Hierarchical spatial modelling for applied population and community ecology

Jeffrey W. Doser, Marc Kéry,
Gesa von Hirschheydt

24-27 June 2024

Multi-species occupancy models (MSOMs)

Jeffrey W. Doser

24-27 June 2024

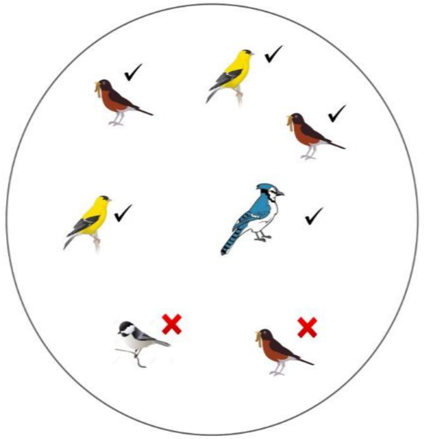
Multi-species detection-nondetection data

eBird iNaturalist

Citizen Science

Acoustic recording units

Camera traps



Point count surveys

Multi-species data from a single visit

Species	Site 1	Site 2	Site 3	Site 4
A	1	0	0	1
B	0	0	1	0
C	1	1	0	0
D	1	0	0	0
E	0	1	1	1
F	0	0	0	1

Multi-species data from multiple visits

Visit 3

Species	Site 1	Site 2	Site 3	Site 4
A	1	0	0	NA
B	0	1	1	NA
C	0	0	0	NA
D	0	0	0	NA
E	0	0	1	NA
F	0	0	0	NA

Visit 2

Species	Site 1	Site 2		
A	0	NA		
B	0	NA		
C	1	NA		
D	0	NA		
E	0	NA	1	1
F	0	NA	0	0

Visit 1

Species	Site 1	Site 2		
A	1	0		
B	0	0		
C	1	1		
D	1	0	0	0
E	0	1	1	1
F	0	0	0	1

Multi-species data from multiple visits

Visit 3

Species	Site 1	Site 2	Site 3	Site 4
A	1	0	0	NA
B	0	1	1	NA
C	0	0	0	NA
D	0	0	0	NA
E	0	0	1	NA
F	0	0	0	NA

Visit 2

Species	Site 1	Site 2		
A	0	NA		
B	0	NA		
C	1	NA		
D	0	NA		
E	0	NA	1	1
F	0	NA	0	0

Visit 1

Species	Site 1	Site 2		
A	1	0		
B	0	0		
C	1	1		
D	1	0	0	0
E	0	1	1	1
F	0	0	0	1

Remember this for when we work with three-dimensional arrays.

Statistical Motivation

- Species of interest (e.g., Species of Greatest Conservation Need) are often the rarest species.

Statistical Motivation

- Species of interest (e.g., Species of Greatest Conservation Need) are often the rarest species.
- Occupancy models are hard to fit when the number of detections is low

Statistical Motivation

- Species of interest (e.g., Species of Greatest Conservation Need) are often the rarest species.
- Occupancy models are hard to fit when the number of detections is low
- Multi-species models can:
  - Improve ability to model rare species

Statistical Motivation

- Species of interest (e.g., Species of Greatest Conservation Need) are often the rarest species.
- Occupancy models are hard to fit when the number of detections is low
- Multi-species models can:
  - Improve ability to model rare species
  - Provide inference at both species and community-levels

Statistical Motivation

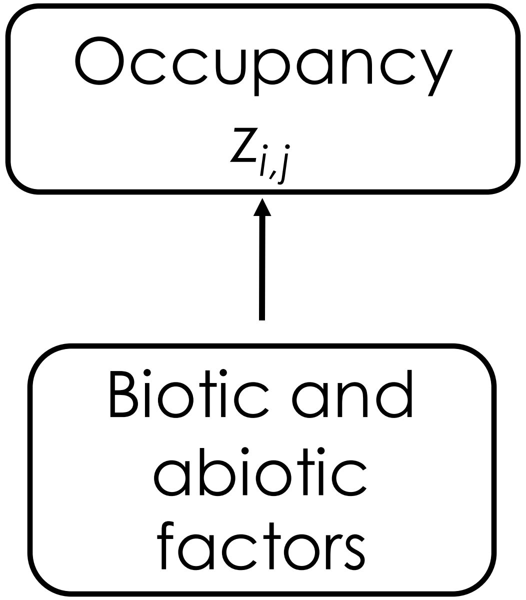
- Species of interest (e.g., Species of Greatest Conservation Need) are often the rarest species.
- Occupancy models are hard to fit when the number of detections is low
- Multi-species models can:
  - Improve ability to model rare species
  - Provide inference at both species and community-levels
  - Use information from other species to improve species-specific estimates

Ecological Motivation

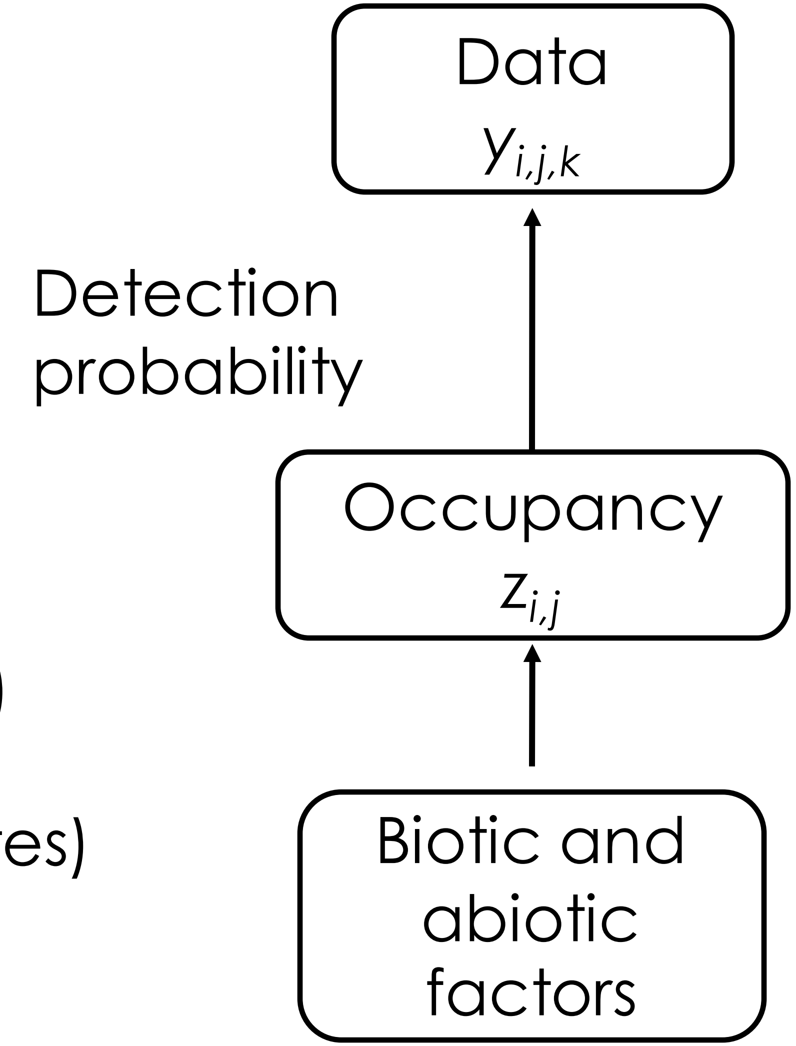
- Management has historically focused on individual species.
- Increased interest in multi-species management.
- Biodiversity conservation.
- Species are not independent of each other.

Multi-species occupancy model

i = 1, …, N (species)
j = 1, …, J (sites)
k = 1, …, K (replicates)

