Bayesian graphical inference in spatial joint species distribution models

Rochebrune 2026

H. Rozier B. Alglave **Pierre Gloaguen** F.Septier

March 23, 2026

Université Bretagne Sud, Vannes, France

Context: Joint species distribution models

Multi-species abundance data

Ecological context

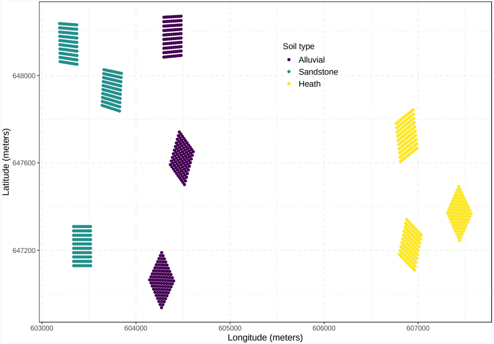
- **Goal** Describe and understand distribution of species within an ecosystem;
- **Mean** Counting individuals of all encountered species through surveys.

Statistical challenges

- Multivariate count data;
- Modelling dependencies:
  - Coming from spatial sampling;
  - Coming from species interactions.

Forest communities in Borneo

Figure 1: Données issues de Sellan et al. (2019).



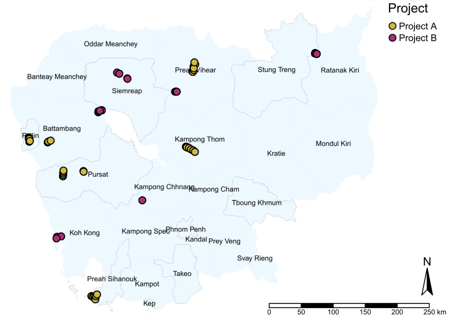
- 900 forest plots (20m × 20m) in Malaysia (Borneo island);
- All trees are counted are counted on all plots: 600 encountered species;

Ecological question

Do variations in plots communities is explained by the soil composition?

Mosquito communities in Cambodia

Figure 2: Source: Rozier et al. (2025).



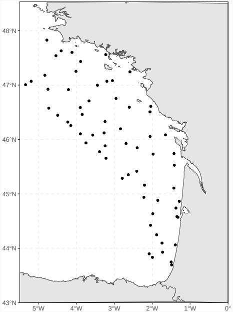
- 76 sites trapping sites across Cambodia
- Survey over time, in resulting 2000 collections;
- On each collection, mosquitoes from 150 species are counted;

Ecological question

What are the spatial and/or temporal patterns of mosquito communities?

Invertebrate communities in the Bay of Biscay

Figure 3: Données issues de Outrequin et al. (2025).

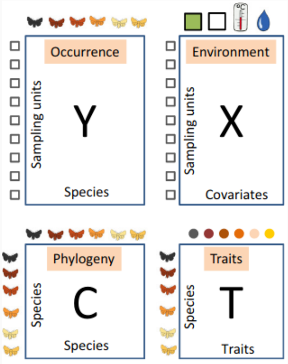


- 64 sites trawling sites across Bay of Biscay (EVOHE Project, 2019);
- Sampling made over different substrats, at different depth;
- 256 invertebrate species encountered.

> What are the spatial patterns of species distributions. Are there some communities?

Joint species distribution data

Figure 4: Source: Ovaskainen and Abrego (2020)



- $\mathbf{Y}$ a $n \times p$ matrix with entries in $\mathbb{N}$;
- $\mathbf{X}$ a $n \times n_{\text{cov}}$ matrix of real values;
- $\mathbf{C}$ a $p \times p$ correlation matrix;
- $\mathbf{T}$ a $p \times n_{\text{trait}}$ matrix of real values;

A multivariate generalized linear mixed model approach

- Latent variable approach: modelling of joint distribution of $(\mathbf{Y}, \mathbf{Z})$.

  For $1 \leq i \leq n, 1 \leq j \leq p,\ Y_{i,j} | Z_{i,j} \overset{\text{ind}}{\sim} \mathcal{D}(\exp(Z_{i,j}), \phi)$. Observation model

- $\mathcal{D}$ is some distribution (Negative binomial, Poisson) with:
  - Expectation $\exp(Z_{i,j})$;
  - Dispersion parameter ϕ.

$$\mathbf{Z} \sim \mathcal{MN}(\mathbf{X}\boldsymbol{\beta}, \Sigma_{\text{Site}}, \Sigma_{\text{Species}}) \text{ . Latent variable}$$

- $\mathcal{MN}(\mathbf{X}\boldsymbol{\beta}, \Sigma_{\text{Site}}, \Sigma_{\text{Species}})$ is the *matrix normal* distribution with:
  - Expectation $\mathbf{X}\boldsymbol{\beta}$ (a $n \times p$ matrix);
  - Covariance between rows Σ_{Site} (a $n \times n$ symmetric positive definite matrix);
  - Covariance between columns Σ_{Species} (a $p \times p$ SPD matrix);
- $\boldsymbol{\beta}$ is an **unknown** $n_{\text{cov}} \times p$ matrix: **response of species to environment**.

