Supporting Information: Alan E. Gelfand and Shinichiro Shirota. 2019. Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. *Ecological Monographs.*

Appendix S4. Model fitting details

Appendix S4.1. Model fitting details for presence/absence data with preferen-tial sampling

Models for $Z(\mathbf{s})$ are discussed in "Preferential sampling" in the main manuscript. Let $\mu(\mathbf{s})$ be the main effect for $Z(\mathbf{s})$, i.e., $Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s})$ where $\epsilon(\mathbf{s}) \sim \mathcal{N}(0, \tau^2)$. The specification for $\mu(\mathbf{s})$ depends on the model, e.g., $\mu(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\alpha} + \delta\eta(\mathbf{s})$ where δ is a preferential parameter for the preferential sampling model (d) in Table 2. We denote $\boldsymbol{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n))^T$, $\boldsymbol{\mu} = (\mu(\mathbf{s}_1), \mu(\mathbf{s}_2), \ldots, \mu(\mathbf{s}_n))^T$ and $\mathbf{X} = (\mathbf{x}(\mathbf{s}_1), \mathbf{x}(\mathbf{s}_2), \ldots, \mathbf{x}(\mathbf{s}_n))^T$.

As for prior specification, we assume $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0})$, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0})$, $\boldsymbol{\omega} \sim \mathcal{N}(\boldsymbol{\omega}_0, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0})$, $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\eta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_0})$, $\delta \sim \mathcal{N}(0, \delta_0)$, $\sigma^2_{\boldsymbol{\omega}} \sim \mathcal{IG}(a_{\boldsymbol{\omega}}, b_{\boldsymbol{\omega}})$, $\sigma^2_{\boldsymbol{\eta}} \sim \mathcal{IG}(a_{\boldsymbol{\eta}}, b_{\boldsymbol{\eta}})$, $\phi_{\boldsymbol{\omega}} \sim \mathcal{U}(l_{\boldsymbol{\omega}}, u_{\boldsymbol{\omega}})$ and $\phi_{\boldsymbol{\eta}} \sim \mathcal{U}(l_{\boldsymbol{\eta}}, u_{\boldsymbol{\eta}})$. The likelihood for $\mathcal{S}$ is

$$\mathcal{L}(\mathcal{S}) \propto \exp\left(-\sum_{i=1}^{I} \lambda(\boldsymbol{u}_i)\Delta_i\right)\prod_{i=1}^{I} \lambda(\boldsymbol{u}_i)^{n_i}, \quad \log\lambda(\boldsymbol{u}_i) = \mathbf{w}^T(\boldsymbol{u}_i)\boldsymbol{\beta} + \eta(\boldsymbol{u}_i) \tag{1}$$

where n_i is the number of $\boldsymbol{s} \in \mathcal{S}$ located within grid i, I is the number of grids and $\sum_{i=1}^{I} n_i = n$. Then, the joint posterior distribution is

$$
\begin{aligned}
p(\boldsymbol{Z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\theta}_{\boldsymbol{\omega}}, \boldsymbol{\theta}_{\boldsymbol{\eta}}, \delta | \mathcal{S}, \mathcal{Y}(\mathcal{S})) \propto\ & \mathcal{L}(\mathcal{S}) \mathcal{N}(\boldsymbol{\alpha} | \boldsymbol{\mu}_{\boldsymbol{\alpha}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0}) \mathcal{N}(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}) \\
& \times \mathcal{N}(\boldsymbol{\omega} | \boldsymbol{\mu}_{\boldsymbol{\omega}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}) \mathcal{N}(\boldsymbol{\eta} | \boldsymbol{\eta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_0}) \mathcal{IG}(\sigma^2_{\boldsymbol{\omega}} | a_{\boldsymbol{\omega}}, b_{\boldsymbol{\omega}}) \mathcal{U}(\phi_{\boldsymbol{\omega}} | l_{\boldsymbol{\omega}}, u_{\boldsymbol{\omega}}) \\
& \times \mathcal{IG}(\sigma^2_{\boldsymbol{\eta}} | a_{\boldsymbol{\eta}}, b_{\boldsymbol{\eta}}) \mathcal{U}(\phi_{\boldsymbol{\eta}} | l_{\boldsymbol{\eta}}, u_{\boldsymbol{\eta}}) \mathcal{N}(\delta | 0, \delta_0) \\
& \times \prod_{i=1}^{n} \mathcal{TN}_{>0}(Z(\mathbf{s}_i) | \mu(\mathbf{s}_i), \tau^2)^{Y(\mathbf{s}_i)} \mathcal{TN}_{\leq 0}(Z(\mathbf{s}_i) | \mu(\mathbf{s}_i), \tau^2)^{1 - Y(\mathbf{s}_i)}
\end{aligned}
\tag{2}
$$

