MCMC samplers for joint species distribution models in spOccupancy

Jeffrey W. Doser

2022 (last update: July 19, 2023)

Contents

1 Introduction

This vignette provides statistical details on the MCMC algorithms used to fit joint species distribution models in `spOccupancy` (i.e., multi-species occupancy models with species correlations). In particular, we discuss the Gibbs samplers for each of the following four models presented in Doser, Finley, and Banerjee (2023):

1. A spatial latent factor multi-species occupancy model using `sfMsPGOcc()` that accommodates residual species correlations, imperfect detection, and spatial autocorrelation.
2. A latent factor multi-species occupancy model using `lfMsPGOcc()` that accommodates residual species correlations and imperfect detection.
3. A spatial latent factor joint species distribution model using `sfJSDM()` that accommodates residual species correlations and spatial autocorrelation.
4. A latent factor joint species distribution model using `lfJSDM()` that accommodates residual species correlations.

2 Pólya-Gamma data augmentation details

We use Pólya-Gamma data augmentation following (Polson, Scott, and Windle 2013) to yield an efficient Gibbs sampler for all joint species distribution models in `spOccupancy`. Traditionally, the species-specific regression coefficients (and intercepts) for occurrence ($\boldsymbol{\beta}_i$) and detection ($\boldsymbol{\alpha}_i$) require a Metropolis update, which can lead to slow convergence and bad mixing of MCMC chains (Clark and Altwegg 2019). Instead, we introduce species-specific Pólya-Gamma latent variables for both the occurrence and detection portions of the spatial factor multi-species occupancy model, which induces efficient Gibbs updates for the species-specific occurrence and detection regression coefficients.

Let $\omega_{i,\beta}(\boldsymbol{s}_j)$ for each species i and location j with coordinates $\boldsymbol{s}_j$ follow a Pólya-Gamma distribution with parameters 1 and 0 (i.e., $\omega_{i,\beta}(\boldsymbol{s}_j) \sim \text{PG}(1,0)$). Given this species-specific latent variable, we can re-express the Bernoulli process model (Equation 1 in Doser, Finley, and Banerjee (2023)) as

$$
\begin{aligned}
\psi_i(\boldsymbol{s}_j)^{z_i(\boldsymbol{s}_j)}(1-\psi_i(\boldsymbol{s}_j))^{1-z_i(\boldsymbol{s}_j)} &= \frac{\exp(\boldsymbol{x}_j^\top\boldsymbol{\beta}_i + \text{w}_i^*(\boldsymbol{s}_j))^{z_i(\boldsymbol{s}_j)}}{1+\exp(\boldsymbol{x}_j^\top\boldsymbol{\beta} + \text{w}_i^*(\boldsymbol{s}_j))} \\
&= \exp(\kappa_i(\boldsymbol{s}_j)[\boldsymbol{x}_j^\top\boldsymbol{\beta}_i + \text{w}_i^*(\boldsymbol{s}_j)])\times \\
&\int \exp(-\frac{\omega_{i,\beta}(\boldsymbol{s}_j)}{2}(\boldsymbol{x}_j^\top\boldsymbol{\beta}_i + \text{w}_i^*(\boldsymbol{s}_j))^2)p(\omega_{i,\beta}(\boldsymbol{s}_j) \mid 1,0)d\omega_{i,\beta}(\boldsymbol{s}_j),
\end{aligned}
\tag{1}
$$

where $\kappa_i(\boldsymbol{s}_j) = z_i(\boldsymbol{s}_j) - 0.5$ and $p(\omega_{i,\beta}(\boldsymbol{s}_j))$ is the probability density function of a Pólya-Gamma distribution with parameters 1 and 0 (Polson, Scott, and Windle 2013). Similarly, we define $\omega_{i,k,\alpha}(\boldsymbol{s}_j) \sim \text{PG}(1,0)$ as a latent variable for each site j, each species i, and each replicate k in the detection portion of the occupancy model, which results in an analogous re-expression of the Bernoulli likelihood for $y_{i,k}(\boldsymbol{s}_j)$ as we showed in Equation (1) for $z_i(\boldsymbol{s}_j)$. These re-expressions of the Bernoulli processes result in Gibbs updates for both the occurrence ($\boldsymbol{\beta}_i$) and detection ($\boldsymbol{\alpha}_i$) regression coefficients when they are assigned normal priors [Polson, Scott, and Windle (2013); clark2019].

3 Spatial factor multi-species occupancy model

3.1 Model description

Let $\boldsymbol{s}_j$ denote the spatial coordinates of site j, for all $j = 1, \dots, J$ sites. Define $z_i(\boldsymbol{s}_j)$ as the true latent presence (1) or absence (0) of species i at site j for $i = 1, \dots, N$ species. We assume $z_i(\boldsymbol{s}_j)$ arises from a Bernoulli process following

$$
z_i(\boldsymbol{s}_j) \sim \text{Bernoulli}(\psi_i(\boldsymbol{s}_j)), \tag{2}
$$

where $\psi_i(\boldsymbol{s}_j)$ is the probability of occurrence for species i at site j. We model $\psi_i(\boldsymbol{s}_j)$ according to

$$\text{logit}(\psi_i(\boldsymbol{s}_j)) = \boldsymbol{x}(\boldsymbol{s}_j)^\top \boldsymbol{\beta}_i + \text{w}^*_i(\boldsymbol{s}_j) \tag{3}$$