The matrix normal distribution

Definition

$\mathbf{M}$: $n \times p$ matrix, $\mathbf{U}$ (resp. $\mathbf{V}$): $n \times n$ (resp. $p \times p$) SPD matrix:

$$\mathbf{Z} \sim \mathcal{MN}(\mathbf{M}, \mathbf{U}, \mathbf{V}) \Leftrightarrow \mathrm{vec}(\mathbf{Z}) \sim \mathcal{N}_{np \times np}(\mathrm{vec}(\mathbf{M}), \mathbf{V} \otimes \mathbf{U}) \,.$$

Properties

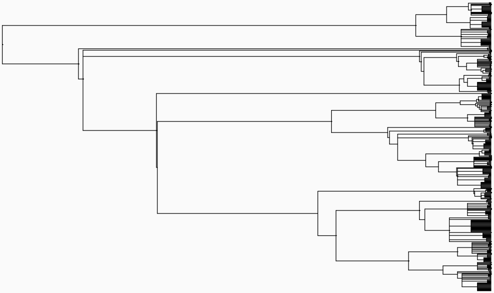
- Separability hypothesis over covariances;
- If $\mathbf{Z} \sim \mathcal{MN}(M, \mathbf{U}, \mathbf{V})$, then $\mathbf{A}Z\mathbf{B}^\mathsf{T} + \mathbf{C} \sim \mathcal{MN}\left(\mathbf{AMB}^\mathsf{T} + \mathbf{C}, \mathbf{AUA}^\mathsf{T}, \mathbf{BVB}^\mathsf{T}\right)$;
  - Useful both for simulation and whitening;

Structuring the niches

- Suppose we have access to other data about species:
- Species traits in a matrix $\mathbf{T}$:

Species	Growth rate	Wood density	Max. Height
Strychnos borneensis	0.008	0.750	19.749
Dysoxylum indet	0.027	0.585	8.588
Memecylon indet	0.013	0.783	8.692
Cratoxylum cochinchinense	0.025	0.670	9.894
Sterculia stipulata	0.027	0.365	10.087

- Phylogeny, giving a correlation matrix $\mathbf{C}$:



Modelling the fixed effects

Structuring the β matrix

- The matrix $\boldsymbol{\beta}$ stacks vector of responses to environment:
  - The environmental niches of species;
- Assume that:
  - The **traits** affect $\boldsymbol{\beta}$ (similar traits lead to similar niche);
  - Columns (species) of $\boldsymbol{\beta}$ are correlated because of **phylogeny**.

Formally, $\boldsymbol{\beta}$ is assumed to be a Matrix Normal random variable such that:

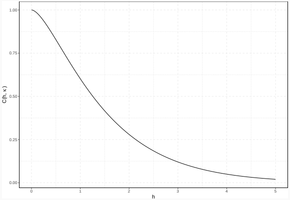
$$\boldsymbol{\beta} \sim \mathcal{MN}\left(\boldsymbol{\Gamma}\mathbf{T}^\top, \eta^2\mathbf{I}_{n_{\mathsf{cov}}}, \rho\mathbf{C} + (1-\rho)\mathbf{I}_p\right) .$$

- $\boldsymbol{\Gamma}$ is a $n_{\mathsf{cov}} \times n_{\mathsf{trait}}$ matrix structuring $\boldsymbol{\beta}$:
  - **Do the species niches are correlated to species traits?**
- $\mathbf{C}$ is the correlation matrix induced by the **phylogeny**;
- $0 \leq \rho \leq 1$ is the weight of **phylogeny** in the columns correlation of $\boldsymbol{\beta}$.