Gibbs sampling for $\boldsymbol{Z}(\cdot)$: Models (a)-(d)

- Sampling $Z(\mathbf{s}_i)|\cdot \sim \mathcal{TN}_{>0}(\mu(\mathbf{s}_i), \tau^2)$ when $Y(\mathbf{s}_i) = 1$ and $Z(\mathbf{s}_i)|\cdot \sim \mathcal{TN}_{\leq 0}(\mu(\mathbf{s}_i), \tau^2)$ otherwise for $i = 1, 2, \ldots, n$ where $\mathcal{TN}_{>0}$ ($\mathcal{TN}_{\leq 0}$) denotes a truncated normal distribution on a positive (nonpositive) domain.

Gibbs sampling for α: Models (a)-(d)

- Sampling $\boldsymbol{\alpha}|\cdot \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$

$$
\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \left(\frac{\mathbf{X}^T (\boldsymbol{Z} - \boldsymbol{\mu}_{(-\mathbf{X}\boldsymbol{\alpha})})}{\tau^2} + \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0} \boldsymbol{\alpha}_0 \right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \left(\frac{\mathbf{X}^T \mathbf{X}}{\tau^2} + \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0} \right)^{-1} \tag{3}
$$

where $\boldsymbol{\mu}_{(-\mathbf{X}\boldsymbol{\alpha})}$ is $\boldsymbol{\mu}$ except for $\mathbf{X}\boldsymbol{\alpha}$.

Metropolis-Hastings update for β: Models (c) and (d)

- MH algorithm for $\boldsymbol{\beta}|\cdot$

$$p(\boldsymbol{\beta}|\cdot) \propto \mathcal{L}(\mathcal{S})\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}) \tag{4}$$

Gibbs sampling for ω: Models (b) and (c)

- Sampling $\boldsymbol{\omega}|\cdot \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\omega}}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$

$$\boldsymbol{\mu}_{\boldsymbol{\omega}} = \boldsymbol{\Sigma}_{\boldsymbol{\omega}} \left(\frac{\boldsymbol{Z} - \boldsymbol{\mu}_{(-\boldsymbol{\omega})}}{\tau^2} + \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}^{-1} \boldsymbol{\omega}_0 \right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\omega}} = \left(\frac{1}{\tau^2} \mathbf{I}_n + \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}^{-1} \right)^{-1} \tag{5}$$

  where $\boldsymbol{\mu}_{(-\boldsymbol{\omega})}$ is $\boldsymbol{\mu}$ except for $\boldsymbol{\omega}$.

Metropolis-Hastings update for η: Models (c) and (d)

- MH algorithm for $\boldsymbol{\eta}|\cdot$

$$p(\boldsymbol{\eta}|\cdot) \propto \mathcal{L}(\mathcal{S})\mathcal{N}(\boldsymbol{Z}|\boldsymbol{\mu}, \tau^2 \mathbf{I}_n)\mathcal{N}(\boldsymbol{\eta}|\boldsymbol{\eta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_0}) \tag{6}$$

Metropolis-Hastings update for $\boldsymbol{\theta}_{\boldsymbol{\omega}} = (\sigma^2_{\boldsymbol{\omega}}, \phi_{\boldsymbol{\omega}})$: Models (b) and (c)

- MH algorithm for $\boldsymbol{\theta}_{\boldsymbol{\omega}}|\cdot$

$$p(\boldsymbol{\theta}_{\boldsymbol{\omega}}|\cdot) \propto \mathcal{N}(\boldsymbol{\omega}|\boldsymbol{\omega}_0, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0})\mathcal{IG}(\sigma^2_{\boldsymbol{\omega}}|a_{\boldsymbol{\omega}}, b_{\boldsymbol{\omega}})\mathcal{U}(\phi_{\boldsymbol{\omega}}|l_{\boldsymbol{\omega}}, u_{\boldsymbol{\omega}}) \tag{7}$$

Metropolis-Hastings update for $\boldsymbol{\theta}_{\boldsymbol{\eta}} = (\sigma^2_{\boldsymbol{\eta}}, \phi_{\boldsymbol{\eta}})$: Models (c) and (d)

