

# Fixed and random effects in HMSC

Prof. Otso Ovaskainen

University of Jyväskylä, Finland

University of Helsinki, Finland

NTNU Trondheim, Norway



JYVÄSKYLÄN YLIOPISTO  
UNIVERSITY OF JYVÄSKYLÄ

# Overview of the structure of HMSC

**HMSC is a multivariate hierarchical generalized linear mixed model fitted with Bayesian inference**

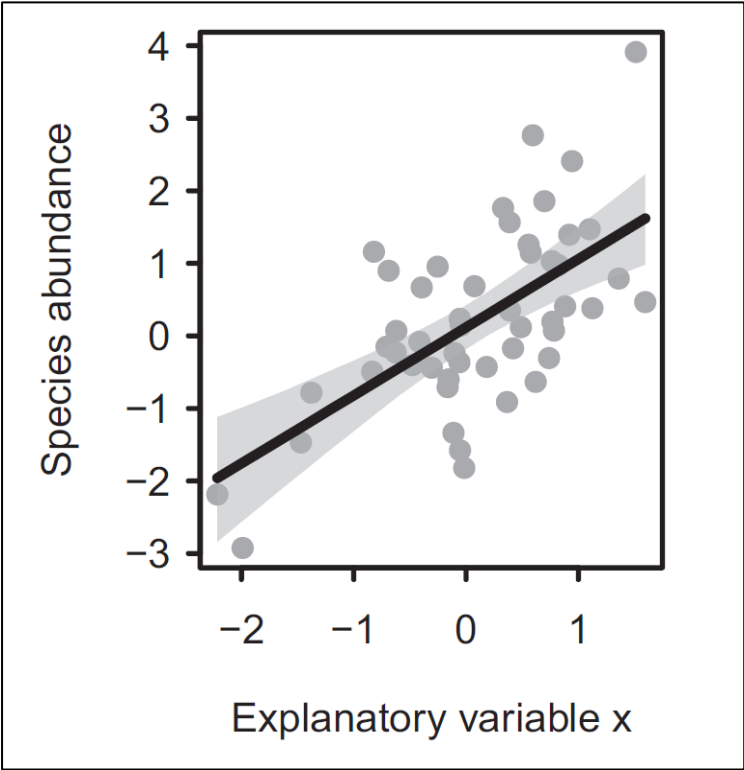
Table 4.1 *Definitions of statistical frameworks as used in this book.*

Name of the statistical framework as used in this book	Number of explanatory variables	Types of explanatory variables	Number of response variables	Types of response variables	Random effects	Introduced in
Linear model	Zero or more	Continuous or categorical	One	Only normally distributed	Does not include	<a href="#">Section 5.2</a>
Generalised linear model	Zero or more	Continuous or categorical	One	Can be non-normally distributed	Does not include	<a href="#">Section 5.3</a>
Linear mixed model	Zero or more	Continuous or categorical	One	Only normally distributed	Can include	<a href="#">Section 5.4</a>
Generalised linear mixed model	Zero or more	Continuous or categorical	One	Can be non-normally distributed	Can include	<a href="#">Section 5.4</a>
Multivariate linear model	Zero or more	Continuous or categorical	One or more	Only normally distributed	Does not include	<a href="#">Section 6.1</a>
Multivariate generalised linear model	Zero or more	Continuous or categorical	One or more	Can be non-normally distributed	Does not include	<a href="#">Section 6.1</a>
Multivariate linear mixed model	Zero or more	Continuous or categorical	One or more	Only normally distributed	Can include	<a href="#">Section 7.3</a>
Multivariate generalised linear mixed model	Zero or more	Continuous or categorical	One or more	Can be non-normally distributed	Can include	<a href="#">Section 7.3</a>

HMSC is a multivariate hierarchical generalized **linear** mixed **model** fitted with Bayesian inference