Modelling residual dependencies

Modelling the dependences between sites

Figure 5: Matern covariance function for $\nu = 1, \sigma^2 = 1, \kappa = 1$



Model on spatial dependences (rowwise dependences)

- For two sites i and ℓ separated by a distance $h_{i,\ell}$, $(\Sigma_{\text{Site}})_{i,\ell} = C(h_{i,\ell}, \kappa)$ with:

$$C(h, \kappa) = \kappa h K_1(\kappa h)\,,$$

  with $K_1(\cdot)$ the modified Bessel function of second kind.
- Matern covariance with:
  - Fixed regularity $\nu = 1$ and marginal variance $\sigma^2 = 1$;
  - Unknown range parameter κ to estimate.

Modelling the dependences between species

Model on species dependences (colwise dependences)

- No "natural" structure induced by a "natural" distance;
- Need for a geometrical/statistical structure to avoid $p \times (p+1)/2$ free parameters;

Low rank structures on covariances

- Reducing number of parameters by assuming a low-rank structure

$$\Sigma_{\text{Species}} = \Lambda\Lambda^{\top},$$

  where Λ is a $p \times q$ matrix with $q \ll p$;
- Probabilistic PCA structure;
- Λ interpreted as species response to q unmeasured covariates;

Modelling the dependences between species (II)

Model on species dependences (colwise dependences)

- No "natural" structure induced by a "natural" distance;
- Need for a geometrical/statistical structure to avoid $p \times (p+1)/2$ free parameters;

Model on species dependences (colwise dependences)

- Modelling the precision matrix $\Omega = \Sigma^{-1}_{\text{Species}}$;
- Conditional independence property: For $1 \leq j, k \leq p$:

$$\Omega_{j,k} = 0 \Leftrightarrow \left(\mathbf{Z}_{\cdot,j} | \mathbf{Z}_{\cdot,-(j,k)}\right) \perp \left(\mathbf{Z}_{\cdot,k} | \mathbf{Z}_{\cdot,-(j,k)}\right) .$$

- Induces a graph between species: Draw an edge between species j and k iff $\omega_{j,k} = 0$;
- *Interpretation:* Ecological interaction network;

Bayesian Model considered

$$Y_{i,j}|Z_{i,j} \overset{\text{ind}}{\sim} \mathcal{NB}(\exp(Z_{i,j}), \phi)\,, \qquad \text{Observation model}$$

$$\mathbf{Z} \sim \mathcal{MN}\left(\mathbf{X}\boldsymbol{\beta}, \Sigma(\kappa), \Omega^{-1}\right). \qquad \text{Latent variable}$$

Priors

$$\boldsymbol{\beta} \sim \mathcal{MN}(0, \mathbf{I}_{n_{\text{cov}}}, \mathbf{I}_p)$$

$$\phi \sim \log\mathcal{N}\left(m, s^2\right)$$

$$\kappa \sim \mathcal{G}\text{amma}(a, b)$$

$$\Omega \sim \text{Continuous spike and slab prior (CSS)}$$

- CSS comes from Wang (2015).

Bayesian priors for precision matrices

A frequentist reference: the graphical Lasso

Learning Ω

- We want to learn an interpretable graph;
- This suggest to have a sparse Ω;
- Natural approach: Penalizing matrices having few zeros.

- Let rows of $\mathbf{Z}$ be centered and independent *i.e.*

$$\mathbf{Z} \sim \mathcal{MN}\left(0, \mathbf{I}_n, \Omega^{-1}\right).$$

- The graphical Lasso learns Ω by maximizing:

$$\log\left(\det(\Omega)\right) - \mathsf{Tr}\left(\frac{1}{n}\mathbf{Z}^\intercal \mathbf{Z}\Omega\right) - \lambda\|\Omega\|_1, \;\; \|\Omega\|_1 = \sum_{1 \leq i,j \leq p} |\omega_{i,j}|.$$

- The λ penalty parameter is chosen using model selection criterion.
- Implemented for JSDM (Chiquet, Mariadassou, and Robin (2021)).

Bayesian graphical lasso

- Prior distribution[1] over Ω;

$$p(\Omega|\lambda) \propto \prod_{i<j} \mathcal{L}\mathsf{ap}(w_{i,j}|\lambda) \prod_{i=1}^{p} \mathcal{E}(w_{i,i}|\lambda) \mathbf{1}_{\mathbb{S}_+}(\Omega)\,.$$

- Leads to a Bayesian posterior proportional to graphical Lasso objective;
- Suitable for efficient Gibbs sampling (Wang (2012)).
- Does not lead to a straightforward rule about what is a "true" 0;

[1] $\mathcal{L}\mathsf{ap}(x|\lambda)$ and $\mathcal{E}(x|\lambda)$ are the p.d.f. of centered Laplace and exponential distributions with parameter λ, $\mathbb{S}_+$ is the set of SPD matrices.

