matrix between the observations in A_1 and A_2 for the rth GP for the hth spatially-varying coefficient. For $j \geq 1$, we have

$$\mathbf{b}_{r,h}(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j, N(\boldsymbol{s}_j)}(\phi_{r,h})\mathbf{C}^{-1}_{N(\boldsymbol{s}_j), N(\boldsymbol{s}_j)}(\phi_{r,h}), \tag{36}$$

where $\mathbf{b}_{r,h}(\boldsymbol{s}_1) = \mathbf{0}$ for all $r = 1, \ldots, q$ and $h = 1, \ldots, \tilde{H}$. Further, we have

$$f_{r,h}(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j, \boldsymbol{s}_j}(\phi_{r,h}) - \mathbf{C}_{\boldsymbol{s}_j, N(\boldsymbol{s}_j)}(\phi_{r,h}) \mathbf{C}_{N(\boldsymbol{s}_j), N(\boldsymbol{s}_j)}^{-1}(\phi_{r,h}) \mathbf{C}_{N(\boldsymbol{s}_j), \boldsymbol{s}_j}(\phi_{r,h}). \tag{37}$$

For any two locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, if $\boldsymbol{s}_1 \in N(\boldsymbol{s}_2)$ and is the lth member of $N(\boldsymbol{s}_2)$, then define $b_{r,h}(\boldsymbol{s}_2, \boldsymbol{s}_1)$ as the lth entry of $\mathbf{b}_{r,h}(\boldsymbol{s}_2)$. Let $U(\boldsymbol{s}_1) = \{\boldsymbol{s}_2 \in S \mid \boldsymbol{s}_1 \in N(\boldsymbol{s}_2)\}$ be the collection of locations $\boldsymbol{s}_2$ for which $\boldsymbol{s}_1$ is a neighbor, where S is the set of all J spatial locations. For every $\boldsymbol{s}_2 \in U(\boldsymbol{s}_1)$, define $a_{r,t}(\boldsymbol{s}_2, \boldsymbol{s}_1) = \mathrm{w}_{r,h}(\boldsymbol{s}_2) - \sum_{\boldsymbol{s} \in N(\boldsymbol{s}_2), \boldsymbol{s} \neq \boldsymbol{s}_2} \mathrm{w}_{r,h}(\boldsymbol{s}) b_{r,h}(\boldsymbol{s}_2, \boldsymbol{s})$. Extending this to matrix notation, let $\boldsymbol{B}_h(\boldsymbol{s}_j)$ be a $q \times mq$ block matrix, with each $q \times q$ diagonal block containing the elements of $\boldsymbol{b}_{r,h}(\boldsymbol{s}_j)$ for each of the $r = 1, \ldots q$ spatial factors for the $h = 1, \ldots, \tilde{H}$ spatially-varying coefficients for each of the specific m neighbors. Let $\boldsymbol{F}_h(\boldsymbol{s}_j)$ be a $q \times q$ diagonal matrix with diagonal elements of $f_{r,h}(\boldsymbol{s}_j)$. Let $\boldsymbol{a}_h(\boldsymbol{s}, \boldsymbol{s}_j)$ contain the values $a_{r,h}(\boldsymbol{s}, \boldsymbol{s}_j)$ for each of the $r = 1, \ldots, q$ latent factors for the hth spatially-varying coefficient. Using this notation, the full conditional distribution for $\mathbf{w}_h(\boldsymbol{s}_j)$ is