Table 4.1 *Definitions of statistical frameworks as used in this book.*

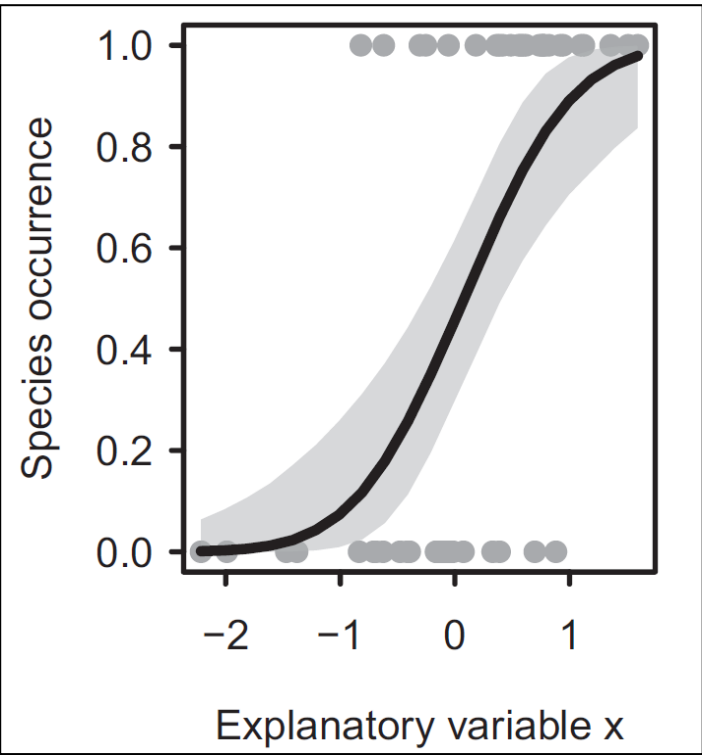
Name of the statistical framework as used in this book	Number of explanatory variables	Types of explanatory variables	Number of response variables	Types of response variables	Random effects	Introduced in
Linear model	Zero or more	Continuous or categorical	One	Only normally distributed	Does not include	<a href="#">Section 5.2</a>



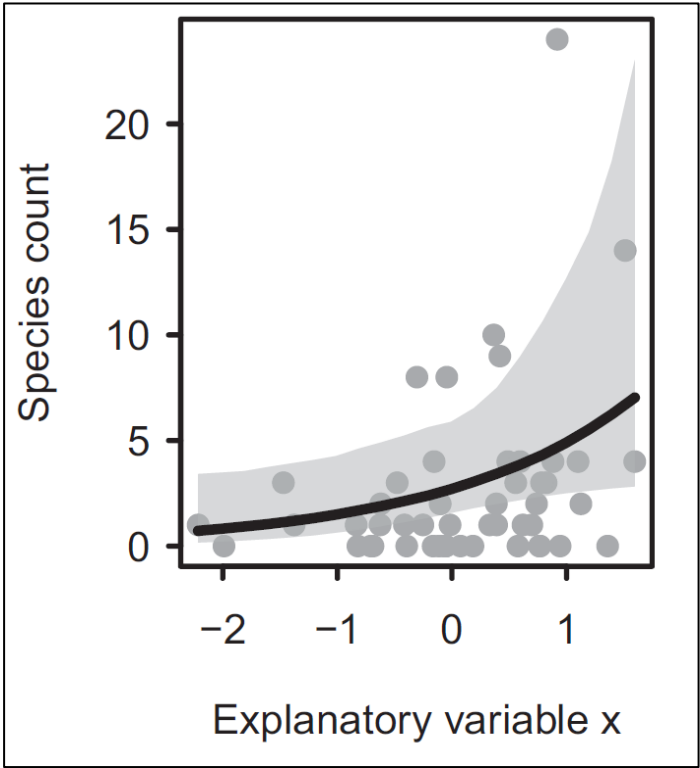
HMSC is a multivariate hierarchical **generalized linear** mixed **model** fitted with Bayesian inference

Table 4.1 *Definitions of statistical frameworks as used in this book.*

Name of the statistical framework as used in this book	Number of explanatory variables	Types of explanatory variables	Number of response variables	Types of response variables	Random effects	Introduced in
Generalised linear model	Zero or more	Continuous or categorical	One	Can be non-normally distributed	Does not include	<a href="#">Section 5.3</a>