Continuous spike and slab prior for precision matrix (I)

- Idea: Modelling[2] a 0 and non-0 regime;

$$
\begin{aligned}
p(\Omega|\theta) = & \mathsf{C}(\theta)^{-1} \times \\
& \left(\prod_{i<j} \left(1-\pi\right)\mathcal{N}\left(w_{i,j}|0, v_0^2\right) + \pi\mathcal{N}\left(w_{i,j}|0, v_1^2\right)\right) \\
& \times \prod_{i=1}^{p} \mathcal{E}(w_{i,i}|\lambda) \Bigg) \times \mathbf{1}_{\mathbb{S}_+}(\Omega)\,,
\end{aligned}
$$

where:

- π can be interpreted as a *a priori* probability of an edge to exist;
- v_0^2 and v_0^2 are *a priori* variances of $\omega_{i,j}$ under the 2 regimes ($v_0^2 \ll v_1^2$);
- $\theta = \{\pi, v_0^2, v_1^2, \lambda\}$.

[2] C(·) is a normalizing constant.

Continuous spike and slab prior for precision matrix (II)

- Hierarchical version: Introduction of a latent binary matrix $\mathbf{G}$:

$$p(\Omega|\theta, \mathbf{G}) = \mathsf{C}(\mathbf{G}, \theta)^{-1} \times \left(\prod_{i<j} \mathcal{N}\left(w_{i,j}|0, v^2_{\mathbf{G}_{i,j}}\right) \times \prod_{i=1}^{p} \mathcal{E}(w_{i,i}|\lambda) \right) \times \mathbf{1}_{\mathbb{S}_+}(\Omega)\,,$$

$$p(\mathbf{G}|\theta) \propto \mathsf{C}(\mathbf{G}, \theta) \times \prod_{i\leq j} \pi^{\mathbf{G}_{i,j}} (1-\pi)^{\mathbf{G}_{i,j}}$$

- $\mathsf{C}(G, \theta)$ is an intractable normalizing term that cancels out in the joint distribution of $(\Omega, \mathbf{G})$;
- “The latent binary variables $\mathbf{G}$ can be viewed as edge-inclusion indicators” (Wang (2015)).

Other bayesian priors for sparse precision matrices

- Lots of methods:
  - Horseshoe prior (Li, Craig, and Bhadra (2019));
  - G-Wishart (Atay-Kayis and Massam (2005));
  - Sparsifying from the Cholesky decomposition (Mastrantonio, Di Loro, and Mingione (2025));
- Nice review with empirical comparisons (Vogels et al. (2024));
- Main challenges: interpretable prior with scalable posterior sampling.

Posterior sampling

Full model and target posterior distribution

Model's equations

$$
\begin{aligned}
Y_{i,j}|Z_{i,j} &\overset{\text{ind}}{\sim} \mathcal{NB}(\exp(Z_{i,j}), \phi)\,, \\
\mathbf{Z} &\sim \mathcal{MN}\left(\mathbf{X}\boldsymbol{\beta}, \Sigma(\kappa), \Omega^{-1}\right), \\
\boldsymbol{\beta} &\sim \mathcal{MN}(0, \mathbf{I}_{n_{\text{cov}}}, \mathbf{I}_p)\,, \\
\phi &\sim \log\mathcal{N}\left(m, s^2\right), \\
\kappa &\sim \mathcal{G}\text{amma}(a, b)\,, \\
p(\Omega|\theta, \mathbf{G}) &= \mathsf{C}(\mathbf{G}, \theta)^{-1}\left(\prod_{i<j}\mathcal{N}\left(w_{i,j}|0, v^2_{\mathbf{G}_{i,j}}\right) \times \prod_{i=1}^{p}\mathcal{E}(w_{i,i}|\lambda)\right)\mathbf{1}_{\mathbb{S}_+}(\Omega)\,, \\
p(\mathbf{G}|\theta) &\propto \mathsf{C}(\mathbf{G}, \theta) \times \prod_{i\leq j}\pi^{\mathbf{G}_{i,j}}(1-\pi)^{\mathbf{G}_{i,j}}
\end{aligned}
$$

$p(\mathbf{Z}, \boldsymbol{\beta}, \phi, \kappa, \Omega, \mathbf{G}|\mathbf{Y})$ Target posterior distribution