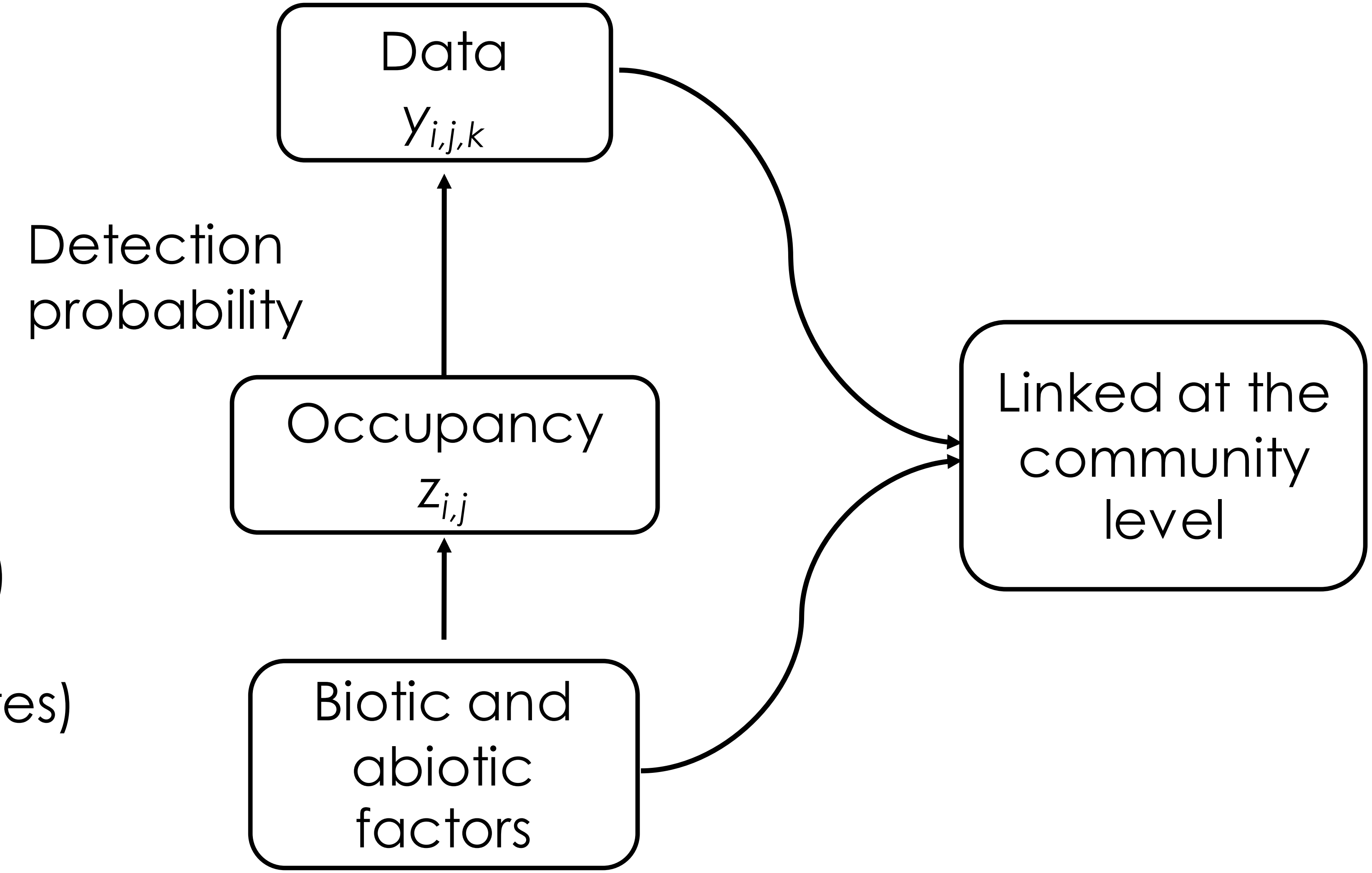


Multi-species occupancy model



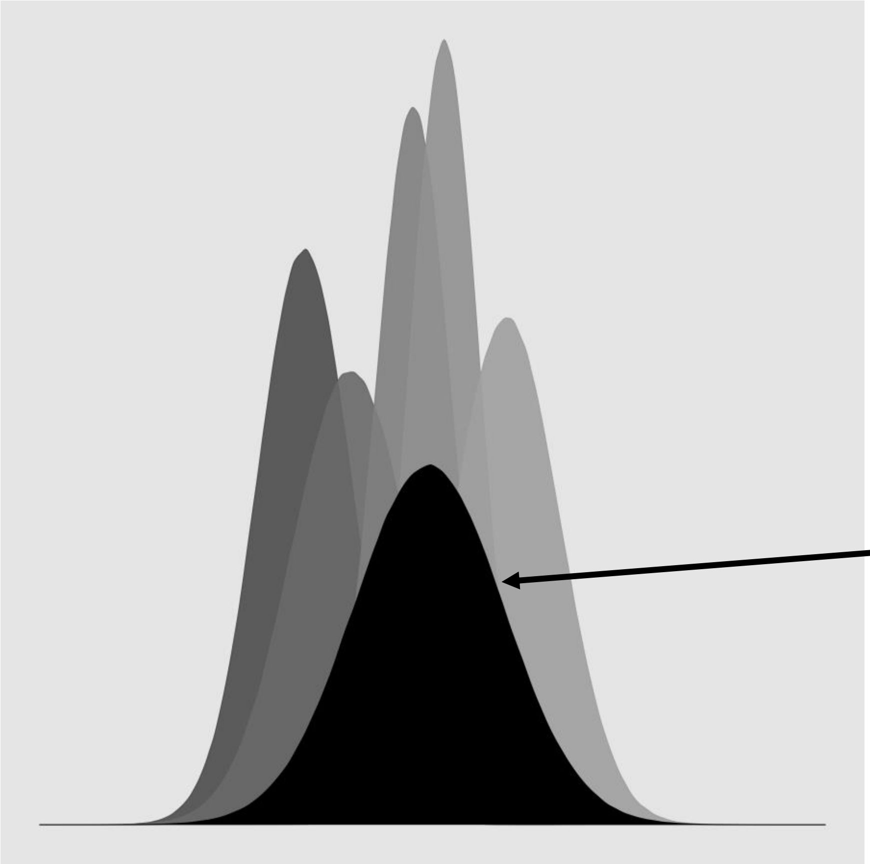
i = 1, …, N (species)
j = 1, …, J (sites)
k = 1, …, K (replicates)

Multi-species occupancy model



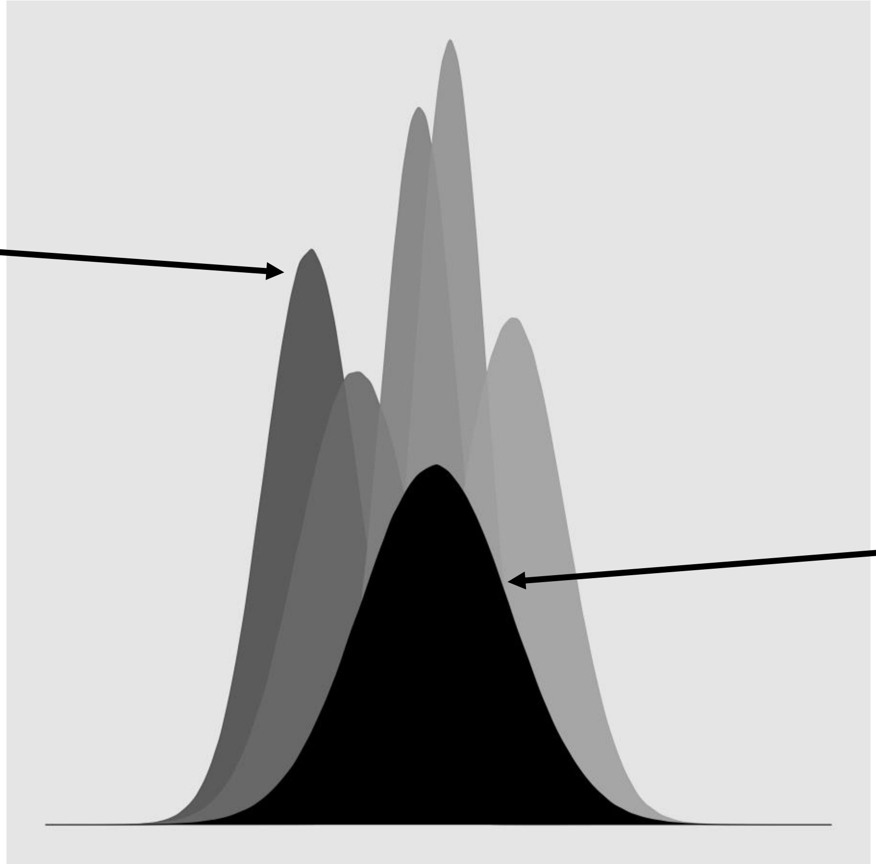
i = 1, …, N (species)
j = 1, …, J (sites)
k = 1, …, K (replicates)

Species-specific and community effects



Community-level
effect

Species-specific and community effects

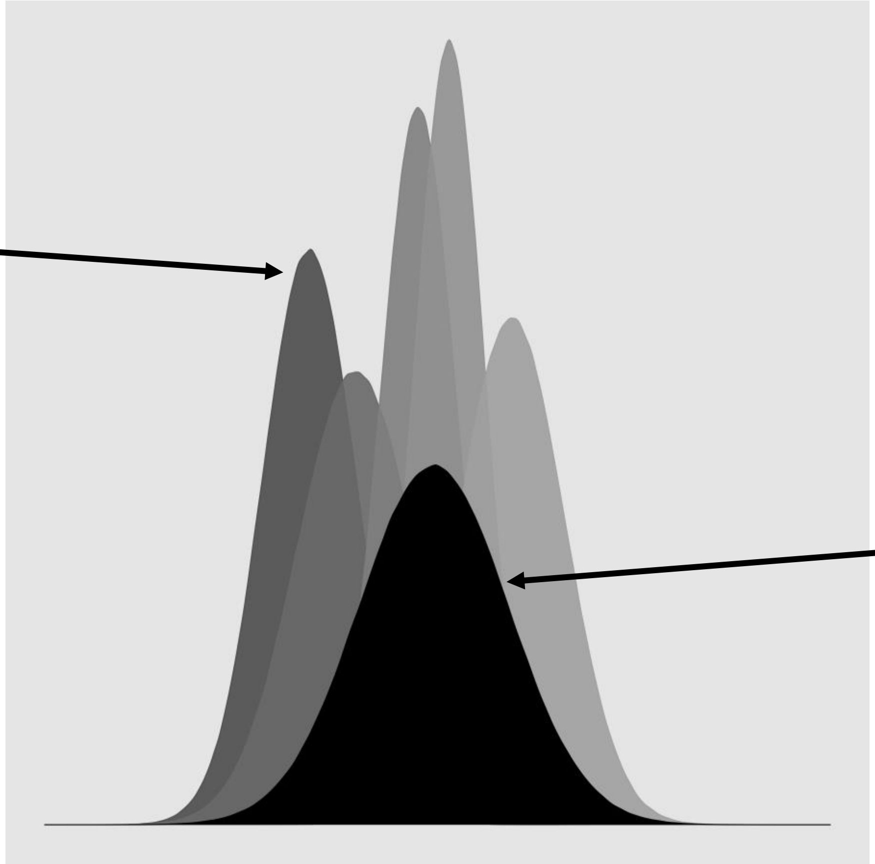


Species-specific effects

Community-level effect

Species-specific and community effects

Species-specific
effects



Community-level
effect

Species-specific effects are drawn from a
common, community-level distribution

Development of the MSOM

JOURNAL ARTICLE

Modelling Species Diversity Through Species Level Hierarchical Modelling FREE

Alan E. Gelfand, Alexandra M. Schmidt, Shanshan Wu, John A. Silander, Jr, Andrew Latimer, Anthony G. Rebelo