where $\boldsymbol{x}_j$ is a $p_\psi \times 1$ vector of an intercept and environmental covariates at site j, $\boldsymbol{\beta}_i$ is a $p_\psi \times 1$ species-specific coefficient vector (including an intercept parameter), and $\text{w}^*_i(\boldsymbol{s}_j)$ is a species-specific latent spatial process. We seek to jointly model the species-specific spatial processes to account for residual correlations between species. We use a spatial factor model (Hogan and Tchernis 2004), a dimension reduction approach that can account for correlations among a large number of species. Specifically, we decompose $\text{w}^*_i(\boldsymbol{s}_j)$ into a linear combination of q latent variables (i.e., factors) and their associated species-specific coefficients (i.e., factor loadings). In particular, we have

$$\text{w}^*_i(\boldsymbol{s}_j) = \boldsymbol{\lambda}_i^\top \mathbf{w}(\boldsymbol{s}_j), \tag{4}$$

where $\boldsymbol{\lambda}_i$ is the ith row of factor loadings from an $N \times q$ matrix $\boldsymbol{\Lambda}$, and $\mathbf{w}(\boldsymbol{s}_j)$ is a $q \times 1$ vector of independent spatial factors at site j. We achieve computational improvements and dimension reduction by setting $q << N$. We account for residual species correlations via their individual responses (i.e., loadings) to the q latent spatial factors.

Following Taylor-Rodriguez et al. (2019) and Tikhonov et al. (2020), we model each $r = 1, \dots, q$ independent spatial process $\text{w}_r(\boldsymbol{s}_j)$ using an NNGP (Datta et al. 2016) to achieve computational efficiency when modeling over a large number of spatial locations. More specifically, we have

$$\text{w}_r(\boldsymbol{s}_j) \sim N(\mathbf{0}, \tilde{\boldsymbol{C}}_r(\boldsymbol{\theta}_r)), \tag{5}$$

where $\tilde{\boldsymbol{C}}_r(\boldsymbol{\theta}_r)$ is the NNGP-derived covariance matrix for the r^{th} spatial process. The vector $\boldsymbol{\theta}_r$ consists of parameters governing the spatial process according to a spatial correlation function (Banerjee, Carlin, and Gelfand 2014). For many correlation functions (e.g., exponential, spherical, Gaussian), $\boldsymbol{\theta}_r$ includes a spatial variance parameter, σ^2_r, and a spatial range parameter, ϕ_r, while the Mat'ern correlation function includes an additional spatial smoothness parameter, ν_r.

We assume all species-specific parameters ($\beta_{i,t}$ for all $t = 1, \dots, p_\psi$) arise from community-level distributions (Dorazio and Royle 2005; Gelfand et al. 2005). Specifically, we assign a normal prior with mean and variance hyperparameters that represent the community-level average and variance among species-specific effects across the community, respectively. For example, we model the non-spatial component of the species-specific occurrence intercept, $\beta_{i,1}$, following

$$\beta_{i,1} \sim N(\mu_{\beta_1}, \tau^2_{\beta_1}), \tag{6}$$

where μ_{β_1} is the average intercept across the community, and $\tau^2_{\beta_1}$ is the variability in the species-specific intercepts across the community.

To estimate $\psi_i(\boldsymbol{s}_j)$ and $z_i(\boldsymbol{s}_j)$ while explicitly accounting for imperfect detection, we obtain $k = 1, \dots, K_j$ sampling replicates at each site j. Let $y_{i,k}(\boldsymbol{s}_j)$ denote the detection (1) or nondetection (0) of species i during replicate k at site j. We model the observed data $y_{i,k}(\boldsymbol{s}_j)$ conditional on the true species-specific occurrence $z_i(\boldsymbol{s}_j)$ at site j following

$$\begin{aligned} y_{i,j,k} &\sim \text{Bernoulli}(\pi_{i,j,k} z_{i,j}), \\ \text{logit}(\pi_{i,j,k}) &= \boldsymbol{v}^\top_{i,j,k} \boldsymbol{\alpha}_i, \end{aligned} \tag{7}$$

where $\pi_{i,j,k}$ is the probability of detecting species i at site j during replicate k (given it is present at site j), which is a function of site and replicate-specific covariates $\boldsymbol{V}$ and a vector of species-specific regression coefficients ($\boldsymbol{\alpha}_i$). Similarly to the occurrence regression coefficients, the species-specific detection coefficients are envisioned as random effects arising from a common community-level distribution:

$$\boldsymbol{\alpha}_i \sim \text{Normal}(\boldsymbol{\mu_\alpha}, \boldsymbol{T}_\alpha), \tag{8}$$

where $\boldsymbol{\mu_\alpha}$ is a vector of community-level mean effects for each detection covariate effect (including the intercept) and $\boldsymbol{T}_\alpha$ is a diagonal matrix with diagonal elements $\boldsymbol{\tau}^2_\alpha$ that represent the variability of each detection covariate effect among species in the community.

We assume normal priors for community-level mean parameters and inverse-Gamma priors for community-level variance parameters. Identifiability of the latent spatial factors requires additional constraints (Hogan and Tchernis 2004). Following Taylor-Rodriguez et al. (2019), we set all elements in the upper triangle of the factor loadings matrix $\boldsymbol{\Lambda}$ equal to 0 and its diagonal elements equal to 1. We additionally fix the spatial variance parameters σ^2_r of each latent spatial processes to 1. We assign standard normal priors for all lower triangular elements in $\boldsymbol{\Lambda}$ and assign each spatial range parameter ϕ_r an independent uniform prior.

3.2 Gibbs sampler

Here we describe the Gibbs sampler for fitting the spatial factor multi-species occupancy model using `sfMsPGOcc()`.