OR

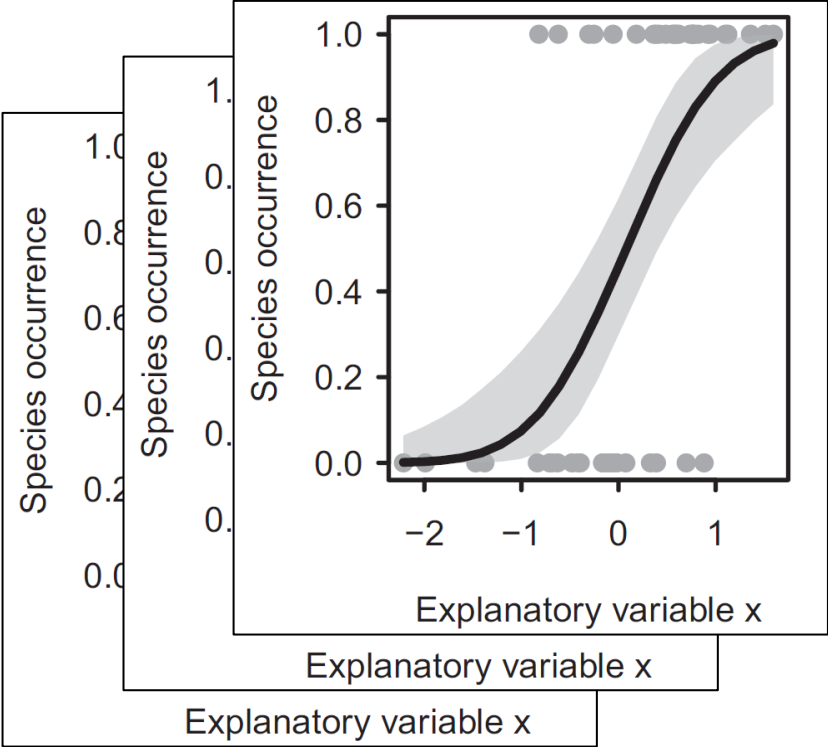


HMSC is a **multivariate** hierarchical **generalized linear** mixed **model** fitted with Bayesian inference

Table 4.1 *Definitions of statistical frameworks as used in this book.*

Name of the statistical framework as used in this book	Number of explanatory variables	Types of explanatory variables	Number of response variables	Types of response variables	Random effects	Introduced in
Multivariate generalised linear model	Zero or more	Continuous or categorical	One or more	Can be non-normally distributed	Does not include	<a href="#">Section 6.1</a>

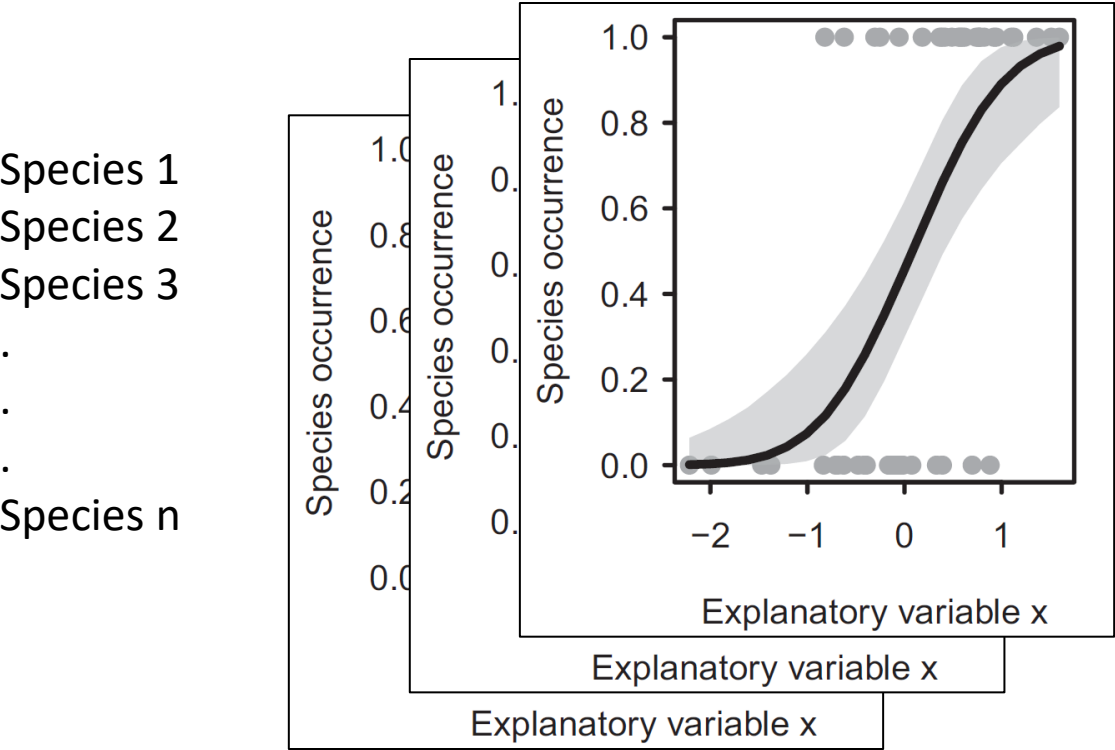
Species 1  
Species 2  
Species 3  
.  
.  
.  
Species n