Journal of the Royal Statistical Society Series C: Applied Statistics, Volume 54, Issue 1, January 2005, Pages 1–20, https://doi.org/10.1111/j.1467-9876.2005.00466.x

Published: 22 October 2004 **Article history** ▾

Primary Article

Estimating Size and Composition of Biological Communities by Modeling the Occurrence of Species

Robert M Dorazio & J. Andrew Royle

Pages 389-398 | Published online: 01 Jan 2012

Cite this article https://doi.org/10.1198/016214505000000015

Multi-species occupancy model

i = 1, …, N (species)
j = 1, …, J (sites)
k = 1, …, K (replicates)

Occupancy (ecological) sub-model

$$z_{i,j} \sim \text{Bernoulli}(\psi_{i,j})$$
$$\text{logit}(\psi_{i,j}) = \beta_{1,i} + \beta_{2,i} \cdot X_{2,j} + \cdots + \beta_{r,i} \cdot X_{r,j}$$
$$\beta_{r,i} \sim \text{Normal}(\mu_{\beta_r}, \tau^2_{\beta,r})$$

Detection (observation) sub-model

$$y_{i,j,k} \sim \text{Bernoulli}(p_{i,j,k} \cdot z_{i,j})$$
$$\text{logit}(p_{i,j,k}) = \alpha_{1,i} + \alpha_{2,i} \cdot V_{2,j,k} + \cdots + \alpha_{r,i} \cdot V_{r,j,k}$$
$$\alpha_{r,i} \sim \text{Normal}(\mu_{\alpha_r}, \tau^2_{\alpha,r})$$

Multi-species occupancy model

i = 1, …, N (species)
j = 1, …, J (sites)
k = 1, …, K (replicates)

Occupancy (ecological) sub-model

$$z_{i,j} \sim \text{Bernoulli}(\psi_{i,j})$$
$$\text{logit}(\psi_{i,j}) = \beta_{1,i} + \beta_{2,i} \cdot X_{2,j} + \cdots + \beta_{r,i} \cdot X_{r,j}$$
$$\boxed{\beta_{r,i} \sim \text{Normal}(\mu_{\beta_r}, \tau^2_{\beta,r})}$$

Detection (observation) sub-model

$$y_{i,j,k} \sim \text{Bernoulli}(p_{i,j,k} \cdot z_{i,j})$$
$$\text{logit}(p_{i,j,k}) = \alpha_{1,i} + \alpha_{2,i} \cdot V_{2,j,k} + \cdots + \alpha_{r,i} \cdot V_{r,j,k}$$
$$\alpha_{r,i} \sim \text{Normal}(\mu_{\alpha_r}, \tau^2_{\alpha,r})$$

Multi-species occupancy model

$$\beta_{r,i} \sim \mathrm{Normal}(\mu_{\beta_r}, \tau^2_{\beta_r})$$

μ_{β_r} Mean effect of covariate across all species

$\tau^2_{\beta_r}$ Variance of the covariate effect among all species

Multi-species occupancy model

$$\beta_{r,i} \sim \text{Normal}(\mu_{\beta_r}, \tau^2_{\beta_r})$$

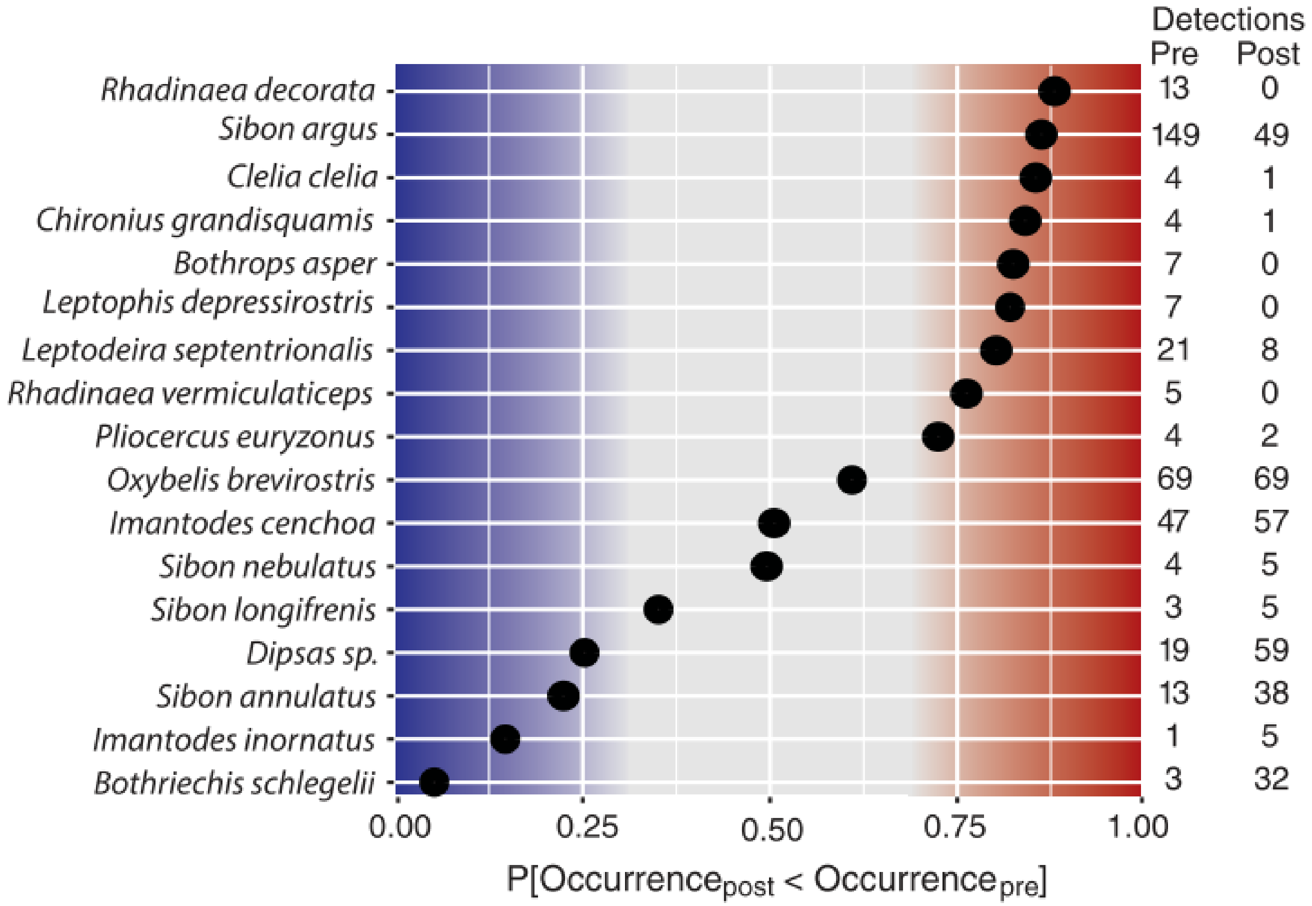
μ_{β_r} Mean effect of covariate across all species

$\tau^2_{\beta_r}$ Variance of the covariate effect among all species

These are just random slopes and intercepts!!

Why multi-species occupancy modeling?

Improved ability to model rare species



Zipkin et al (2020)

Deriving community-level metrics

- Recall the Bayesian approach allows us to generate estimates of quantities (with uncertainty) that aren't directly parameters in the model. These are called *Derived Quantities*.
- In an *MSOM*, we can generate a variety of biodiversity metrics (species richness, composition, diversity) as derived quantities
- Usually involves manipulating the species-specific occupancy estimates at each site z_{ij} in different ways

Potential downsides compared to single-species models

- Longer model run times

Potential downsides compared to single-species models

- Longer model run times
- Coding often involves working with multi-dimensional arrays (but `spOccupancy` simplifies this!)

Potential downsides compared to single-species models

- Longer model run times
- Coding often involves working with multi-dimensional arrays (but `spOccupancy` simplifies this!)
- Defining a "community" is not always straightforward:
  - Pacifici et al. 2014 Ecology and Evolution

Potential downsides compared to single-species models

- Longer model run times
- Coding often involves working with multi-dimensional arrays (but `spOccupancy` simplifies this!)
- Defining a "community" is not always straightforward:
  - Pacifici et al. 2014 Ecology and Evolution
- May not be ideal for the "rarest of the rare" species:
  - Erickson and Smith, 2023 Ecography

Fitting the MSOM in `spOccupancy`

- `msPGOcc()` (multi-species Pólya-Gamma occupancy model)
- Same exact arguments as `PGOcc()`
- Detection-nondetection data are now supplied as a three-dimensional array

Fitting the MSOM in `spOccupancy`

- `msPGOcc()` (multi-species Pólya-Gamma occupancy model)
- Same exact arguments as `PGOcc()`
- Detection-nondetection data are now supplied as a three-dimensional array

See [vignette here] for example of formatting data for multi-species occupancy model

Extending the MSOM

Extending the MSOM

- As of now, our MSOM only shares information across species via the random intercepts/slopes.

Extending the MSOM

- As of now, our MSOM only shares information across species via the random intercepts/slopes.
- Could the co-occurrence patterns across different species provide us with improved ecological insights?

Extending the MSOM

- As of now, our MSOM only shares information across species via the random intercepts/slopes.
- Could the co-occurrence patterns across different species provide us with improved ecological insights?
- Can we extend the MSOM to account for residual species correlations?