- MH algorithm for $\boldsymbol{\theta}_{\boldsymbol{\eta}}|\cdot$

$$p(\boldsymbol{\theta}_{\boldsymbol{\eta}}|\cdot) \propto \mathcal{N}(\boldsymbol{\eta}|\boldsymbol{\eta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_0})\mathcal{IG}(\sigma^2_{\boldsymbol{\eta}}|a_{\boldsymbol{\eta}}, b_{\boldsymbol{\eta}})\mathcal{U}(\phi_{\boldsymbol{\eta}}|l_{\boldsymbol{\eta}}, u_{\boldsymbol{\eta}}) \tag{8}$$

Gibbs sampling for δ: Models (c) and (d)

- Sampling $\delta|\cdot \sim \mathcal{N}(\mu_\delta, \sigma^2_\delta)$

$$\mu_\delta = \sigma^2_\delta \frac{\boldsymbol{\eta}^T(\boldsymbol{Z} - \boldsymbol{\mu}_{(-\delta\boldsymbol{\eta})})}{\tau^2}, \quad \sigma^2_\delta = \frac{1}{\frac{\boldsymbol{\eta}^T\boldsymbol{\eta}}{\tau^2} + \frac{1}{\delta_0}} \tag{9}$$

where $\boldsymbol{\mu}_{(-\delta\boldsymbol{\eta})}$ is $\boldsymbol{\mu}$ except for $\delta\boldsymbol{\eta}$.

For large n, we implement nearest neighbor Gaussian processes (NNGP, **?**). We explain for $\boldsymbol{\omega}$ below, but the same discussion can be applied for $\boldsymbol{\eta}$. The NNGP expresses the joint density of $\boldsymbol{\omega}$ as the product of approximate conditional densities by projecting on to *neighbors* instead of on to the full set of locations, i.e.,

$$\begin{aligned} \pi(\boldsymbol{\omega}) &= \pi(\omega(\mathbf{s}_1))\pi(\omega(\mathbf{s}_2)|\omega(\mathbf{s}_1)) \cdots \pi(\omega(\mathbf{s}_i)|\boldsymbol{\omega}_{<i}) \cdots \pi(\omega(\mathbf{s}_n)|\boldsymbol{\omega}_{<n}) \\ &\approx \pi(\omega(\mathbf{s}_1))\pi(\omega(\mathbf{s}_2)|\omega(\mathbf{s}_1)) \cdots \pi(\omega(\mathbf{s}_i)|\boldsymbol{\omega}_{N_i}) \cdots \pi(\omega(\mathbf{s}_n)|\boldsymbol{\omega}_{N_n}) = \tilde{\pi}(\boldsymbol{\omega}) \end{aligned} \tag{10}$$

where $\boldsymbol{\omega}_{<i} = \{\omega(\mathbf{s}_1), \omega(\mathbf{s}_2), \ldots, \omega(\mathbf{s}_{i-1})\}$ and N_i is the set of indices of neighbors of $\mathbf{s}_i$, $\boldsymbol{\omega}_{N_i} \subseteq \boldsymbol{\omega}_{<i}$. $\tilde{\pi}(\boldsymbol{\omega})$ is a proper multivariate joint density (**?**). As for neighbor selections, choosing N_i to be any subset of $\{1, 2, \ldots, i-1\}$ ensures a valid probability density. Sampling from $\tilde{\pi}(\boldsymbol{\omega})$ is sequentially implemented for $i = 1, \ldots, n$ as follows

$$\begin{aligned} &\omega(\mathbf{s}_i) \sim \mathcal{N}(\boldsymbol{\omega}_0 + \boldsymbol{B}_i \boldsymbol{\omega}(\mathcal{S}_{N_i}), F_i), \\ \text{where} \quad &\boldsymbol{B}_i = \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}(i, N_i)\boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}(N_i, N_i)^{-1}, \quad F_i = \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}(i, i) - \boldsymbol{B}_i \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}(N_i, i) \end{aligned} \tag{11}$$

where $\boldsymbol{\omega}(\mathcal{S}_{N_i}) = (\omega(\mathbf{s}_{N_i(1)}), \ldots, \omega(\mathbf{s}_{N_i(k)}))^\top$ and $N_i(j)$ is j-th component of N_i. Given this expression, Gibbs sampling for $\boldsymbol{\omega}$ is available within the generalized spatial linear model framework (**?**).