$$\mathbf{w}_h(\boldsymbol{s}_j) \mid \cdot \sim N_q(\boldsymbol{\mu}_{j,h} \boldsymbol{\Sigma}_{j,h}, \boldsymbol{\Sigma}_{j,h}) \text{ where,}$$
$$\boldsymbol{\mu}_{j,h} = \boldsymbol{F}_h(\boldsymbol{s}_j)^{-1} \boldsymbol{B}_h(\boldsymbol{s}_j) \mathbf{w}_h(N(\boldsymbol{s}_j)) + \sum_{\boldsymbol{s} \in U(\boldsymbol{s}_j)} \boldsymbol{B}_h(\boldsymbol{s}, \boldsymbol{s}_j)^\top \boldsymbol{F}_h(\boldsymbol{s}_j)^{-1} \boldsymbol{a}_h(\boldsymbol{s}, \boldsymbol{s}_j) +$$
$$\boldsymbol{\Lambda}_h^\top \tilde{\boldsymbol{X}}_{h,N}(\boldsymbol{s}_j) \boldsymbol{Z}_N^*(\boldsymbol{s}_j) \boldsymbol{S}_N(\boldsymbol{s}_j)(\boldsymbol{S}_N(\boldsymbol{s}_j)^{-1}(\boldsymbol{z}(\boldsymbol{s}_j) - 0.5\mathbf{1}_N) - \boldsymbol{Z}_N^*(\boldsymbol{s}_j)(\boldsymbol{X}(\boldsymbol{s}_j)^\top \boldsymbol{\beta} + \sum_{k \neq h}^{\tilde{H}} \tilde{\boldsymbol{X}}_{k,N}(\boldsymbol{s}_j) \boldsymbol{\Lambda}_k \mathbf{w}_k(\boldsymbol{s}_j))) \text{ and}$$
$$\boldsymbol{\Sigma}_{j,h} = (\boldsymbol{F}_h(\boldsymbol{s}_j))^{-1} + \sum_{\boldsymbol{s} \in U(\boldsymbol{s}_j)} \boldsymbol{B}_h(\boldsymbol{s}, \boldsymbol{s}_j)^\top \boldsymbol{F}_h(\boldsymbol{s}_j)^{-1} \boldsymbol{B}_h(\boldsymbol{s}, \boldsymbol{s}_j) +$$
$$(\boldsymbol{\Lambda}_h^\top \tilde{\boldsymbol{X}}_{h,N}(\boldsymbol{s}_j) \boldsymbol{Z}_N^*(\boldsymbol{s}_j)) \boldsymbol{S}_N(\boldsymbol{s}_j)(\boldsymbol{\Lambda}_h^\top \tilde{\boldsymbol{X}}_{h,N}(\boldsymbol{s}_j) \boldsymbol{Z}_N^*(\boldsymbol{s}_j))^\top)^{-1}, \tag{38}$$

where $\mathbf{w}_h(N(\boldsymbol{s}_j))$ is a stacked $mq \times 1$ vector of the m realizations of each of the r NNGPs at the locations in $N(\boldsymbol{s}_j)$, $\boldsymbol{S}_N(\boldsymbol{s}_j)$ is an $N \times N$ diagonal matrix with the Pólya-Gamma occupancy variables for each species i at site j along the diagonal elements, $\tilde{\boldsymbol{X}}_{h,N}(\boldsymbol{s}_j)$ is a $N \times N$ diagonal matrix with diagonal elements all equal to the value of the hth spatially-varying covariate at location $\boldsymbol{s}_j$, $\boldsymbol{Z}_N^*(\boldsymbol{s}_j)$ is a $N \times N$ diagonal matrix with diagonal elements equal to the auxiliary range binary data points at site $\boldsymbol{s}_j$ for each of the N species, $\boldsymbol{X}(\boldsymbol{s}_j)^\top$ is a $N \times (NH)$ block-diagonal matrix with the ith diagonal block the $\boldsymbol{x}(\boldsymbol{s}_j)$ vector of H covariates (including the intercept), and $\boldsymbol{\beta}$ is the $(NH) \times 1$ stacked vector of species-specific regression coefficients (including the intercept).

3.2.10 Update latent spatial factor loadings for each SVC ($\boldsymbol{\Lambda}_h$)

Recall we set all diagonal elements of $\boldsymbol{\Lambda}_h$ to 1 and all upper triangular elements equal to 0 in order to ensure identifiability of the latent spatial factors. Given this requirement, let $q_i = \min\{i - 1, q\}$ for $2 \leq i \leq N$, and let $\tilde{\boldsymbol{\lambda}}_{i,h} = (\lambda_{i,h,1}, \ldots, \lambda_{i,h,q_i})^\top$ be the vector representing the unrestricted elements in the ith row of $\boldsymbol{\Lambda}_h$. Define $\mathbf{W}_h$ as the $J \times q$ matrix of latent spatial factors for the hth SVC, and let $\mathbf{W}_{1:i,h}$ be the first i columns of $\mathbf{W}_h$. Using this notation, the full conditional density for $\tilde{\boldsymbol{\lambda}}_{i,h}$ is $N_q(\boldsymbol{\Omega}_{\tilde{\boldsymbol{\lambda}}_{i,h}} \boldsymbol{\mu}_{\tilde{\boldsymbol{\lambda}}_{i,h}}, \boldsymbol{\Omega}_{\tilde{\boldsymbol{\lambda}}_{i,h}})$, where