Joint species distribution models (JSDMs)

- A form of multi-species model that has historically ignored imperfect detection

Joint species distribution models (JSDMs)

- A form of multi-species model that has historically ignored imperfect detection
- Explicitly account for *residual species correlations* by estimating a covariance matrix

Joint species distribution models (JSDMs)

- A form of multi-species model that has historically ignored imperfect detection
- Explicitly account for *residual species correlations* by estimating a covariance matrix
- Residual correlations can come from
  - Biotic interactions (e.g., competition, facilitation, predation)

Joint species distribution models (JSDMs)

- A form of multi-species model that has historically ignored imperfect detection
- Explicitly account for *residual species correlations* by estimating a covariance matrix
- Residual correlations can come from
  - Biotic interactions (e.g., competition, facilitation, predation)
  - Missing covariates

Joint species distribution models (JSDMs)

- A form of multi-species model that has historically ignored imperfect detection
- Explicitly account for *residual species correlations* by estimating a covariance matrix
- Residual correlations can come from
    - Biotic interactions (e.g., competition, facilitation, predation)
    - Missing covariates
- Estimating residual correlations can provide additional insights for hypothesis testing and underlying drivers of community patterns

Historical differences between JSDMs and MSOMs

Historical differences between JSDMs and MSOMs

Joint Species Distribution Models

- Don’t account for imperfect detection
- Account for residual species correlations
- Sometimes treat species-specific effects as random effects
- Key references:
  - Latimer et al. (2009)
  - Ovaskainen et al. (2010)
  - Warton et al. (2015)

Multi-species Occupancy Models

- Account for imperfect detection
- Don't account for residual species correlations
- Treat species specific-effects as random effects
- Key references:
  - Dorazio and Royle (2005)
  - Gelfand et al. (2005)
  - Devarajan et al. (2020)

Merging JSDMs and MSOMs

Ecology, 100(8), 2019, e02754

Joint species distribution models with species correlations and imperfect detection

Mathias W. Tobler,[1,5] Marc Kéry,[2] Francis K. C. Hui,[3] Gurutzeta Guillera-Arroita,[4] Peter Knaus,[2] and Thomas Sattler[2]

Merging JSDMs and MSOMs

Ecology, 100(8), 2019, e02754

Joint species distribution models with species correlations and imperfect detection

Mathias W. Tobler,[1,5] Marc Kéry,[2] Francis K. C. Hui,[3] Gurutzeta Guillera-Arroita,[4] Peter Knaus,[2] and Thomas Sattler[2]

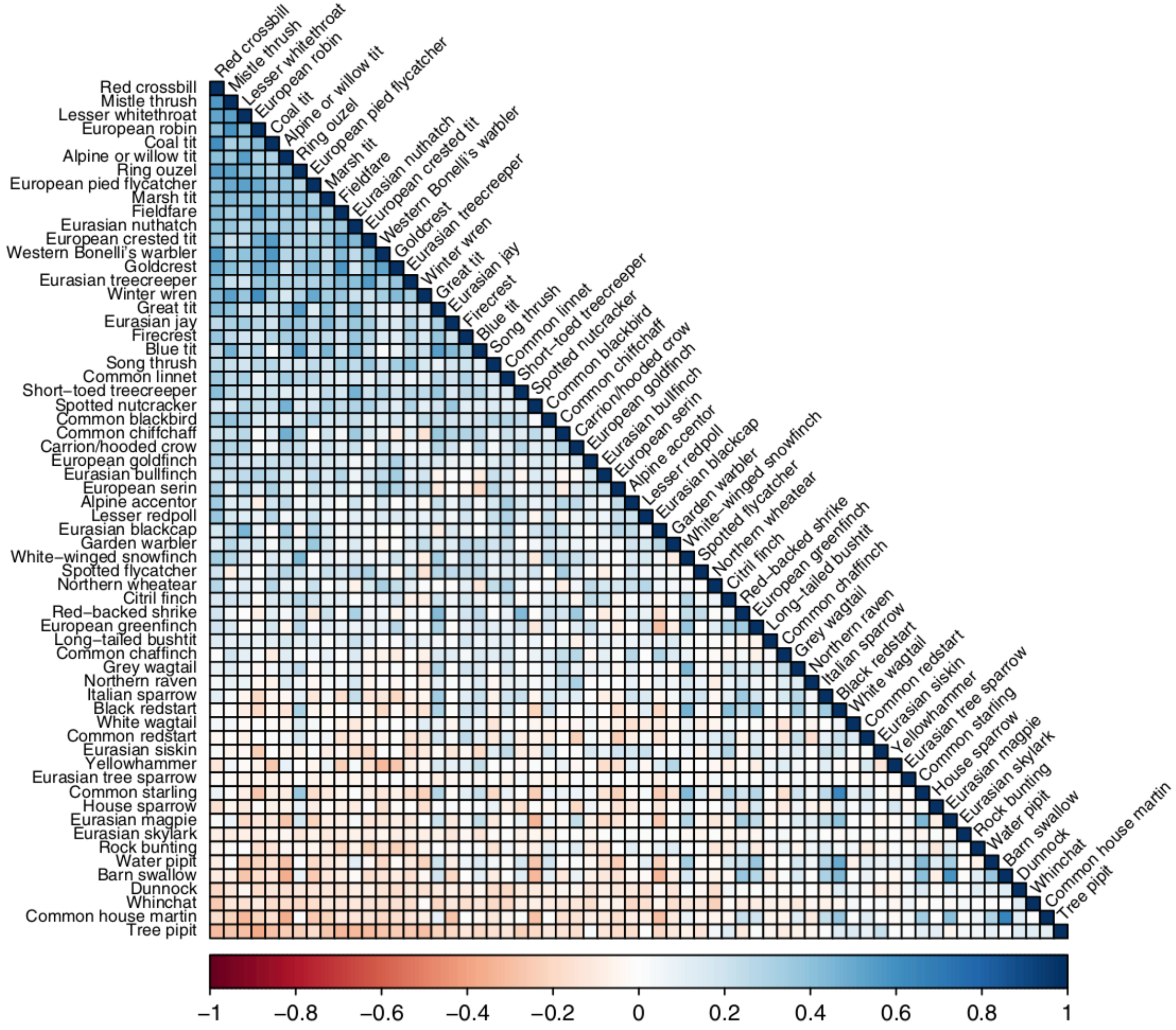
Joint species distribution models that account for imperfect detection

=

Multi-species occupancy models that account for species correlations

JSDMs with imperfect detection

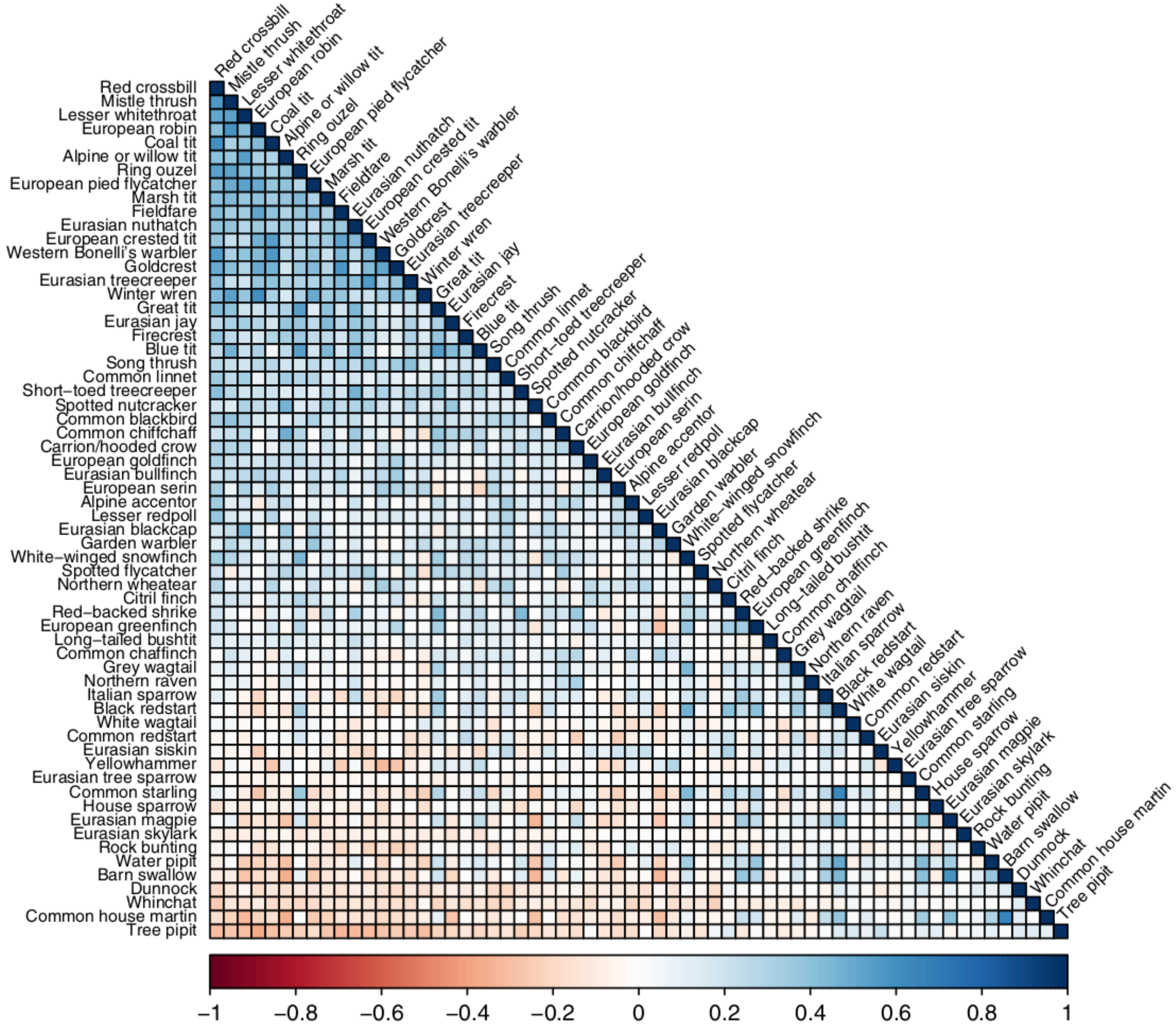
- Residual correlation matrices of occupancy probability can provide insights on underlying biotic/abiotic factors