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

$$\boldsymbol{\mu_\beta} \mid \cdot \sim N([\boldsymbol{\Sigma}^{-1}_\beta + N\boldsymbol{T}^{-1}_\beta]^{-1}\Big[\sum_{i=1}^{N}(\boldsymbol{T}^{-1}_\beta\boldsymbol{\beta}_i) + \boldsymbol{\Sigma}^{-1}_\beta\boldsymbol{\mu}_{0,\beta}\Big], [\boldsymbol{\Sigma}^{-1}_\beta + N\boldsymbol{T}^{-1}_\beta]^{-1}). \tag{9}$$

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

$$\boldsymbol{\mu_\alpha} \mid \cdot \sim N([\boldsymbol{\Sigma}^{-1}_\alpha + N\boldsymbol{T}^{-1}_\alpha]^{-1}\Big[\sum_{i=1}^{N}(\boldsymbol{T}^{-1}_\alpha\boldsymbol{\alpha}_i) + \boldsymbol{\Sigma}^{-1}_\alpha\boldsymbol{\mu}_{0,\alpha}\Big], [\boldsymbol{\Sigma}^{-1}_\alpha + N\boldsymbol{T}^{-1}_\alpha]^{-1}). \tag{10}$$

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

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

$$\tau^2_{t,\beta} \mid \cdot \sim \text{IG}(a_{\tau_{t,\beta}} + \frac{N}{2}, b_{\tau_{t,\beta}} + \frac{\sum_{i=1}^{N}(\beta_{i,t} - \mu_{\beta_t})^2}{2}). \tag{11}$$

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

Let $\tau^2_{t,\alpha}$ denote the community-level variance for the t^{th} detection parameter ($t = 1, \dots, p_\pi$). We assign an inverse gamma normal prior to $\tau^2_{t,\alpha}$ with shape parameter $a_{\tau_{t,\alpha}}$ and scale parameter $b_{\tau_{t,\alpha}}$. Our full conditional is then

$$\tau^2_{t,\alpha} \mid \cdot \sim \text{IG}(a_{\tau_{t,\alpha}} + \frac{N}{2}, b_{\tau_{t,\alpha}} + \frac{\sum_{i=1}^N (\alpha_{i,t} - \mu_{\alpha_t})^2}{2}). \tag{12}$$

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. Our full conditional is

$$\omega_{i,\beta}(\boldsymbol{s}_j) \mid \cdot \sim \text{PG}(1, \boldsymbol{x}(\boldsymbol{s}_j)^\top \boldsymbol{\beta}_i + \text{w}^*_i(\boldsymbol{s}_j)). \tag{13}$$

3.2.6 Update 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. Following Polson, Scott, and Windle (2013), we have

$$\omega_{i,k,\alpha}(\boldsymbol{s}_j) \mid \cdot \sim \text{PG}(1, \boldsymbol{v}(\boldsymbol{s}_j)^\top \boldsymbol{\alpha}_i). \tag{14}$$

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

$$\boldsymbol{\beta}_i \mid \cdot \sim \text{Normal}\Big([\boldsymbol{T}^{-1}_\beta + \boldsymbol{X}^\top \boldsymbol{S}_\beta \boldsymbol{X}]^{-1}[\boldsymbol{X}^\top(\boldsymbol{z}_i - 0.5\mathbf{1}_J - \boldsymbol{S}_\beta \mathbf{w}^*_i) + \boldsymbol{T}^{-1}_\beta \boldsymbol{\mu}_{\boldsymbol{\beta}}], [\boldsymbol{T}^{-1}_\beta + \boldsymbol{X}^\top \boldsymbol{S}_\beta \boldsymbol{X}]^{-1}\Big), \tag{15}$$

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, and $\mathbf{w}^*_i$ is the $J \times 1$ vector of spatial random effects for species i.

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

Next, we sample the species-specific detection regression coefficients for species i ($\boldsymbol{\alpha}_i$) from

$$\boldsymbol{\alpha}_i \mid \cdot \sim \text{Normal}\Big([\boldsymbol{T}^{-1}_\alpha + \tilde{\boldsymbol{V}}^\top \boldsymbol{S}_\alpha \tilde{\boldsymbol{V}}]^{-1}[\tilde{\boldsymbol{V}}^\top(\tilde{\boldsymbol{y}}_i - 0.5\mathbf{1}_{J^*_i}) + \boldsymbol{T}^{-1}_\alpha \boldsymbol{\mu}_{\boldsymbol{\alpha}}], [\boldsymbol{T}^{-1}_\alpha + \tilde{\boldsymbol{V}}^\top \boldsymbol{S}_\alpha \tilde{\boldsymbol{V}}]^{-1}\Big). \tag{16}$$

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

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

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 $\mathbf{w}_r(N(\boldsymbol{s}_j))$ denote the m realizations of the r^{th} NNGP at the locations in $N(\boldsymbol{s}_j)$. Let $C(\cdot, \phi_r)$ denote the correlation function of the original Gaussian Process (GP) from which the r^{th} NNGP is derived. For any two sets A_1 and A_2, define $\mathrm{C}_{A_1,A_2}(\phi_r)$ as the correlation matrix between the observations in A_1 and A_2 for the r^{th} GP. For $j \geq 1$, we have

$$\mathbf{b}_r(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j, N(\boldsymbol{s}_j)}(\phi_r)\mathbf{C}^{-1}_{N(\boldsymbol{s}_j),N(\boldsymbol{s}_j)}(\phi_r), \tag{17}$$