Gibbs approach

Table 2: Steps of the Gibbs algorithm

Conditional variable	**Sampling Method**
$\mathbf{Z}\vert\mathbf{Y}, \boldsymbol{\beta}, \kappa, \Omega$	MALA (rowwise)
$\boldsymbol{\beta}\vert\mathbf{Z}, \kappa, \Omega$	Exact (gaussian)
$\phi\vert\mathbf{Y}, \mathbf{Z}$	MALA
$\Omega\vert\mathbf{Z}, \mathbf{G}, \boldsymbol{\beta}, \kappa$	Exact
$\mathbf{G}\vert\Omega$	Exact
$\kappa\vert\mathbf{Z}, \boldsymbol{\beta}, \Omega$	MALA

Gibbs sampling of $\Omega | \mathbf{Z}, \mathbf{G}, \boldsymbol{\beta}, \kappa$

- *Row whitening:* Compute $\tilde{\mathbf{Z}} = \L_\kappa^{-1}(\mathbf{Z} - \boldsymbol{\beta})$, where $\mathbf{L}_\kappa \mathbf{L}_\kappa^\intercal = \Sigma(\kappa)$;
- Consider $\mathbf{S} = \tilde{\mathbf{Z}}^\intercal \tilde{\mathbf{Z}}$;
- Sample new columns of Ω iteratively:
- To sample from the (say) last column conditionally to the $(p-1)$ other, partition:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{-p,-p} & \vec{s}_{-p,p} \\ \vec{s}^\intercal_{-p,p} & s_{p,p} \end{pmatrix}$$

- Set:

$$\Omega_{\text{new}} = \begin{pmatrix} \Omega_{-p,-p} & \vec{u} \\ \vec{u}^\intercal & v + \vec{u}^\intercal \Omega_{-p,-p} \vec{u} \end{pmatrix},$$

  where

$$v \sim \mathcal{G}\text{amma}\left(\frac{n}{2} + 1, \frac{s_{p,p} + \lambda}{2}\right),$$
$$\vec{u} \sim \mathcal{N}_{p-1}(-\mathbf{C}\vec{s}_{-p,p}, \mathbf{C}),$$

  where $\mathbf{C}$ is a matrix depending on $\mathbf{G}$ (easy to compute).
- Starting from an SPD matrix, this gives an SPD matrix.

Sampling of $\mathrm{G}|\Omega$

- Using the proposed hierarchical prior, conditional of $\mathbf{G}|\Omega$ sampling is straightforward.
- Sample independently for $1 \leq i < j \leq p$:

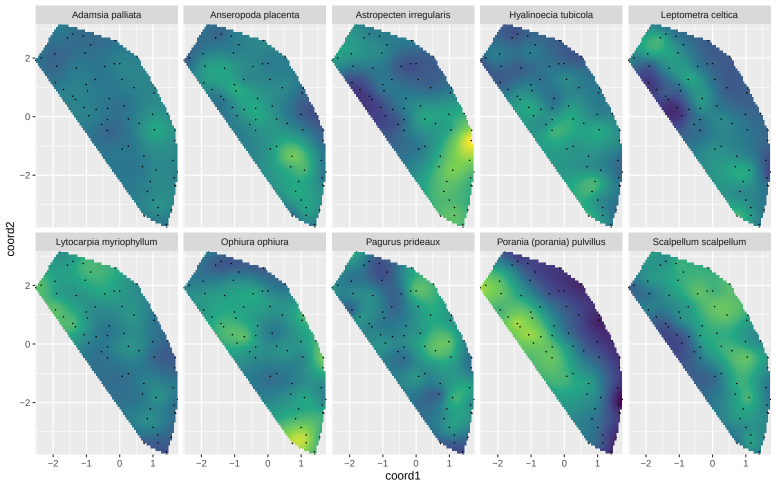
$$\mathbb{P}(\mathbf{G}_{i,j} = 1|\Omega) = \frac{\mathcal{N}\left(\omega_{i,j}|0, v_1^2\right)\pi}{\mathcal{N}(\omega_{i,j}|0, v_1^2)\pi + \mathcal{N}(\omega_{i,j}|0, v_0^2)(1-\pi)}\,.$$