Tobler et al. (2019)

JSDMs with imperfect detection

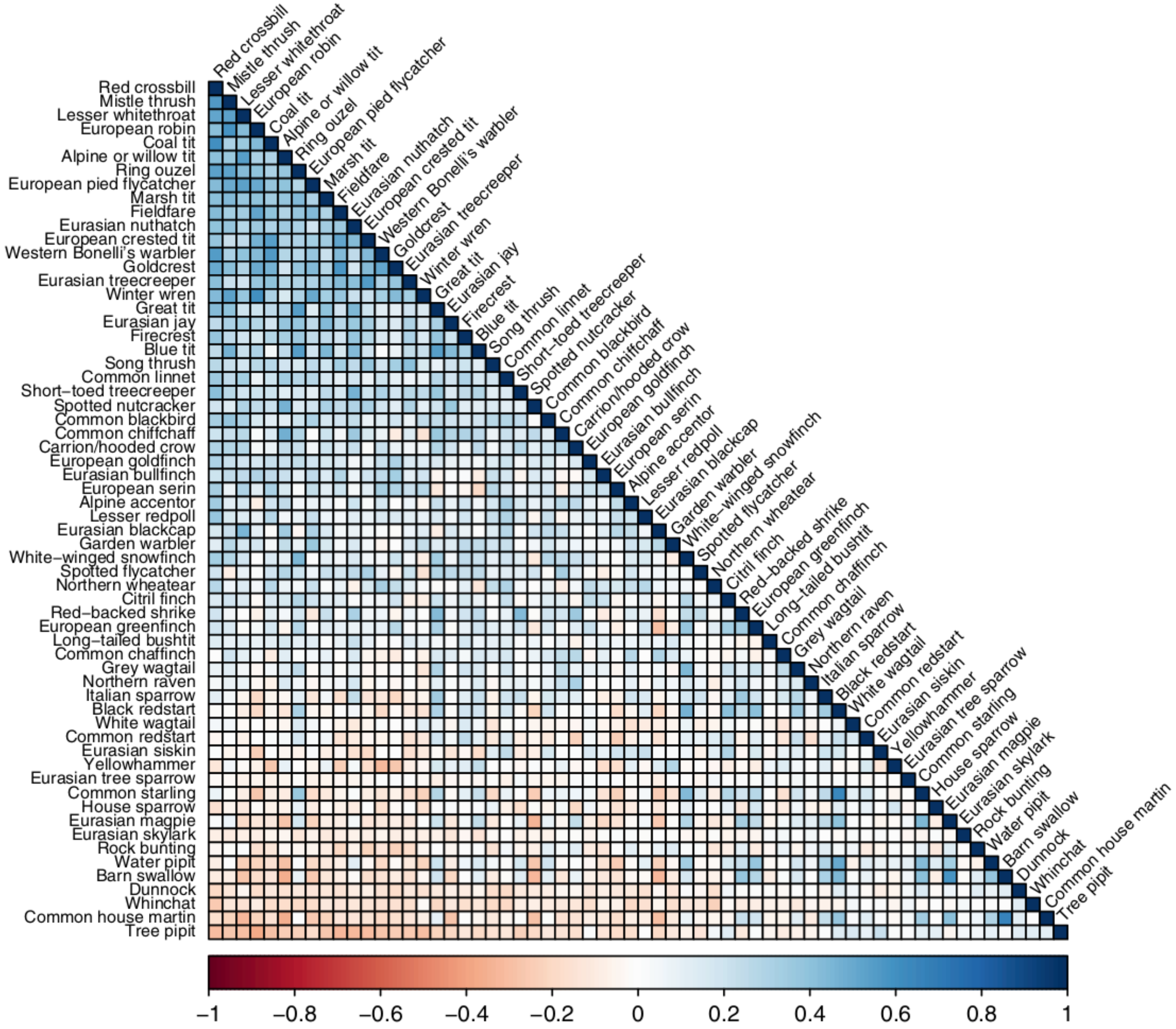
- Residual correlation matrices of occupancy probability can provide insights on underlying biotic/abiotic factors
- We could try to explicitly estimate every single correlation between each pair of species in the community.



Tobler et al. (2019)

JSDMs with imperfect detection

- Residual correlation matrices of occupancy probability can provide insights on underlying biotic/abiotic factors
- We could try to explicitly estimate every single correlation between each pair of species in the community.
- Quickly becomes a lot of parameters and very difficult to estimate.
- Instead, we will use a dimension reduction approach called **Factor Modelling.**



Tobler et al. (2019)

Latent factor multi-species occupancy model

Latent factor multi-species occupancy model

- Conceptually: account for residual species correlations by estimating species-specific responses (loadings) to a set of latent variables (factors).

Latent factor multi-species occupancy model

- Conceptually: account for residual species correlations by estimating species-specific responses (loadings) to a set of latent variables (factors).
- Can view the latent factors as "missing covariates"

Latent factor multi-species occupancy model

- Conceptually: account for residual species correlations by estimating species-specific responses (loadings) to a set of latent variables (factors).
- Can view the latent factors as "missing covariates"
- Each species has a unique effect of each latent factor. This effect is called a **factor loading**.

Latent factor multi-species occupancy model

- Conceptually: account for residual species correlations by estimating species-specific responses (loadings) to a set of latent variables (factors).
- Can view the latent factors as "missing covariates"
- Each species has a unique effect of each latent factor. This effect is called a **factor loading**.
- Species with similar factor loadings will tend to co-occur together.
- Can ultimately derive an interspecies covariance (or correlation) matrix

Latent factor multi-species occupancy model

- Conceptually: account for residual species correlations by estimating species-specific responses (loadings) to a set of latent variables (factors).
- Can view the latent factors as "missing covariates"
- Each species has a unique effect of each latent factor. This effect is called a **factor loading**.
- Species with similar factor loadings will tend to co-occur together.
- Can ultimately derive an interspecies covariance (or correlation) matrix
- This is a form of factor analysis (similar to PCA)

Latent factor multi-species occupancy model

Example: one occupancy covariate and two latent factors

Occupancy (ecological) sub-model

$$
\begin{aligned}
z_{i,j} &\sim \text{Bernoulli}(\psi_{i,j}) \\
\text{logit}(\psi_{i,j}) &= \beta_{1,i} + \beta_{2,i} \cdot X_{2,j} + \lambda_{i,1} \cdot \text{w}_{1,j} + \lambda_{i,2} \cdot \text{w}_{2,j} \\
\beta_{r,i} &\sim \text{Normal}(\mu_{\beta_r}, \tau^2_{\beta_r})
\end{aligned}
$$

Detection (observation) sub-model

$$
\begin{aligned}
y_{i,j,k} &\sim \text{Bernoulli}(p_{i,j,k} \cdot z_{i,j}) \\
\text{logit}(p_{i,j,k}) &= \alpha_{1,i} + \alpha_{2,i} \cdot V_{2,j,k} + \cdots + \alpha_{r,i} \cdot V_{r,j,k} \\
\alpha_{r,i} &\sim \text{Normal}(\mu_{\alpha_r}, \tau^2_{\alpha,r})
\end{aligned}
$$

Latent factor multi-species occupancy model

Example: one occupancy covariate and two latent factors

Occupancy (ecological) sub-model

$$
\begin{aligned}
z_{i,j} &\sim \text{Bernoulli}(\psi_{i,j}) \\
\text{logit}(\psi_{i,j}) &= \beta_{1,i} + \beta_{2,i} \cdot X_{2,j} + \boxed{\lambda_{i,1} \cdot \text{w}_{1,j} + \lambda_{i,2} \cdot \text{w}_{2,j}} \\
\beta_{r,i} &\sim \text{Normal}(\mu_{\beta_r}, \tau^2_{\beta_r})
\end{aligned}
$$

Detection (observation) sub-model

$$
\begin{aligned}
y_{i,j,k} &\sim \text{Bernoulli}(p_{i,j,k} \cdot z_{i,j}) \\
\text{logit}(p_{i,j,k}) &= \alpha_{1,i} + \alpha_{2,i} \cdot V_{2,j,k} + \cdots + \alpha_{r,i} \cdot V_{r,j,k} \\
\alpha_{r,i} &\sim \text{Normal}(\mu_{\alpha_r}, \tau^2_{\alpha,r})
\end{aligned}
$$

Latent factor multi-species occupancy model

- Example: one covariate and two factors ("missing covariates")

$$\text{logit}(\psi_{i,j}) = \beta_{1,i} + \beta_{2,i} \cdot X_{2,j} + \lambda_{i,1} \cdot \mathrm{w}_{1,j} + \lambda_{i,2} \cdot \mathrm{w}_{2,j}$$

Latent factor multi-species occupancy model

- Example: one covariate and two factors ("missing covariates")

$$\mathrm{logit}(\psi_{i,j}) = \beta_{1,i} + \beta_{2,i} \cdot X_{2,j} + \lambda_{i,1} \cdot \mathrm{w}_{1,j} + \lambda_{i,2} \cdot \mathrm{w}_{2,j}$$