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

$$f_r(\boldsymbol{s}_j) = \mathbf{C}_{\boldsymbol{s}_j,\boldsymbol{s}_j}(\phi_r) - \mathbf{C}_{\boldsymbol{s}_j,N(\boldsymbol{s}_j)}(\phi_r)\mathbf{C}^{-1}_{N(\boldsymbol{s}_j),N(\boldsymbol{s}_j)}(\phi_r)\mathbf{C}_{N(\boldsymbol{s}_j),\boldsymbol{s}_j}(\phi_r), \tag{18}$$

where $f_r(\boldsymbol{s}_1) = 0$ for all $r = 1, \ldots, q$. For any two locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, if $\boldsymbol{s}_1 \in N(\boldsymbol{s}_2)$ and is the l^{th} member of $N(\boldsymbol{s}_2)$, then define $b_r(\boldsymbol{s}_2, \boldsymbol{s}_1)$ as the l^{th} entry of $\mathbf{b}_r(\boldsymbol{s}_2)$. 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 $a_r(\boldsymbol{s}_2, \boldsymbol{s}_1) = \mathrm{w}_r(\boldsymbol{s}_2) - \sum_{\boldsymbol{s} \in N(\boldsymbol{s}_2), \boldsymbol{s} \neq \boldsymbol{s}_2} \mathrm{w}_r(\boldsymbol{s}) b_r(\boldsymbol{s}_2, \boldsymbol{s})$. Extending this to matrix notation, let $\boldsymbol{B}(\boldsymbol{s}_j)$ be a $q \times mq$ block matrix, with each $q \times q$ diagonal block containing the elements of $\boldsymbol{b}_r(\boldsymbol{s}_j)$ for each of the $r = 1, \ldots q$ spatial factors for each of the specific m neighbors. Let $\boldsymbol{F}(\boldsymbol{s}_j)$ be a $q \times q$ diagonal matrix with diagonal elements of $f_r(\boldsymbol{s}_j)$. Let $\boldsymbol{a}(\boldsymbol{s}, \boldsymbol{s}_j)$ contain the values $a_r(\boldsymbol{s}, \boldsymbol{s}_j)$ for each of the $r = 1, \ldots, q$ latent factors. Using this notation, the full conditional for $\mathbf{w}(\boldsymbol{s}_j)$ is

$$\begin{gathered}\mathbf{w}(\boldsymbol{s}_j) \mid \cdot N_q(\boldsymbol{\mu}_j \boldsymbol{\Sigma}_j, \boldsymbol{\Sigma}_j) \text{ where,} \\ \boldsymbol{\mu}_j = \boldsymbol{F}(\boldsymbol{s}_j)^{-1}\boldsymbol{B}(\boldsymbol{s}_j)\mathbf{w}(N(\boldsymbol{s}_j)) + \textstyle\sum_{\boldsymbol{s} \in U(\boldsymbol{s}_j)} \boldsymbol{B}(\boldsymbol{s}, \boldsymbol{s}_j)^\top \boldsymbol{F}(\boldsymbol{s}_j)^{-1}\boldsymbol{a}(\boldsymbol{s}, \boldsymbol{s}_j) + \\ \boldsymbol{\Lambda}^\top \boldsymbol{S}_{j,\beta}((\boldsymbol{z}(\boldsymbol{s}_j) - 0.5\mathbf{1}_N)\boldsymbol{S}^{-1}_{j,\beta} - \boldsymbol{X}(\boldsymbol{s}_j)^\top \boldsymbol{\beta}) \text{ and} \\ \boldsymbol{\Sigma}_j = \left(\boldsymbol{F}(\boldsymbol{s}_j)^{-1} + \textstyle\sum_{\boldsymbol{s} \in U(\boldsymbol{s}_j)} \boldsymbol{B}(\boldsymbol{s}, \boldsymbol{s}_j)^\top \boldsymbol{F}(\boldsymbol{s}_j)^{-1}\boldsymbol{B}(\boldsymbol{s}, \boldsymbol{s}_j) + \boldsymbol{\Lambda}^\top \boldsymbol{S}_{j,\beta}\boldsymbol{\Lambda}\right)^{-1},\end{gathered} \tag{19}$$

where $\mathbf{w}(N(\boldsymbol{s}_j))$ is a stacked $mq \times 1$ vector of the m realizations of each of the r NNGPs at the locations in $N(\boldsymbol{s}_j)$, $\boldsymbol{S}_{j,\beta}$ is an $N \times N$ diagonal matrix with the Pólya-Gamma auxiliary variables for each species i at site j along the diagonal elements, $\boldsymbol{X}(\boldsymbol{s}_j)^\top$ is a $N \times (Np_\psi)$ block-diagonal matrix with the ith diagonal block the length $\boldsymbol{x}(\boldsymbol{s}_j)$ vector of p_ψ spatially-varying covariates, and $\boldsymbol{\beta}$ is the $(Np_\psi) \times 1$ stacked vector of species-specific regression coefficients (including the intercept).