Appendix S4.2. Model fitting details for the presence/absence presence-only data fusion

Essentially the same discussion as in Appnedix S4.1 can be applied for the data fusion models (c')-(f) in "Fusing presence/absence and presence-only data: Model fitting and inference for data fusion" in the main manuscript. As for prior specification, we assume $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0})$, $\boldsymbol{\beta}_{PA} \sim \mathcal{N}(\boldsymbol{\beta}_{PA,0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{PA,0}})$, $\boldsymbol{\beta}_{PO} \sim \mathcal{N}(\boldsymbol{\beta}_{PO,0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{PO,0}})$, $\boldsymbol{\omega} \sim \mathcal{N}(\boldsymbol{\omega}_0, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0})$, $\boldsymbol{\eta}_{PA} \sim \mathcal{N}(\boldsymbol{\eta}_{PA,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PA,0}})$, $\boldsymbol{\eta}_{PO} \sim \mathcal{N}(\boldsymbol{\eta}_{PO,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PO,0}})$, $\delta_{PA} \sim \mathcal{N}(0, \delta_{PA,0})$, $\delta_{PO} \sim \mathcal{N}(0, \delta_{PO,0})$, $\sigma^2_{\boldsymbol{\omega}} \sim \mathcal{IG}(a_{\boldsymbol{\omega}}, b_{\boldsymbol{\omega}})$, $\sigma^2_{\boldsymbol{\eta}_{PA}} \sim \mathcal{IG}(a_{\boldsymbol{\eta}_{PA}}, b_{\boldsymbol{\eta}_{PA}})$, $\sigma^2_{\boldsymbol{\eta}_{PO}} \sim \mathcal{IG}(a_{\boldsymbol{\eta}_{PO}}, b_{\boldsymbol{\eta}_{PO}})$, $\phi_{\boldsymbol{\omega}} \sim \mathcal{U}(l_{\boldsymbol{\omega}}, u_{\boldsymbol{\omega}})$, $\phi_{\boldsymbol{\eta}_{PA}} \sim \mathcal{U}(l_{\boldsymbol{\eta}_{PA}}, u_{\boldsymbol{\eta}_{PA}})$ and $\phi_{\boldsymbol{\eta}_{PO}} \sim \mathcal{U}(l_{\boldsymbol{\eta}_{PO}}, u_{\boldsymbol{\eta}_{PO}})$. The likelihoods for $\mathcal{S}_{PA}$ and $\mathcal{S}_{PO}$ are

$$\mathcal{L}(\mathcal{S}_{PA}) \propto \exp\left(-\sum_{i=1}^{I} \lambda_{PA}(\boldsymbol{u}_i)\Delta_i\right) \prod_{i=1}^{I} \lambda_{PA}(\boldsymbol{u}_i)^{n_i}, \quad \log \lambda_{PA}(\boldsymbol{u}_i) = \mathbf{w}^T(\boldsymbol{u}_i)\boldsymbol{\beta}_{PA} + \eta_{PA}(\boldsymbol{u}_i) \tag{12}$$

$$\mathcal{L}(\mathcal{S}_{PO}) \propto \exp\left(-\sum_{i=1}^{I} \lambda_{PO}(\boldsymbol{u}_i)\Delta_i\right) \prod_{i=1}^{I} \lambda_{PO}(\boldsymbol{u}_i)^{m_i}, \quad \log \lambda_{PO}(\boldsymbol{u}_i) = \mathbf{w}^T(\boldsymbol{u}_i)\boldsymbol{\beta}_{PO} + \eta_{PO}(\boldsymbol{u}_i) \tag{13}$$

where $\sum_{i=1}^{I} n_i = n$ and $\sum_{i=1}^{I} m_i = m$.