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{\lambda}}_{i,h}} = \begin{cases} (\tilde{\boldsymbol{X}}_{h,J} \boldsymbol{Z}_i^* \mathbf{W}_{h,1:(i-1)})^\top \boldsymbol{S}_{i,\beta}(\boldsymbol{S}_{i,\beta}^{-1}(\boldsymbol{z}_i - 0.5\mathbf{1}_J) - \boldsymbol{Z}_i^*(\boldsymbol{X}\boldsymbol{\beta}_i + [\sum_{k \neq h}^{\tilde{H}} \tilde{\boldsymbol{X}}_{k,J} \mathbf{W}_k \boldsymbol{\lambda}_{k,i}] + \tilde{\boldsymbol{X}}_{h,J} \dot{\mathbf{w}}_{h,i})) & \text{if} \quad 2 \leq i \leq q \\ (\tilde{\boldsymbol{X}}_{h,J} \boldsymbol{Z}_i^* \mathbf{W}_h)^\top \boldsymbol{S}_{i,\beta}(\boldsymbol{S}_{i,\beta}^{-1}(\boldsymbol{z}_i - 0.5\mathbf{1}_J) - \boldsymbol{Z}_i^*(\boldsymbol{X}\boldsymbol{\beta}_i + [\sum_{k \neq h}^{\tilde{H}} \tilde{\boldsymbol{X}}_{k,J} \mathbf{W}_k \boldsymbol{\lambda}_{k,i}])) & \text{if} \quad i > q \end{cases} \tag{39}$$

$$\boldsymbol{\Omega}_{\tilde{\boldsymbol{\lambda}}_{i,h}} = \begin{cases} ((\tilde{\boldsymbol{X}}_{h,J} \boldsymbol{Z}_i^* \mathbf{W}_{h,1:(i-1)})^\top \boldsymbol{S}_{i,\beta}(\tilde{\boldsymbol{X}}_{h,J} \boldsymbol{Z}_i^* \mathbf{W}_{h,1:(i-1)}) + I_{i-1})^{-1} & \text{if} \quad 2 \leq i \leq q \\ ((\tilde{\boldsymbol{X}}_{h,J} \boldsymbol{Z}_i^* \mathbf{W}_h)^\top \boldsymbol{S}_{i,\beta}(\tilde{\boldsymbol{X}}_{h,J} \boldsymbol{Z}_i^* \mathbf{W}_h) + I_q)^{-1} & \text{if} \quad i > q \end{cases}, \tag{40}$$

where $\dot{\mathbf{w}}_{h,i}$ is the ith column of $\mathbf{W}_h$, and all other variables are as defined previously.

3.2.11 Update spatial range parameters (ϕ)

We use a Metropolis within Gibbs step to sample $\boldsymbol{\phi}$. The full conditional posterior density for $\phi_{r,h}$ for each $r = 1, \ldots, q$ and $h = 1, \ldots, \tilde{H}$ is proportional to

$$
\begin{aligned}
p(\phi_{r,h} \mid \cdot) &\propto p(\phi_{r,h}) p(\mathbf{w}_{r,h} \mid \phi_{r,h}) \\
&\propto p(\phi_{r,h}) \times \prod_{j=1}^{J} N\left(\mathrm{w}_{r,h}(\boldsymbol{s}_j) \mid \mathbf{b}_{r,h}(\boldsymbol{s}_j) \mathbf{w}_{r,h}(N(\boldsymbol{s}_j)), f_{r,h}(\boldsymbol{s}_j).\right)
\end{aligned}
\tag{41}
$$

We sample $\phi_{r,h}$ using a random walk Metropolis step. We use a normal proposal distribution along with a Jacobian transformation.

3.2.12 Update latent occurrence values ($z_i(\boldsymbol{s}_j)$)

Finally, we sample the latent occurrence states for each species. We set $z_i(\boldsymbol{s}_j) = 1$ for all sites where there is at least one detection of species i, and so we only need to sample $z_i(\boldsymbol{s}_j)$ at sites where there are no detections. We set $z_i(\boldsymbol{s}_j) = 0$ for all sites where $z_i^*(\boldsymbol{s}_j) = 0$. Thus, for all locations with no detections of the species i where $z_i^*(\boldsymbol{s}_j) = 1$, we sample $z_i(\boldsymbol{s}_j)$ according to

$$
z_i(\boldsymbol{s}_j) \mid \cdot \sim \text{Bernoulli}\left(\frac{\psi_i(\boldsymbol{s}_j) \prod_{k=1}^{K_j}(1 - p_{i,k}(\boldsymbol{s}_j))}{1 - \psi_i(\boldsymbol{s}_j) + \psi_i(\boldsymbol{s}_j) \prod_{k=1}^{K_j}(1 - p_{i,k}(\boldsymbol{s}_j))} \right).
\tag{42}
$$

References

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand. 2014. *Hierarchical Modeling and Analysis for Spatial Data.* Chapman; Hall/CRC.

Clark, James S, Alan E Gelfand, Christopher W Woodall, and Kai Zhu. 2014. "More Than the Sum of the Parts: Forest Climate Response from Joint Species Distribution Models." *Ecological Applications* 24 (5): 990–99.

Datta, Abhirup, Sudipto Banerjee, Andrew O Finley, and Alan E Gelfand. 2016. "Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets." *Journal of the American Statistical Association* 111 (514): 800–812.