3.2.10 Update latent spatial factor loadings ($\boldsymbol{\Lambda}$)

Recall we set all diagonal elements of $\boldsymbol{\Lambda}$ 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 $\tilde{\boldsymbol{\lambda}}_i = (\lambda_{i,1}, \ldots, \lambda_{i,q_i})^\top$ be the vector representing the unrestricted elements in the i^{th} row of $\boldsymbol{\Lambda}$. Define $\mathbf{W}$ as the $J \times q$ matrix of latent spatial factors, and let $\mathbf{W}_{1:i}$ be the first i columns of $\mathbf{W}$. Using this notation, the full conditional density for $\tilde{\boldsymbol{\lambda}}_i$ is $N_q(\boldsymbol{\Omega}_{\tilde{\boldsymbol{\lambda}}_i}\boldsymbol{\mu}_{\tilde{\boldsymbol{\lambda}}_i}, \boldsymbol{\Omega}_{\tilde{\boldsymbol{\lambda}}_i})$, where

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{\lambda}}_i} = \begin{cases} \mathbf{W}^\top_{1:(i-1)}\boldsymbol{S}_{i,\beta}(\boldsymbol{S}^{-1}_{i,\beta}(\boldsymbol{z}_i - 0.5\mathbf{1}_J) - \boldsymbol{X}_i^\top \boldsymbol{\beta}_i - \dot{\mathbf{w}}_i) & \text{if} \quad 2 \leq i \leq q \\ \mathbf{W}^\top \boldsymbol{S}_{i,\beta}(\boldsymbol{S}^{-1}_{i,\beta}(\boldsymbol{z}_i - 0.5\mathbf{1}_J) - \boldsymbol{X}_i^\top \boldsymbol{\beta}_i) & \text{if} \quad i > q \end{cases}, \text{ and} \tag{20}$$

$$\boldsymbol{\Omega}_{\tilde{\boldsymbol{\lambda}}_i} = \begin{cases} (\mathbf{W}^\top_{1:(i-1)}\boldsymbol{S}_{i,\beta}\mathbf{W}_{1:(i-1)} + I_{i-1})^{-1} & \text{if} \quad 2 \leq i \leq q \\ (\mathbf{W}^\top \boldsymbol{S}_{i,\beta}\mathbf{W} + I_q)^{-1} & \text{if} \quad i > q \end{cases}, \tag{21}$$

where $\boldsymbol{S}_{i,\beta}$ is a $J \times J$ matrix with diagonal elements consisting of the latent Pólya-Gamma auxiliary variables for species i, $\dot{\mathbf{w}}_i$ is the i^{th} column of $\mathbf{W}$, and $\boldsymbol{X}_i^\top$ is an $N \times p_\psi$ matrix of spatially-varying covariates for species i (which we assume are equivalent for all i species).

3.2.11 Update spatial range parameters ($\boldsymbol{\phi}$)

We use a Metropolis within Gibbs step to sample $\boldsymbol{\phi}$. The full conditional posterior density for ϕ_r for each $r = 1, \dots, q$ is proportional to

$$\begin{gathered} p(\phi_r \mid \cdot) \propto p_r(\phi_r) p(\mathbf{w}_r \mid \phi_r) \\ \propto p(\phi_r) \times \prod_{j=1}^{J} N\left(\mathrm{w}_r(\boldsymbol{s}_j) \mid \mathbf{b}_r(\boldsymbol{s}_j)^\top \mathbf{w}_r(N(\boldsymbol{s}_j)), f_r(\boldsymbol{s}_j).\right) \end{gathered} \tag{22}$$

We sample ϕ_r 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. Thus, for all locations with no detections of the species i, we sample $z_i(\boldsymbol{s}_j)$ according to

$$z_i(\boldsymbol{s}_j) \mid \cdot \sim \text{Bernoulli}\left(\frac{\psi_i(\boldsymbol{s}_j) \prod_{k=1}^{K_j}(1 - \pi_{i,k}(\boldsymbol{s}_j))}{1 - \psi_i(\boldsymbol{s}_j) + \psi_i(\boldsymbol{s}_j) \prod_{k=1}^{K_j}(1 - \pi_{i,k}(\boldsymbol{s}_j))}\right). \tag{23}$$

4 Latent factor multi-species occupancy model

The `spOccupancy` function `lfMsPGOcc()` fits a latent factor multi-species occupancy model. The latent factor multi-species occupancy model is identical to the spatial factor multi-species occupancy model, except we do not assume any spatial structure for the latent factors. Instead, we assign each of the $r = 1, \dots, q$ latent factors a standard normal prior. This model is analogous to the model of (Tobler et al. 2019), except we use a logistic link function and Pólya-Gamma latent variables rather than a probit link function, as well as different restrains on the factor loadings matrix.

4.1 Model description

Let $z_{i,j}$ be the true presence (1) or absence of some species i at site j for a total of $i = 1, \dots, N$ species and $j = 1, \dots, J$ sites. We assume $z_{i,j}$ arises from a Bernoulli process following

$$\begin{aligned} &z_{i,j} \sim \text{Bernoulli}(\psi_{i,j}), \\ &\text{logit}(\psi_{i,j}) = \boldsymbol{x}_j^\top \boldsymbol{\beta}_i + \mathrm{w}^*_{i,j}, \end{aligned} \tag{24}$$

where $\psi_{i,j}$ is the probability of occurrence of species i at site j, which is a function of site-specific covariates $\boldsymbol{x}_j$, a vector of species-specific regression coefficients ($\boldsymbol{\beta}_i$) for those covariates, and a latent process $\mathrm{w}^*_{i,j}$. We incorporate residual species correlations through the formulation of the latent process $\mathrm{w}^*_{i,j}$. We use a factor modeling approach, which is a dimension reduction approach that can account for correlations among a large number of species. Specifically, we decompose $\mathrm{w}^*_{i,j}$ into a linear combination of q latent variables (i.e., factors) and their associated species-specific coefficients (i.e., factor loadings). Thus, we have

$$\mathrm{w}^*_{i,j} = \boldsymbol{\lambda}_i^\top \mathbf{w}_j, \tag{25}$$

where $\boldsymbol{\lambda}_i$ is the ith row of factor loadings from an $N \times q$ matrix $\boldsymbol{\Lambda}$, and $\mathbf{w}_j$ is a $q \times 1$ vector of independent latent factors at site j. We achieve computational improvements by setting $q << N$. We account for residual species correlations via their individual responses (i.e., loadings) to the q latent spatial factors. We can envision the latent variables $\mathbf{w}_j$ as unmeasured site-specific covariates that are treated as random variables in the model estimation procedure. For the non-spatial latent factor model, we assign a standard normal prior distribution to the latent factors (i.e., we assume each latent factor is independent and arises from a normal distribution with mean 0 and standard deviation 1).