"Missing covariates" that account for residual species correlations

Latent factor multi-species occupancy model

- Example: one covariate and two factors ("missing covariates")

$$\text{logit}(\psi_{i,j}) = \beta_{1,i} + \beta_{2,i} \cdot X_{2,j} + \lambda_{i,1} \cdot \mathrm{w}_{1,j} + \lambda_{i,2} \cdot \mathrm{w}_{2,j}$$

Effects of the missing covariates

Factor loadings matrix

- N x q matrix.
- Note that N is typically much bigger than q, resulting in a "tall and skinny" matrix.
- Example: 10 species and 3 factors

q= 3 factors (columns)

N = 10 species (rows)

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} \\ \lambda_{2,1} & \lambda_{2,2} & \lambda_{2,3} \\ \lambda_{3,1} & \lambda_{3,2} & \lambda_{3,3} \\ \lambda_{4,1} & \lambda_{4,2} & \lambda_{4,3} \\ \lambda_{5,1} & \lambda_{5,2} & \lambda_{5,3} \\ \lambda_{6,1} & \lambda_{6,2} & \lambda_{6,3} \\ \lambda_{7,1} & \lambda_{7,2} & \lambda_{7,3} \\ \lambda_{8,1} & \lambda_{8,2} & \lambda_{8,3} \\ \lambda_{9,1} & \lambda_{9,2} & \lambda_{9,3} \\ \lambda_{10,1} & \lambda_{10,2} & \lambda_{10,3} \end{bmatrix}$$

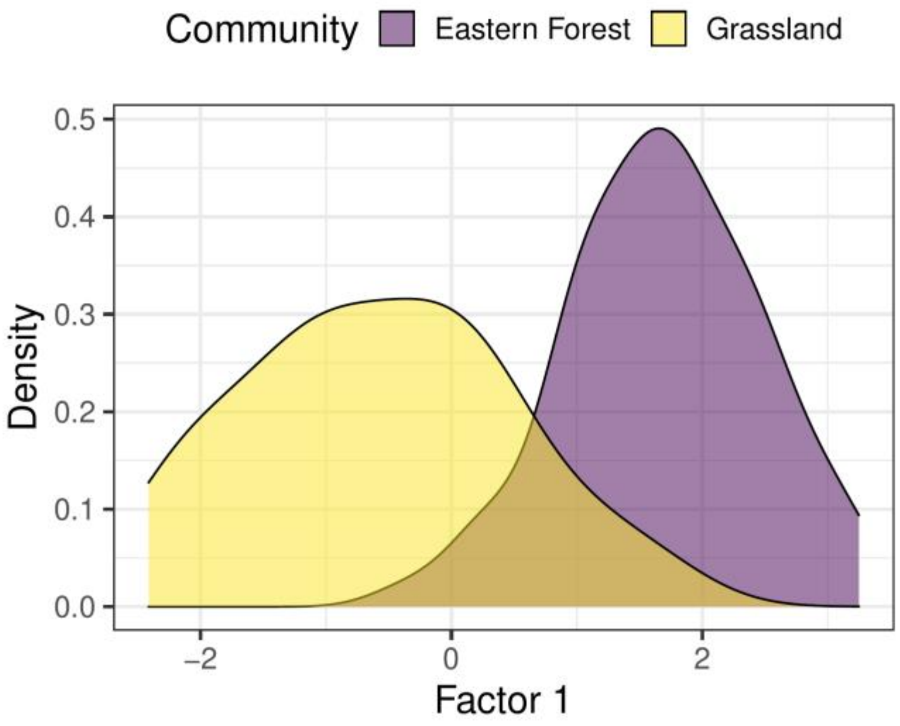
Deriving an interspecies covariance matrix

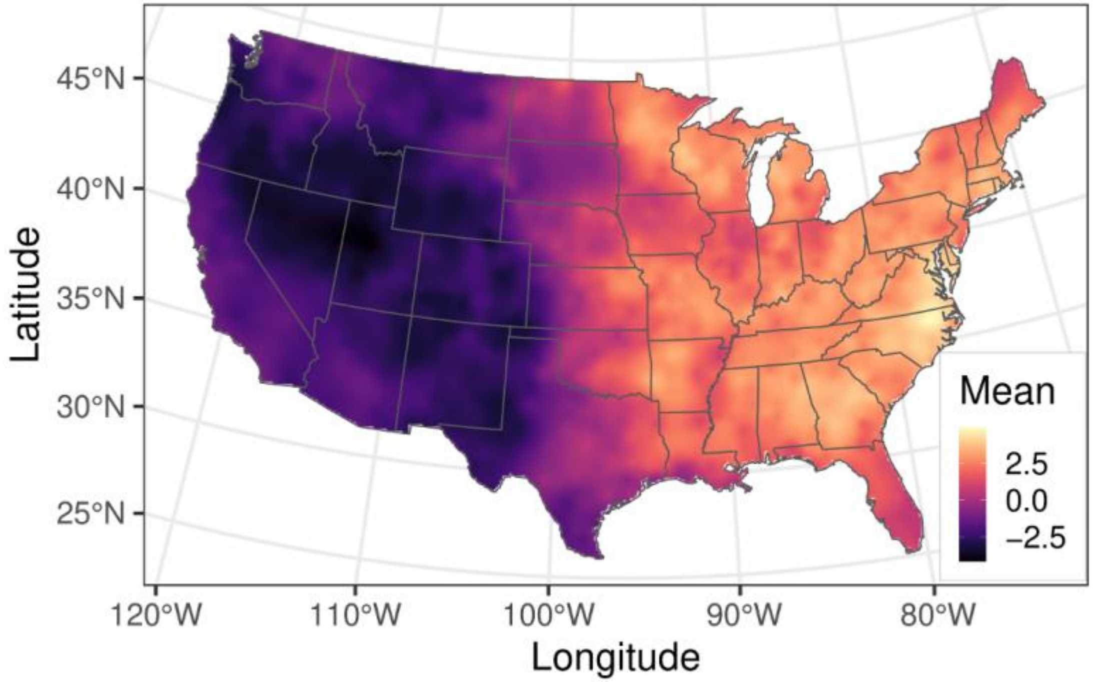
$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^\top$$

$$\begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} \\ \lambda_{2,1} & \lambda_{2,2} & \lambda_{2,3} \\ \lambda_{3,1} & \lambda_{3,2} & \lambda_{3,3} \\ \lambda_{4,1} & \lambda_{4,2} & \lambda_{4,3} \\ \lambda_{5,1} & \lambda_{5,2} & \lambda_{5,3} \\ \lambda_{6,1} & \lambda_{6,2} & \lambda_{6,3} \\ \lambda_{7,1} & \lambda_{7,2} & \lambda_{7,3} \\ \lambda_{8,1} & \lambda_{8,2} & \lambda_{8,3} \\ \lambda_{9,1} & \lambda_{9,2} & \lambda_{9,3} \\ \lambda_{10,1} & \lambda_{10,2} & \lambda_{10,3} \end{bmatrix} \begin{bmatrix} \lambda_{1,1} & \lambda_{2,1} & \lambda_{3,1} & \lambda_{4,1} & \lambda_{5,1} & \lambda_{6,1} & \lambda_{7,1} & \lambda_{8,1} & \lambda_{9,1} & \lambda_{10,1} \\ \lambda_{1,2} & \lambda_{2,2} & \lambda_{3,2} & \lambda_{4,2} & \lambda_{5,2} & \lambda_{6,2} & \lambda_{7,2} & \lambda_{8,2} & \lambda_{9,2} & \lambda_{10,2} \\ \lambda_{1,3} & \lambda_{2,3} & \lambda_{3,3} & \lambda_{4,3} & \lambda_{5,3} & \lambda_{6,3} & \lambda_{7,3} & \lambda_{8,3} & \lambda_{9,3} & \lambda_{10,3} \end{bmatrix}$$

Interpretation of the latent factors and factor loadings

Can be used as a model-based ordination technique by assessing species with similar factor loadings