Application to Bay of Biscay invertebrates

Estimation of the spatial field per species

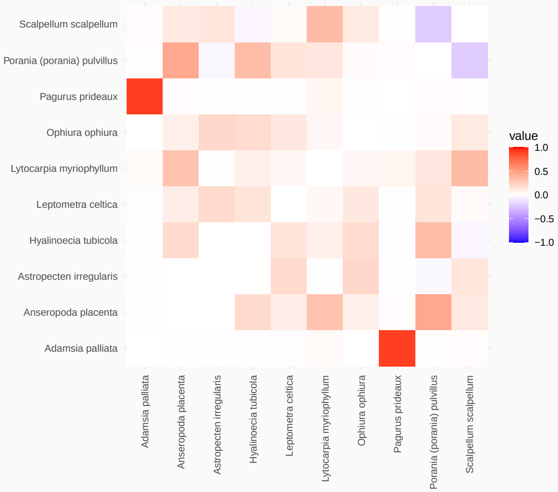
- Inference on 10 species;

Figure 6: Estimated spatial field for 10 species (posterior mean)



Estimation of the interaction network (partial correlation)

Figure 7: Estimated partial correlation matrix between species (posterior mean)



Estimation of $\mathbf{G}$

Figure 8: Probability of $\mathbf{G}_{i,j} = 1$ depending threshold

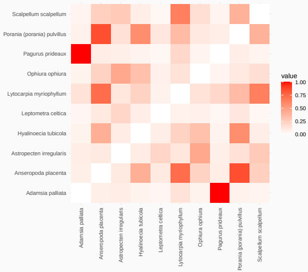


Figure 9: Number of edges

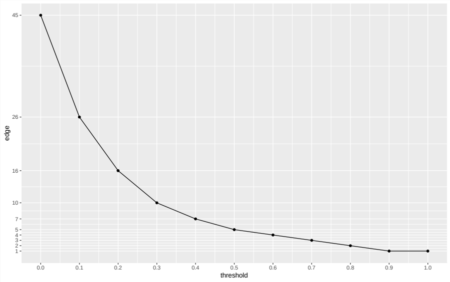


Figure 10: *Pagurus prideaux* and *Adamsia palliata*

Conclusions

- Exact Bayesian inference for network inference for spatial data;
- Estimation of two sources of dependence on a separable context;
- Works on n around hundreds, and p around dozens;

Perspectives

- Assess ecological pertinence on on various different ecological data (trees, mosquitoes, invertebrate, fishes);
- Proposing some priors based on ecological assumptions;

Atay-Kayis, Aliye, and Hélene Massam. 2005. "A Monte Carlo Method for Computing the Marginal Likelihood in Nondecomposable Gaussian Graphical Models." *Biometrika* 92 (2): 317–35.

Chiquet, Julien, Mahendra Mariadassou, and Stéphane Robin. 2021. "The Poisson-Lognormal Model as a Versatile Framework for the Joint Analysis of Species Abundances." *Frontiers in Ecology and Evolution* 9: 588292.

Li, Yunfan, Bruce A Craig, and Anindya Bhadra. 2019. "The Graphical Horseshoe Estimator for Inverse Covariance Matrices." *Journal of Computational and Graphical Statistics* 28 (3): 747–57.

Mastrantonio, Gianluca, Pierfrancesco Alaimo Di Loro, and Marco Mingione. 2025. "A New Hierarchical Distribution on Arbitrary Sparse Precision Matrices." *arXiv Preprint arXiv:2506.09607*.

Outrequin, Thomas, Hervé Le Bris, Jacques Grall, and Pascal Laffargue. 2025. "Mesoscale Assemblages of Fish and Megainvertebrates as Evidence of Benthiscapes on Continental Shelves." *Journal of Marine Systems* 250: 104088.

Ovaskainen, Otso, and Nerea Abrego. 2020. *Joint Species Distribution Modelling: With Applications in r*. Cambridge University Press.

Rozier, Heloise, Pierre Gloaguen, François Septier, and Sébastien Boyer. 2025. "Evaluating Experimental Design to Sample Mosquitoes." *Scientific Reports* 15 (1): 43552.

Sellan, Giacomo, Jill Thompson, Noreen Majalap, and Francis Q Brearley. 2019. "Soil Characteristics Influence Species Composition and Forest Structure Differentially Among Tree Size Classes in a Bornean Heath Forest." *Plant and Soil* 438 (1): 173–85.

Vogels, Lucas, Reza Mohammadi, Marit Schoonhoven, and Ș İlker Birbil. 2024. "Bayesian Structure Learning in Undirected Gaussian Graphical Models: Literature Review with Empirical Comparison." *Journal of the American Statistical Association* 119 (548): 3164–82.

Wang, Hao. 2012. "Bayesian Graphical Lasso Models and Efficient Posterior Computation."

———. 2015. "Scaling It up: Stochastic Search Structure Learning in Graphical Models." *Bayesian Analysis* 7 (2): 351–77.