We envision the species-specific regression coefficients ($\boldsymbol{\beta}_i$) as random effects arising from a common community-level distribution:

$$\boldsymbol{\beta}_i \sim \text{Normal}(\boldsymbol{\mu_\beta}, \boldsymbol{T}_\beta), \tag{26}$$

where $\boldsymbol{\mu_\beta}$ is a vector of community-level mean effects for each occurrence covariate effect (including the intercept) and $\boldsymbol{T}_\beta$ is a diagonal matrix with diagonal elements $\boldsymbol{\tau}^2_\beta$ that represent the variability of each occurrence covariate effect among species in the community.

We do not directly observe $z_{i,j}$, but rather we observe an imperfect representation of the latent occurrence process. Let $y_{i,j,k}$ be the observed detection (1) or nondetection (0) of a species i of interest at site j during replicate k for each of $k = 1, \dots, K_j$ replicates at each site j. We envision the detection-nondetection data as arising from a Bernoulli process conditional on the true latent occurrence process:

$$\begin{aligned} &y_{i,j,k} \sim \text{Bernoulli}(p_{i,j,k} z_{i,j}), \\ &\text{logit}(p_{i,j,k}) = \boldsymbol{v}^\top_{i,j,k} \boldsymbol{\alpha}_i, \end{aligned} \tag{27}$$

where $p_{i,j,k}$ is the probability of detecting species i at site j during replicate k (given it is present at site j), which is a function of site and replicate-specific covariates $\boldsymbol{V}$ and a vector of species-specific regression coefficients ($\boldsymbol{\alpha}_i$). Similarly to the occurrence regression coefficients, the species-specific detection coefficients are envisioned as random effects arising from a common community-level distribution:

$$\boldsymbol{\alpha}_i \sim \text{Normal}(\boldsymbol{\mu_\alpha}, \boldsymbol{T}_\alpha), \tag{28}$$

where $\boldsymbol{\mu_\alpha}$ is a vector of community-level mean effects for each detection covariate effect (including the intercept) and $\boldsymbol{T}_\alpha$ is a diagonal matrix with diagonal elements $\boldsymbol{\tau}^2_\alpha$ that represent the variability of each detection covariate effect among species in the community.

We assign multivariate normal priors for the community-level occurrence ($\boldsymbol{\mu_\beta}$) and detection ($\boldsymbol{\mu_\alpha}$) means, and assign independent inverse-Gamma priors on the community-level occurrence (τ^2_β) and detection (τ^2_α) variance parameters. To ensure identifiability of the latent factors, we set all elements in the upper triangle of the factor loadings matrix $\boldsymbol{\Lambda}$ equal to 0 and its diagonal elements equal to 1. Analogous to the spatial factor multi-species occupancy model, we introduce Pólya-Gamma auxiliary variables for both the occurrence and detection components of the model to induce a Gibbs update for the species-specific occurrence and detection random effects.

4.2 Gibbs sampler

The Gibbs sampler for the latent factor multi-species occupancy model is identical to the sampler for the spatial factor multi-species occupancy model, with two exceptions: the spatial range parameters are no longer in the model, and the update for the latent factors is different. See Section 3.2 for the Gibbs updates for all parameters besides the latent factors.

4.2.1 Update latent factors ($\mathbf{w}_j$)

Let $\mathbf{w}_j$ denote the q latent factors at site j. Our full conditional is

$$\begin{gathered} \mathbf{w}_j \mid \cdot N_q(\boldsymbol{\mu}_j \boldsymbol{\Sigma}_j, \boldsymbol{\Sigma}_j) \text{ where,} \\ \boldsymbol{\mu}_j = \boldsymbol{\Lambda}^\top \boldsymbol{S}_{j,\beta}((\boldsymbol{z}_j - 0.5\mathbf{1}_N)\boldsymbol{S}^{-1}_{j,\beta} - \boldsymbol{X}^\top_j \boldsymbol{\beta}) \text{ and} \\ \boldsymbol{\Sigma}_j = \left(I_q + \boldsymbol{\Lambda}^\top \boldsymbol{S}_{j,\beta} \boldsymbol{\Lambda}\right)^{-1}, \end{gathered} \tag{29}$$

where $\boldsymbol{S}_{j,\beta}$ is an $N \times N$ diagonal matrix with the Pólya-Gamma auxiliary variables for each species i at site j along the diagonal elements, $\boldsymbol{X}_j^\top$ is a $N \times (Np_\psi)$ block-diagonal matrix with the ith diagonal block the length $\boldsymbol{x}_j$ vector of p_ψ spatially-varying covariates, $\boldsymbol{\beta}$ is the $(Np_\psi) \times 1$ stacked vector of species-specific regression coefficients (including the intercept), and I_q is the $q \times q$ identity matrix.