Dorazio, Robert M, and J Andrew Royle. 2005. "Estimating Size and Composition of Biological Communities by Modeling the Occurrence of Species." *Journal of the American Statistical Association* 100 (470): 389–98.

Doser, Jeffrey W, Andrew O Finley, and Sudipto Banerjee. 2023. "Joint Species Distribution Models with Imperfect Detection for High-Dimensional Spatial Data." *Ecology* 104 (9): e4137.

Doser, Jeffrey W, Andrew O Finley, Marc Kéry, and Elise F Zipkin. 2022. "spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models." *Methods in Ecology and Evolution* 13 (8): 1670–78.

Doser, Jeffrey W, Andrew O Finley, Sarah P Saunders, Marc Kéry, Aaron S Weed, and Elise F Zipkin. 2024. "Modeling Complex Species-Environment Relationships Through Spatially-Varying Coefficient Occupancy Models." *Journal of Agricultural, Biological, and Environmental Statistics.*

Gelfand, Alan E, Hyon-Jung Kim, CF Sirmans, and Sudipto Banerjee. 2003. "Spatial Modeling with Spatially Varying Coefficient Processes." *Journal of the American Statistical Association* 98 (462): 387–96.

Gelfand, Alan E., Alexandra M. Schmidt, Sudipto Banerjee, and C. F. Sirmans. 2004. "Nonstationary Multivariate Process Modeling Through Spatially Varying Coregionalization." *Test* 13 (2): 263–312. https://doi.org/10.1007/BF02595775.

Gelfand, Alan E, Alexandra M Schmidt, Shanshan Wu, John A Silander Jr, Andrew Latimer, and Anthony G Rebelo. 2005. "Modelling Species Diversity Through Species Level Hierarchical Modelling." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54 (1): 1–20.

Hogan, Joseph W, and Rusty Tchernis. 2004. "Bayesian Factor Analysis for Spatially Correlated Data, with Application to Summarizing Area-Level Material Deprivation from Census Data." *Journal of the American Statistical Association* 99 (466): 314–24.

Johnson, Devin S, Paul B Conn, Mevin B Hooten, Justina C Ray, and Bruce A Pond. 2013. "Spatial Occupancy Models for Large Data Sets." *Ecology* 94 (4): 801–8.

Latimer, Andrew M, Shanshan Wu, Alan E Gelfand, and John A Silander Jr. 2006. "Building Statistical Models to Analyze Species Distributions." *Ecological Applications* 16 (1): 33–50.

MacKenzie, Darryl I, James D Nichols, Gideon B Lachman, Sam Droege, J Andrew Royle, and Catherine A Langtimm. 2002. "Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One." *Ecology* 83 (8): 2248–55.

Polson, Nicholas G, James G Scott, and Jesse Windle. 2013. "Bayesian Inference for Logistic Models Using pólya–Gamma Latent Variables." *Journal of the American Statistical Association* 108 (504): 1339–49.

Shirota, S., A. E. Gelfand, and S. Banerjee. 2019. "Spatial Joint Species Distribution Modeling Using Dirichlet Processes." *Statistica Sinica* 29: 1127–54.

Socolar, Jacob B, Simon C Mills, Torbjørn Haugaasen, James J Gilroy, and David P Edwards. 2022. "Biogeographic Multi-Species Occupancy Models for Large-Scale Survey Data." *Ecology and Evolution* 12 (10): e9328.

Taylor-Rodriguez, Daniel, Andrew O Finley, Abhirup Datta, Chad Babcock, Hans-Erik Andersen, Bruce D Cook, Douglas C Morton, and Sudipto Banerjee. 2019. "Spatial factor models for high-dimensional and large spatial data: An application in forest variable mapping." *Statistica Sinica* 29: 1155.

Tobler, Mathias W, Marc Kéry, Francis KC Hui, Gurutzeta Guillera-Arroita, Peter Knaus, and Thomas Sattler. 2019. "Joint Species Distribution Models with Species Correlations and Imperfect Detection." *Ecology* 100 (8): e02754.

Tyre, Andrew J, Brigitte Tenhumberg, Scott A Field, Darren Niejalke, Kirsten Parris, and Hugh P Possingham. 2003. "Improving Precision and Reducing Bias in Biological Surveys: Estimating False-Negative Error Rates." *Ecological Applications* 13 (6): 1790–1801.

Vecchia, Aldo V. 1988. "Estimation and Model Identification for Continuous Spatial Processes." *Journal of the Royal Statistical Society: Series B (Methodological)* 50 (2): 297–312.

Zipkin, Elise F, J Andrew Royle, Deanna K Dawson, and Scott Bates. 2010. "Multi-Species Occurrence Models to Evaluate the Effects of Conservation and Management Actions." *Biological Conservation* 143 (2): 479–84.