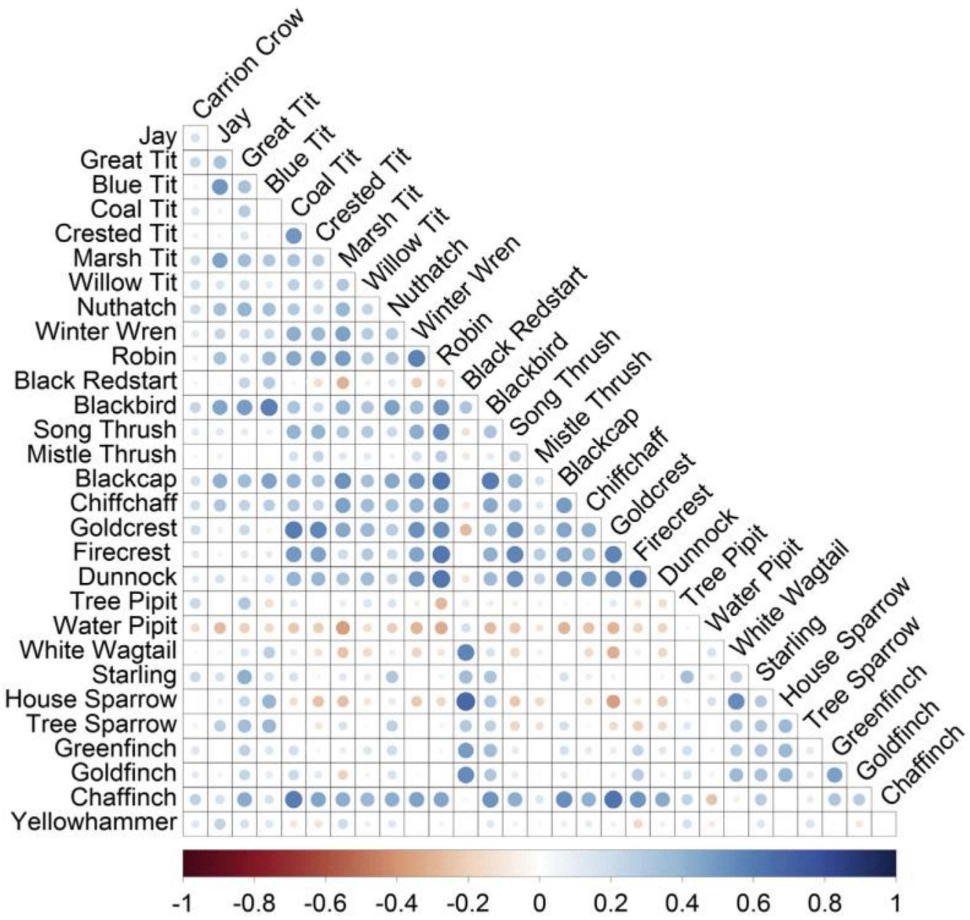




Doser, Finley, Banerjee (2023) Ecology

Interpretation of residual correlation matrix

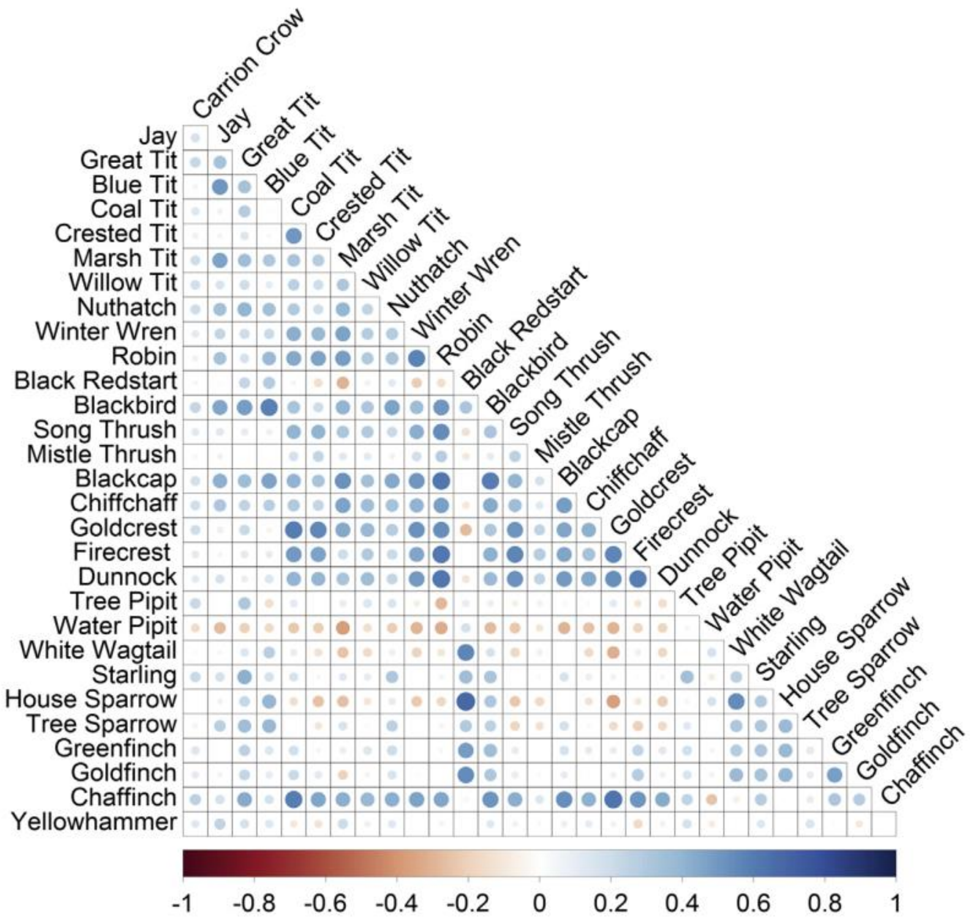
- Can provide insights on missing covariates and/or biological processes driving co-occurrence patterns



Kéry and Royle (2021) Chapter 8

Interpretation of residual correlation matrix

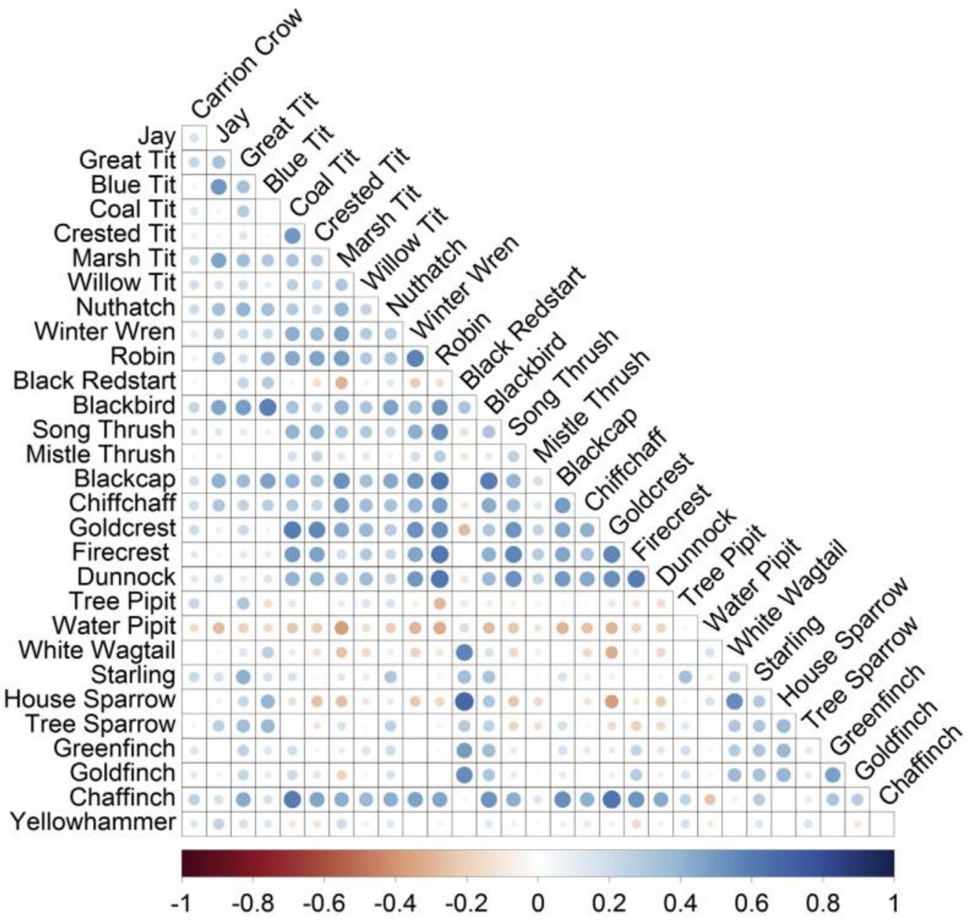
- Can provide insights on missing covariates and/or biological processes driving co-occurrence patterns
- Need to be cautious in interpreting the correlations: these are not true biological interactions!



Kéry and Royle (2021) Chapter 8

Interpretation of residual correlation matrix

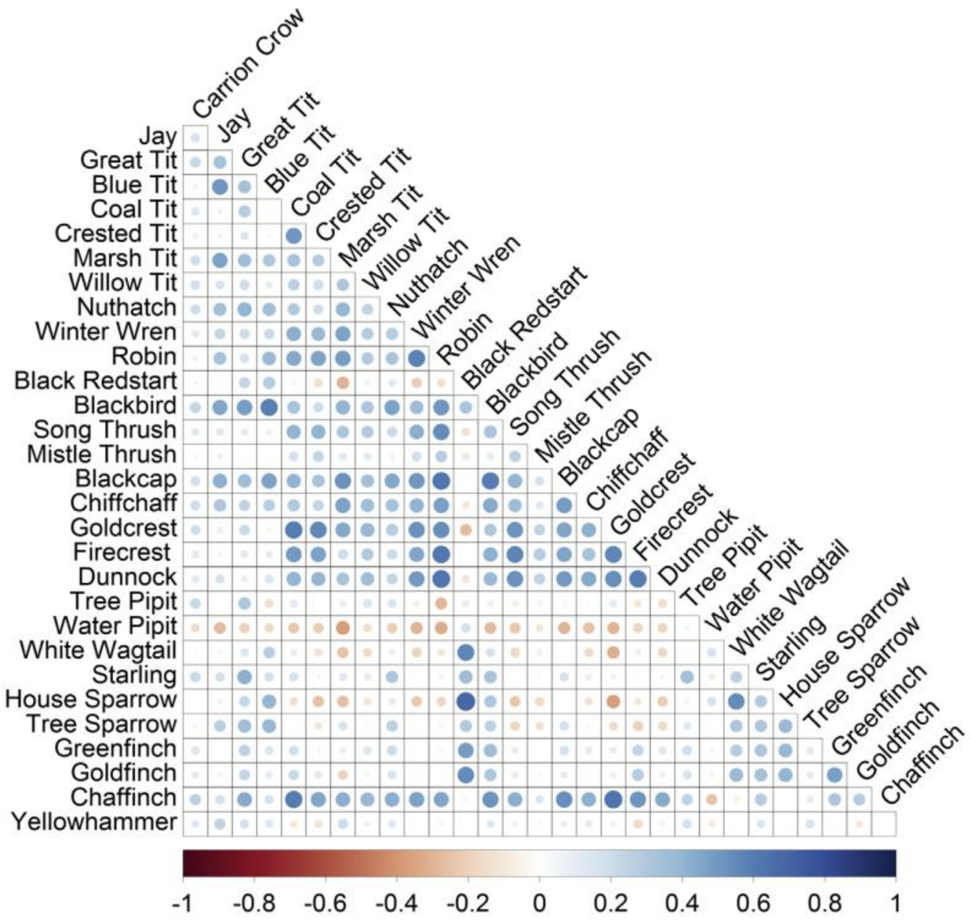
- Can provide insights on missing covariates and/or biological processes driving co-occurrence patterns
- Need to be cautious in interpreting the correlations: these are not true biological interactions!
- Interpretation depends on scale



Kéry and Royle (2021) Chapter 8

Interpretation of residual correlation matrix

- Can provide insights on missing covariates and/or biological processes driving co-occurrence patterns
- Need to be cautious in interpreting the correlations: these are not true biological interactions!
- Interpretation depends on scale
- See Poggiato et al. (2021) *TREE* and Chapter 8 Kéry and Royle (2021)



Kéry and Royle (2021) Chapter 8

Pros and cons of latent factor MSOM vs. Regular MSOM

Pros

- Often improves model fit, predictive performance, and sometimes even precision of estimates.
- Arguably more biologically realistic since we know species are not independent of each other.
- Can be used for model-based ordination.
- Helps generate hypotheses.