5 Spatial factor joint species distribution model

The `spOccupancy` function `sfJSDM()` fits a spatial factor joint species distribution model. The spatial factor JSDM (`sfJSDM()`) is a joint species distribution model that ignores imperfect detection but accounts for species residual correlations and spatial autocorrelation. As in the spatial factor multi-species occupancy model, we account for species correlations using a spatial factor model, where the spatial factors arise from q independent NNGPs. This is analogous to the NNGP model presented by Tikhonov et al. (2020), and is similar to other spatially-explicit JSDMs (Thorson et al. 2015; Ovaskainen et al. 2016). Because this model does not account for imperfect detection, we eliminate the detection sub-model and rather directly model a simplified version of the replicated detection-nondetection data, denoted as $y^*_i(\boldsymbol{s}_j)$, where $y^*_i(\boldsymbol{s}_j) = I(\sum_{k=1}^{K_j} y_{i,k}(\boldsymbol{s}_j) > 0)$, with $I(\cdot)$ an indicator function denoting whether or not species i was detected during at least one of the K_j replicates at site j. Note that in the following description, we will describe the covariate effects as effecting the probability of occurrence. However, since we do not explicitly account for imperfect detection, the estimated probability is really a confounded process of occurrence and detection, and thus all covariate effects should be interpreted as combined effects on occurrence and detection.

5.1 Model description

Let $\boldsymbol{s}_j$ denote the spatial coordinates of site j, for all $j = 1, \dots, J$ sites. Define $y^*_i(\boldsymbol{s}_j)$ as the detection (1) or nondetection (0) of species i at site j. We assume $y^*_i(\boldsymbol{s}_j)$ arises from a Bernoulli process following

$$y^*_i(\boldsymbol{s}_j) \sim \text{Bernoulli}(\psi_i(\boldsymbol{s}_j)), \tag{30}$$

where $\psi_i(\boldsymbol{s}_j)$ is the probability of occurrence for species i at site j. We model $\psi_i(\boldsymbol{s}_j)$ according to

$$\text{logit}(\psi_i(\boldsymbol{s}_j)) = \boldsymbol{x}(\boldsymbol{s}_j)^\top \boldsymbol{\beta}_i + \text{w}^*_i(\boldsymbol{s}_j) \tag{31}$$

where $\boldsymbol{x}_j$ is a $p_\psi \times 1$ vector of an intercept and environmental covariates at site j, $\boldsymbol{\beta}_i$ is a $p_\psi \times 1$ species-specific coefficient vector (including an intercept parameter), and $\text{w}^*_i(\boldsymbol{s}_j)$ is a species-specific latent spatial process. Analogous to the spatial factor multi-species occupancy model, we model $\text{w}^*_i(\boldsymbol{s}_j)$ using a spatial facotr modeling approach, where we have

$$\text{w}^*_i(\boldsymbol{s}_j) = \boldsymbol{\lambda}_i^\top \mathbf{w}(\boldsymbol{s}_j), \tag{32}$$

where $\boldsymbol{\lambda}_i$ is the ith row of factor loadings from an $N \times q$ matrix $\boldsymbol{\Lambda}$, and $\mathbf{w}(\boldsymbol{s}_j)$ is a $q \times 1$ vector of independent spatial factors at site j. We achieve computational improvements and dimension reduction by setting $q << N$. We account for residual species correlations via their individual responses (i.e., loadings) to the q latent spatial factors.

We model each $r = 1, \dots, q$ independent spatial process $\text{w}_r(\boldsymbol{s}_j)$ using an NNGP (Datta et al. 2016) to achieve computational efficiency when modeling over a large number of spatial locations. More specifically, we have

$$\text{w}_r(\boldsymbol{s}_j) \sim N(\boldsymbol{0}, \tilde{\boldsymbol{C}}_r(\boldsymbol{\theta}_r)), \tag{33}$$

where $\tilde{\boldsymbol{C}}_r(\boldsymbol{\theta}_r)$ is the NNGP-derived covariance matrix for the r^{th} spatial process. The vector $\boldsymbol{\theta}_r$ consists of parameters governing the spatial process according to a spatial correlation function (Banerjee, Carlin, and Gelfand 2014). For many correlation functions (e.g., exponential, spherical, Gaussian), $\boldsymbol{\theta}_r$ includes a spatial variance parameter, σ^2_r, and a spatial range parameter, ϕ_r, while the Mat'ern correlation function includes an additional spatial smoothness parameter, ν_r.

We assume all species-specific parameters ($\beta_{i,t}$ for all $t = 1, \dots, p_\psi$) arise from community-level distributions (Dorazio and Royle 2005; Gelfand et al. 2005). Specifically, we assign a normal prior with mean and variance hyperparameters that represent the community-level average and variance among species-specific effects across the community, respectively. For example, we model the non-spatial component of the species-specific occurrence intercept, $\beta_{i,1}$, following

$$\beta_{i,1} \sim N(\mu_{\beta_1}, \tau^2_{\beta_1}), \tag{34}$$

where μ_{β_1} is the average intercept across the community, and $\tau^2_{\beta_1}$ is the variability in the species-specific intercepts across the community.

5.2 Gibbs sampler

The Gibbs sampler for the spatial factor joint species distribution model is analogous to the updates for the occurrence parameters in the spatial factor multi-species occupancy model, with all instances of $\boldsymbol{z}_i(\boldsymbol{s}_j)$ replaced by $y^*_i(\boldsymbol{s}_j)$. See Section 3.2.