The joint posterior distribution is

$$
\begin{aligned}
p(\boldsymbol{Z}, \boldsymbol{\alpha}, &\boldsymbol{\beta}_{PA}, \boldsymbol{\beta}_{PO}, \boldsymbol{\omega}, \boldsymbol{\eta}_{PA}, \boldsymbol{\eta}_{PO}, \boldsymbol{\theta}_{\boldsymbol{\omega}}, \boldsymbol{\theta}_{\boldsymbol{\eta}_{PA}}, \boldsymbol{\theta}_{\boldsymbol{\eta}_{PO}}, \delta_{PA}, \delta_{PO} | \mathcal{S}_{PA}, \mathcal{S}_{PO}, \mathcal{Y}(\mathcal{S}_{PA})) \\
&\propto \mathcal{L}(\mathcal{S}_{PA})\mathcal{L}(\mathcal{S}_{PO})\mathcal{N}(\boldsymbol{\alpha}|\boldsymbol{\mu}_{\boldsymbol{\alpha}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0})\mathcal{N}(\boldsymbol{\beta}_{PA}|\boldsymbol{\beta}_{PA,0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{PA,0}})\mathcal{N}(\boldsymbol{\beta}_{PO}|\boldsymbol{\beta}_{PO,0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{PO,0}}) \\
&\times \mathcal{N}(\boldsymbol{\omega}|\boldsymbol{\mu}_{\boldsymbol{\omega}_0}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0})\mathcal{N}(\boldsymbol{\eta}_{PA}|\boldsymbol{\eta}_{PA,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PA,0}})\mathcal{N}(\boldsymbol{\eta}_{PO}|\boldsymbol{\eta}_{PO,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PO,0}}) \\
&\times \mathcal{IG}(\sigma^2_{\boldsymbol{\omega}}|a_{\boldsymbol{\omega}}, b_{\boldsymbol{\omega}})\mathcal{U}(\phi_{\boldsymbol{\omega}}|l_{\boldsymbol{\omega}}, u_{\boldsymbol{\omega}})\mathcal{IG}(\sigma^2_{\boldsymbol{\eta}_{PA}}|a_{\boldsymbol{\eta}_{PA}}, b_{\boldsymbol{\eta}_{PA}})\mathcal{U}(\phi_{\boldsymbol{\eta}_{PA}}|l_{\boldsymbol{\eta}_{PA}}, u_{\boldsymbol{\eta}_{PA}}) \\
&\times \mathcal{IG}(\sigma^2_{\boldsymbol{\eta}_{PO}}|a_{\boldsymbol{\eta}_{PO}}, b_{\boldsymbol{\eta}_{PO}})\mathcal{U}(\phi_{\boldsymbol{\eta}_{PO}}|l_{\boldsymbol{\eta}_{PO}}, u_{\boldsymbol{\eta}_{PO}})\mathcal{N}(\delta_{PA}|0, \delta_{PA,0})\mathcal{N}(\delta_{PO}|0, \delta_{PO,0}) \\
&\times \prod_{i=1}^{n} \mathcal{TN}_{>0}(Z(\mathbf{s}_i)|\mu(\mathbf{s}_i), \tau^2)^{Y(\mathbf{s}_i)}\mathcal{TN}_{\le 0}(Z(\mathbf{s}_i)|\mu(\mathbf{s}_i), \tau^2)^{1-Y(\mathbf{s}_i)}
\end{aligned}
\tag{14}
$$

Gibbs sampling for $Z(\cdot)$: Models (c')-(f)

- Sampling $Z(\mathbf{s}_i)|\cdot \sim \mathcal{TN}_{>0}(\mu(\mathbf{s}_i), \tau^2)$ when $Y(\mathbf{s}_i) = 1$ and $Z(\mathbf{s}_i)|\cdot \sim \mathcal{TN}_{\le 0}(\mu(\mathbf{s}_i), \tau^2)$ otherwise for $i = 1, 2, \ldots, n$ where $\mathcal{TN}_{>0}$ ($\mathcal{TN}_{\le 0}$) is truncated normal distribution on positive (nonpositive) domain.

Gibbs sampling for α: Models (c')-(f)

- Sampling $\boldsymbol{\alpha}|\cdot \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\left(\frac{\mathbf{X}^T(\boldsymbol{Z} - \boldsymbol{\mu}_{(-\mathrm{x}\boldsymbol{\alpha})})}{\tau^2} + \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0}^{-1}\boldsymbol{\alpha}_0\right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \left(\frac{\mathbf{X}^T\mathbf{X}}{\tau^2} + \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0}^{-1}\right)^{-1} \tag{15}$$

where $\boldsymbol{\mu}_{(-\mathrm{x}\boldsymbol{\alpha})}$ is $\boldsymbol{\mu}$ except for $\mathbf{X}\boldsymbol{\alpha}$.