Pros and cons of latent factor MSOM vs. Regular MSOM

Pros

- Often improves model fit, predictive performance, and sometimes even precision of estimates.
- Arguably more biologically realistic since we know species are not independent of each other.
- Can be used for model-based ordination.
- Helps generate hypotheses.

Cons

- Usually requires more data.
- Need to choose the number of factors to include.
- Potential for overinterpretation of residual covariance matrix
- Convergence and model fitting can be very tricky. Need to place restrictions on the factor loadings

How to choose the number of factors (q)?

- Similar to performing a PCA and choosing the number of components to look at.

How to choose the number of factors (q)?

- Similar to performing a PCA and choosing the number of components to look at.
- More factors means more parameters to estimate.

How to choose the number of factors (q)?

- Similar to performing a PCA and choosing the number of components to look at.
- More factors means more parameters to estimate.
- I often start out with between 5-10 depending on size of community.

How to choose the number of factors (q)?

- Similar to performing a PCA and choosing the number of components to look at.
- More factors means more parameters to estimate.
- I often start out with between 5-10 depending on size of community.
- Things to consider:
    - Size of community and how much variation you expect across species.

How to choose the number of factors (q)?

- Similar to performing a PCA and choosing the number of components to look at.
- More factors means more parameters to estimate.
- I often start out with between 5-10 depending on size of community.
- Things to consider:
    - Size of community and how much variation you expect across species.
    - Amount of computational power/time.

How to choose the number of factors (q)?

- Similar to performing a PCA and choosing the number of components to look at.
- More factors means more parameters to estimate.
- I often start out with between 5-10 depending on size of community.
- Things to consider:
    - Size of community and how much variation you expect across species.
    - Amount of computational power/time.
    - Size of data set and rarity of species in data set.

How to choose the number of factors (q)?

- Similar to performing a PCA and choosing the number of components to look at.
- More factors means more parameters to estimate.
- I often start out with between 5-10 depending on size of community.
- Things to consider:
    - Size of community and how much variation you expect across species.
    - Amount of computational power/time.
    - Size of data set and rarity of species in data set.
- See guidance here.

Priors and constraints for the factor loadings

- Without any restrictions, there are infinitely many possible values for the factor loadings and factors.
- We fix the factor loadings on the diagonal to 1 and all values in the "upper triangle" to 0.
- All others have a Normal(0, 1) prior distribution.
- All other priors same as before.

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \lambda_{1,3} \\ \lambda_{2,1} & \lambda_{2,2} & \lambda_{2,3} \\ \lambda_{3,1} & \lambda_{3,2} & \lambda_{3,3} \\ \lambda_{4,1} & \lambda_{4,2} & \lambda_{4,3} \\ \lambda_{5,1} & \lambda_{5,2} & \lambda_{5,3} \\ \lambda_{6,1} & \lambda_{6,2} & \lambda_{6,3} \\ \lambda_{7,1} & \lambda_{7,2} & \lambda_{7,3} \\ \lambda_{8,1} & \lambda_{8,2} & \lambda_{8,3} \\ \lambda_{9,1} & \lambda_{9,2} & \lambda_{9,3} \\ \lambda_{10,1} & \lambda_{10,2} & \lambda_{10,3} \end{bmatrix}$$

Priors and constraints for the factor loadings

- Without any restrictions, there are infinitely many possible values for the factor loadings and factors.
- We fix the factor loadings on the diagonal to 1 and all values in the "upper triangle" to 0.
- All others have a Normal(0, 1) prior distribution.
- All other priors same as before.

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{2,1} & 1 & 0 \\ \lambda_{3,1} & \lambda_{3,2} & 1 \\ \lambda_{4,1} & \lambda_{4,2} & \lambda_{4,3} \\ \lambda_{5,1} & \lambda_{5,2} & \lambda_{5,3} \\ \lambda_{6,1} & \lambda_{6,2} & \lambda_{6,3} \\ \lambda_{7,1} & \lambda_{7,2} & \lambda_{7,3} \\ \lambda_{8,1} & \lambda_{8,2} & \lambda_{8,3} \\ \lambda_{9,1} & \lambda_{9,2} & \lambda_{9,3} \\ \lambda_{10,1} & \lambda_{10,2} & \lambda_{10,3} \end{bmatrix}$$

Ordering of species in the data set

- Notice the first q species (rows) have constraints in their factor loadings.
- This means the order of species in our data set could impact how well our model converges/fits.
- In particular, we may need to carefully choose the first q species.

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{2,1} & 1 & 0 \\ \lambda_{3,1} & \lambda_{3,2} & 1 \\ \lambda_{4,1} & \lambda_{4,2} & \lambda_{4,3} \\ \lambda_{5,1} & \lambda_{5,2} & \lambda_{5,3} \\ \lambda_{6,1} & \lambda_{6,2} & \lambda_{6,3} \\ \lambda_{7,1} & \lambda_{7,2} & \lambda_{7,3} \\ \lambda_{8,1} & \lambda_{8,2} & \lambda_{8,3} \\ \lambda_{9,1} & \lambda_{9,2} & \lambda_{9,3} \\ \lambda_{10,1} & \lambda_{10,2} & \lambda_{10,3} \end{bmatrix}$$

Choosing the first q species

- Put a common species first
- For the remaining q – 1 factors, place species that you a priori believe may have different occurrence patterns than the first species, as well as the other species placed before it.
- Adjust species ordering after an initial model fit.

Convergence diagnostics and other considerations when fitting spatial occupancy models

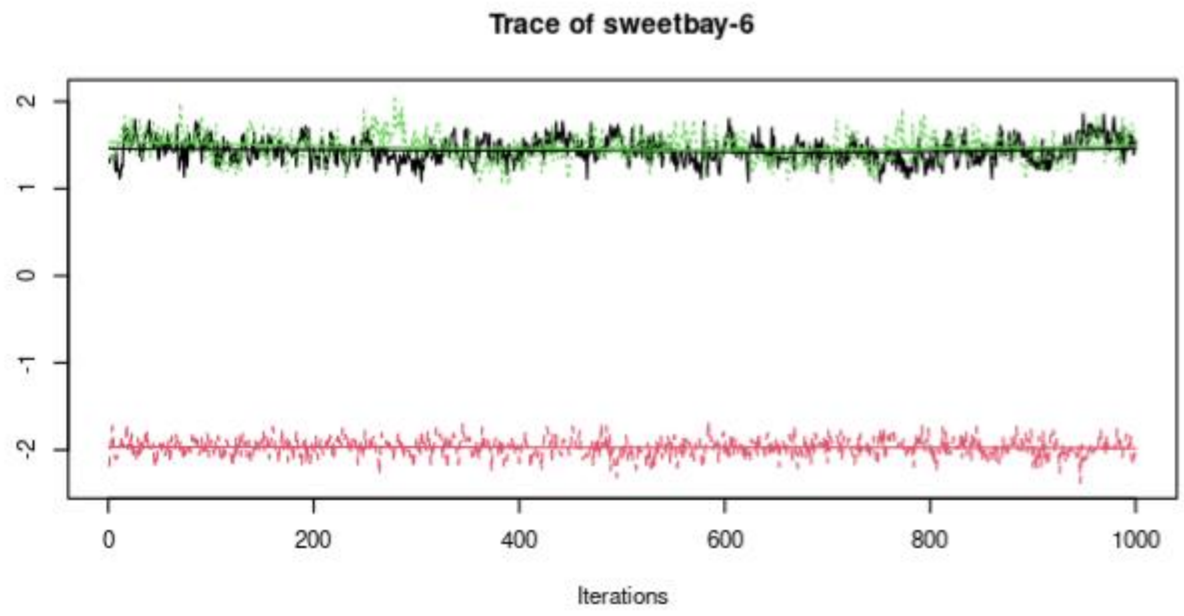
Jeffrey W. Doser

2023 (last update: March 25, 2024)

Source: vignettes/modelConsiderations.Rmd

Initial values can be very important!!

- In general, the more complex the model, the more important the initial values can be.
- Drastically different initial values in latent factor models can lead to chain "flip-flopping", where any individual chain looks converged, but across chains there are drastically different values.
- May need to manually set initial values or only run a single chain. See guidance here.



Fitting the latent factor MSOM in `spOccupancy`

- Function `lfMsPGOcc()` (latent factor multi-species Pólya-Gamma occupancy model)
- Fits the model developed by Tobler et al. (2019) with some adaptations to make it faster
- Same arguments as `msPGOcc()` with the addition of n.factors to specify the number of latent factors to use.

Exercise: Landscape scale effects on occupancy of tropical amphibians

`5-amphibian-multi-species-occ.R`