6 Latent factor joint species distribution model

The `spOccupancy` function `lfJSDM()` fits a latent factor joint species distribution model. The latent factor JSDM (`lfJSDM()`) is a standard joint species distribution model that ignores imperfect detection and spatial autocorrelation but accounts for species residual correlations. As in the latent factor multi-species occupancy model, we account for species correlations using a latent factor model, where the latent factors arise from standard normal distributions. This model is analogous to many varieties of non-spatial JSDMs that leverage a factor modeling approach for dimension reduction (Hui 2016; Ovaskainen et al. 2017). The model is identical to the spatial factor joint species distribution model implemented in `sfJSDM()`, except the latent factors are assumed to arise from standard normal distributions instead of a latent spatial process. This model is analogous to the latent factor multi-species occupancy model, except here we do not account for imperfect detection.

6.1 Model description

Define $y^*_{i,j}$ as the detection (1) or nondetection (0) of species i at site j for $i = 1, \dots, N$ species at $j = 1, \dots, J$ sites. We assume $y^*_{i,j}$ arises from a Bernoulli process following

$$y^*_{i,j} \sim \text{Bernoulli}(\psi_{i,j}), \tag{35}$$

where $\psi_{i,j}$ is the probability of occurrence for species i at site j. We model $\psi_{i,j}$ according to

$$\text{logit}(\psi_{i,j}) = \boldsymbol{x}_j^\top \boldsymbol{\beta}_i + \text{w}^*_{i,j} \tag{36}$$

where $\boldsymbol{x}_j$ is a $p_\psi \times 1$ vector of an intercept and environmental covariates at site j, $\boldsymbol{\beta}_i$ is a $p_\psi \times 1$ species-specific coefficient vector (including an intercept parameter), and $\text{w}^*_{i,j}$ is a species-specific latent process. Analogous to the latent factor multi-species occupancy model, we model $\text{w}^*_{i,j}$ using a factor modeling approach, where we have

$$\text{w}^*_{i,j} = \boldsymbol{\lambda}_i^\top \mathbf{w}_j, \tag{37}$$

where $\boldsymbol{\lambda}_i$ is the ith row of factor loadings from an $N \times q$ matrix $\boldsymbol{\Lambda}$, and $\mathbf{w}_j$ is a $q \times 1$ vector of independent latent factors at site j. We achieve computational improvements and dimension reduction by setting $q << N$.

We account for residual species correlations via their individual responses (i.e., loadings) to the q latent factors. We can envision the latent variables $\mathbf{w}_j$ as unmeasured site-specific covariates that are treated as random variables in the model estimation procedure. Analogous to the latent factor multi-species occupancy model, we assign a standard normal prior distribution to the latent factors (i.e., we assume each latent factor is independent and arises from a normal distribution with mean 0 and standard deviation 1).

We assume all species-specific parameters ($\beta_{i,t}$ for all $t = 1, \dots, p_\psi$) arise from community-level distributions (Dorazio and Royle 2005; Gelfand et al. 2005). Specifically, we assign a normal prior with mean and variance hyperparameters that represent the community-level average and variance among species-specific effects across the community, respectively. For example, we model the non-spatial component of the species-specific occurrence intercept, $\beta_{i,1}$, following

$$\beta_{i,1} \sim N(\mu_{\beta_1}, \tau^2_{\beta_1}), \tag{38}$$

where μ_{β_1} is the average intercept across the community, and $\tau^2_{\beta_1}$ is the variability in the species-specific intercepts across the community.

6.2 Gibbs sampler

The Gibbs sampler for the latent factor joint species distribution model is analogous to the updates for all occurrence parameters in the spatial factor multi-species occupancy model except for the latent factors $\boldsymbol{w}_j$, with all instances of $\boldsymbol{z}_i(\boldsymbol{s}_j)$ replaced by $y^*_i(\boldsymbol{s}_j)$. See Section 3.2. Additionally, the updates for the latent factors are identical to the updates in the latent factor multi-species occupancy model, again with all instances of $\boldsymbol{z}_{i,j}$ replaced by $y^*_i(\boldsymbol{s}_j)$. See Section 4.2.

References

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

Clark, Allan E, and Res Altwegg. 2019. “Efficient Bayesian analysis of occupancy models with logit link functions.” *Ecology and Evolution* 9 (2): 756–68.

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*, e4137.

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.

Hui, Francis KC. 2016. “Boral–Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r.” *Methods in Ecology and Evolution* 7 (6): 744–50.

Ovaskainen, Otso, David B Roy, Richard Fox, and Barbara J Anderson. 2016. “Uncovering Hidden Spatial Structure in Species Communities with Spatially Explicit Joint Species Distribution Models.” *Methods in Ecology and Evolution* 7 (4): 428–36.

Ovaskainen, Otso, Gleb Tikhonov, Anna Norberg, F Guillaume Blanchet, Leo Duan, David Dunson, Tomas Roslin, and Nerea Abrego. 2017. “How to Make More Out of Community Data? A Conceptual Framework and Its Implementation as Models and Software.” *Ecology Letters* 20 (5): 561–76.

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.

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.

Thorson, James T, Mark D Scheuerell, Andrew O Shelton, Kevin E See, Hans J Skaug, and Kasper Kristensen. 2015. "Spatial Factor Analysis: A New Tool for Estimating Joint Species Distributions and Correlations in Species Range." *Methods in Ecology and Evolution* 6 (6): 627–37.

Tikhonov, Gleb, Li Duan, Nerea Abrego, Graeme Newell, Matt White, David Dunson, and Otso Ovaskainen. 2020. "Computationally Efficient Joint Species Distribution Modeling of Big Spatial Data." *Ecology* 101 (2): e02929.

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.