Metropolis-Hastings update for $\boldsymbol{\beta}_{PA}$: Models (e) and (f)

- MH algorithm for $\boldsymbol{\beta}_{PA}|\cdot$

$$p(\boldsymbol{\beta}_{PA}|\cdot) \propto \mathcal{L}(\mathcal{S}_{PA})\mathcal{N}(\boldsymbol{\beta}_{PA}|\boldsymbol{\beta}_{PA,0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{PA,0}}) \tag{16}$$

Metropolis-Hastings update for $\boldsymbol{\beta}_{PO}$: Models (c')-(f)

- MH algorithm for $\boldsymbol{\beta}_{PO}|\cdot$

$$p(\boldsymbol{\beta}_{PO}|\cdot) \propto \mathcal{L}(\mathcal{S}_{PO})\mathcal{N}(\boldsymbol{\beta}_{PO}|\boldsymbol{\beta}_{PO,0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{PO,0}}) \tag{17}$$

Gibbs sampling for ω: Models (c') and (f)

- Sampling $\boldsymbol{\omega}|\cdot \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\omega}}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}})$

$$\boldsymbol{\mu}_{\boldsymbol{\omega}} = \boldsymbol{\Sigma}_{\boldsymbol{\omega}} \left(\frac{\boldsymbol{Z} - \boldsymbol{\mu}_{(-\boldsymbol{\omega})}}{\tau^2} + \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}^{-1} \boldsymbol{\omega}_0 \right), \quad \boldsymbol{\Sigma}_{\boldsymbol{\omega}} = \left(\frac{1}{\tau^2} \mathbf{I}_n + \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0}^{-1} \right)^{-1} \tag{18}$$

where $\boldsymbol{\mu}_{(-\boldsymbol{\omega})}$ is $\boldsymbol{\mu}$ except for $\boldsymbol{\omega}$.

Metropolis-Hastings update for $\boldsymbol{\eta}_{PA}$: Models (e) and (f)

- MH algorithm for $\boldsymbol{\eta}_{PA}|\cdot$

$$p(\boldsymbol{\eta}_{PA}|\cdot) \propto \mathcal{L}(\mathcal{S}_{PA}) \mathcal{N}(\boldsymbol{Z}|\boldsymbol{\mu}, \tau^2 \mathbf{I}_n) \mathcal{N}(\boldsymbol{\eta}_{PA}|\boldsymbol{\eta}_{PA,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PA,0}}) \tag{19}$$

Metropolis-Hastings update for $\boldsymbol{\eta}_{PO}$: Models (c')-(f)

- MH algorithm for $\boldsymbol{\eta}_{PO}|\cdot$

$$p(\boldsymbol{\eta}_{PO}|\cdot) \propto \mathcal{L}(\mathcal{S}_{PO}) \mathcal{N}(\boldsymbol{Z}|\boldsymbol{\mu}, \tau^2 \mathbf{I}_n) \mathcal{N}(\boldsymbol{\eta}_{PO}|\boldsymbol{\eta}_{PO,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PO,0}}) \tag{20}$$

Metropolis-Hastings update for $\boldsymbol{\theta}_{\boldsymbol{\omega}} = (\sigma^2_{\boldsymbol{\omega}}, \phi_{\boldsymbol{\omega}})$: Models (b) and (c')

- MH algorithm for $\boldsymbol{\theta}_{\boldsymbol{\omega}}|\cdot$

$$p(\boldsymbol{\theta}_{\boldsymbol{\omega}}|\cdot) \propto \mathcal{N}(\boldsymbol{\omega}|\boldsymbol{\omega}_0, \boldsymbol{\Sigma}_{\boldsymbol{\omega}_0})\mathcal{IG}(\sigma^2_{\boldsymbol{\omega}}|a_{\boldsymbol{\omega}}, b_{\boldsymbol{\omega}})\mathcal{U}(\phi_{\boldsymbol{\omega}}|l_{\boldsymbol{\omega}}, u_{\boldsymbol{\omega}}) \tag{21}$$

Metropolis-Hastings update for $\boldsymbol{\theta}_{\boldsymbol{\eta}_{PA}} = (\sigma^2_{\boldsymbol{\eta}_{PA}}, \phi_{\boldsymbol{\eta}_{PA}})$: Models (e) and (f)

- MH algorithm for $\boldsymbol{\theta}_{\boldsymbol{\eta}_{PA}}|\cdot$

$$p(\boldsymbol{\theta}_{\boldsymbol{\eta}_{PA}}|\cdot) \propto \mathcal{N}(\boldsymbol{\eta}_{PA}|\boldsymbol{\eta}_{PA,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PA,0}})\mathcal{IG}(\sigma^2_{\boldsymbol{\eta}_{PA}}|a_{\boldsymbol{\eta}_{PA}}, b_{\boldsymbol{\eta}_{PA}})\mathcal{U}(\phi_{\boldsymbol{\eta}_{PA}}|l_{\boldsymbol{\eta}_{PA}}, u_{\boldsymbol{\eta}_{PA}}) \tag{22}$$

Metropolis-Hastings update for $\boldsymbol{\theta}_{\boldsymbol{\eta}_{PO}} = (\sigma^2_{\boldsymbol{\eta}_{PO}}, \phi_{\boldsymbol{\eta}_{PO}})$: Models (c')-(f)

- MH algorithm for $\boldsymbol{\theta}_{\boldsymbol{\eta}_{PO}}|\cdot$

$$p(\boldsymbol{\theta}_{\boldsymbol{\eta}_{PO}}|\cdot) \propto \mathcal{N}(\boldsymbol{\eta}_{PO}|\boldsymbol{\eta}_{PO,0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}_{PO,0}})\mathcal{IG}(\sigma^2_{\boldsymbol{\eta}_{PO}}|a_{\boldsymbol{\eta}_{PO}}, b_{\boldsymbol{\eta}_{PO}})\mathcal{U}(\phi_{\boldsymbol{\eta}_{PO}}|l_{\boldsymbol{\eta}_{PO}}, u_{\boldsymbol{\eta}_{PO}}) \tag{23}$$

Gibbs sampling for δ_{PA}: Models (e) and (f)

- Sampling $\delta_{PA}|\cdot \sim \mathcal{N}(\mu_{\delta_{PA}}, \sigma^2_{\delta_{PA}})$

$$\mu_{\delta_{PA}} = \sigma^2_{\delta_{PA}} \frac{\boldsymbol{\eta}_{PA}^T(\boldsymbol{Z} - \boldsymbol{\mu}_{(-\delta_{PA}\boldsymbol{\eta}_{PA})})}{\tau^2}, \quad \sigma^2_{\delta_{PA}} = \frac{1}{\frac{\boldsymbol{\eta}_{PA}^T\boldsymbol{\eta}_{PA}}{\tau^2} + \frac{1}{\delta_{PA,0}}} \tag{24}$$

where $\boldsymbol{\mu}_{(-\delta_{PA}\boldsymbol{\eta}_{PA})}$ is $\boldsymbol{\mu}$ except for $\delta_{PA}\boldsymbol{\eta}_{PA}$.

Gibbs sampling for δ_{PO}: Models (c')-(f)

- Sampling $\delta_{PO}|\cdot \sim \mathcal{N}(\mu_{\delta_{PO}}, \sigma^2_{\delta_{PO}})$

$$\mu_{\delta_{PO}} = \sigma^2_{\delta_{PO}} \frac{\boldsymbol{\eta}_{PO}^T(\boldsymbol{Z} - \boldsymbol{\mu}_{(-\delta_{PO}\boldsymbol{\eta}_{PO})})}{\tau^2}, \quad \sigma^2_{\delta_{PO}} = \frac{1}{\frac{\boldsymbol{\eta}_{PO}^T\boldsymbol{\eta}_{PO}}{\tau^2} + \frac{1}{\delta_{PO,0}}} \tag{25}$$

where $\boldsymbol{\mu}_{(-\delta_{PO}\boldsymbol{\eta}_{PO})}$ is $\boldsymbol{\mu}$ except for $\delta_{PO}\boldsymbol{\eta}_{PO}$.

As discussed in Appendix S4.1, NNGP completion is available for $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}_{PA,0}}$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}_{PO,0}}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\omega}_0}$.

Appendix S4.3. Model fitting with partially observed presence-only data

Under the gridding in "Preferential sampling" in the main manuscript, suppose we have n_i presence locations $(\mathbf{s}_{i,1}, \mathbf{s}_{i,2}, ..., \mathbf{s}_{i,n_i})$ within grid cell i for $i = 1, 2, \ldots, I$ where I is the number of grids. Following the discussion above, $U(\mathbf{s}_{i,j})T(\mathbf{s}_{i,j}) \equiv 1$, $0 \leq j \leq n_i, 1 \leq i \leq I$. Then, the corresponding likelihood function becomes

$$L(\mathcal{S}) \propto \exp\left(-\int_D \lambda(\mathbf{s})U(\mathbf{s})T(\mathbf{s})\, d\mathbf{s}\right)\prod_{i=1}^{I}\prod_{j=1}^{n_i}\lambda(\mathbf{s}_{i,j}) \quad \text{with} \tag{26}$$

$$\log\, \lambda(\mathbf{s}) = \mathbf{w}^T(\mathbf{s})\boldsymbol{\beta} + \omega(\mathbf{s}) \tag{27}$$

and $\omega(\mathbf{s})$, a zero-mean stationary, isotropic Gaussian process (GP) over D. Although we have only finitely many presence locations, the integral in $L(\mathcal{S})$ involves the uncountable random field $\lambda_D = \{\lambda(\mathbf{s}) \,:\, \mathbf{s} \in D\}$. Fortunately, we have a natural approximation to the stochastic integral at the scale of grid cells. That is, though we have geo-coded locations for the observed sites, with covariate information at grid cell level, we only attempt to explain the point pattern at grid cell level. Also, with many unsampled cells, many $n_i = 0$.

A computational advantage accrues to working at grid cell level; we can employ a product Poisson likelihood approximation rather than the point pattern likelihood in (**??**). That is, for cell i, suppose $\mathbf{s}_i$ is the centroid. Then, given the set $\{\lambda(\mathbf{s}_i), i = 1, 2, ..., I$, the n_i are independent and $n_i \sim \text{Po}(\Delta\lambda(\mathbf{s}_i)q_i)$ where Δ is the area of cell i. Approximation of the point pattern likelihood using such a *tiled* surface over a lattice embedding the region was discussed in **?**. There it is shown that the approximation can be justified in the sense that the resulting approximate posterior converges to the true posterior as the partition gets finer.

Notice that, for any cell with $q_i = 0$ (which can happen if either $p_i = 0$ or $u_i = 0$) there is no contribution from A_i in the product Poisson likelihood. Let $\mathbf{W} = (\mathbf{w}(\mathbf{s}_1), \mathbf{w}(\mathbf{s}_2), \ldots, \mathbf{w}(\mathbf{s}_m))^T$ Since, from (**??**), log$\lambda(\mathbf{s})$ follows a GP, the posterior distribution takes the form

$$\begin{aligned} p(\boldsymbol{\lambda}(\mathbf{s}_{1:m}), \boldsymbol{\beta}, \boldsymbol{\theta} | \mathcal{S}) \quad \propto \quad & \exp\Big(-\sum_{i=1}^{I} \lambda(\mathbf{s}_i)\Delta_i q_i\Big)\prod_{i=1}^{m}\lambda^{n_i}(\mathbf{s}_i) \\ & \times\ \mathcal{N}_m(\log \boldsymbol{\lambda}(\mathbf{s}_{1:m}) | \mathbf{W}\boldsymbol{\beta}, \boldsymbol{\Sigma}_\omega(\boldsymbol{\theta}))\pi(\boldsymbol{\beta})\pi(\boldsymbol{\theta}) \end{aligned} \tag{28}$$

where $\mathcal{N}_m(\cdot|\cdot)$ denotes an $m\ (< I)$ dimensional Gaussian density with mean $\mathbf{W}\boldsymbol{\beta}$ and covariance $\boldsymbol{\Sigma}_\omega(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$ denotes the parameters in the covariance function of $\omega(\mathbf{s})$ in (**??**).

With regard to displays of intensity surfaces, the $\lambda_i p_i$ surface will capture the (lack of) sampling effort. The $\lambda_i u_i$ surface reveals the effect of transformation. The $\lambda(\mathbf{s})$ surface is most interesting since it offers insight into the expected pattern of presences over all of D. Posterior draws of $\lambda_{1:I}$ can be used to infer about the potential intensity, displaying say the posterior mean surface. We can also learn about the potential density $g(\mathbf{s})$ in this discretized setting as $g_i = \lambda_i / \sum_{k=1}^{I} \lambda_k$, and the corresponding density under transformation as $g_{u,i} = \lambda_i u_i / \sum_{k=1}^{I} \lambda_k u_k$.

References

Benes, V., K. Bodlák, J. Møller, and R. P. Waagepetersen (2002). Bayesian analysis of log Gaussian Cox processes for disease mapping. Technical report, Department of Mathematical Sciences, Aalborg University.

Datta, A., S. Banerjee, A. O. Finley, and A. E. Gelfand (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association 111*, 800–812.