HMSC is a **multivariate** hierarchical **generalized linear mixed model** fitted with Bayesian inference

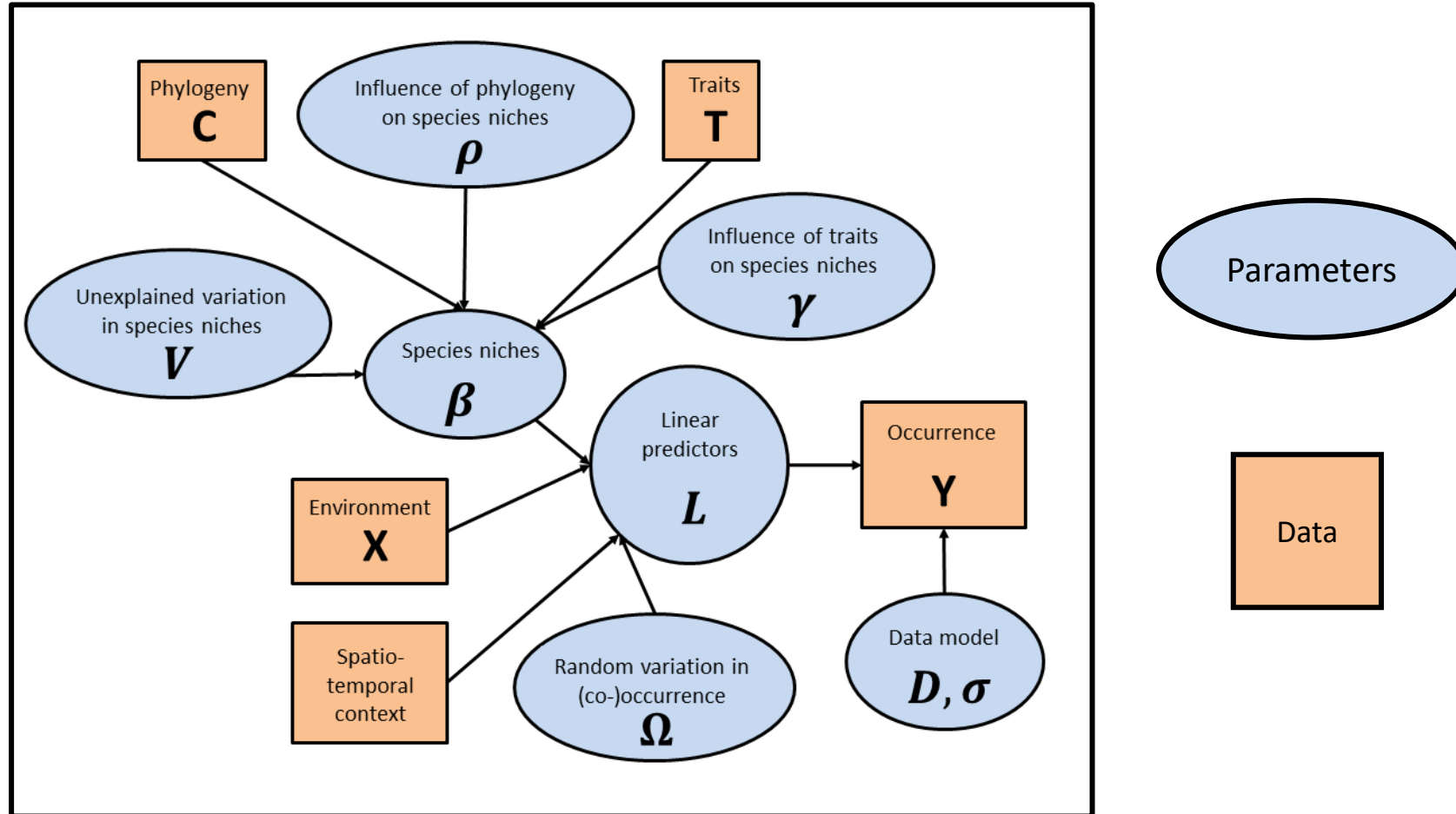
Table 4.1 *Definitions of statistical frameworks as used in this book.*

Name of the statistical framework as used in this book	Number of explanatory variables	Types of explanatory variables	Number of response variables	Types of response variables	Random effects	Introduced in
Multivariate generalised linear mixed model	Zero or more	Continuous or categorical	One or more	Can be non-normally distributed	Can include	<a href="#">Section 7.3</a>

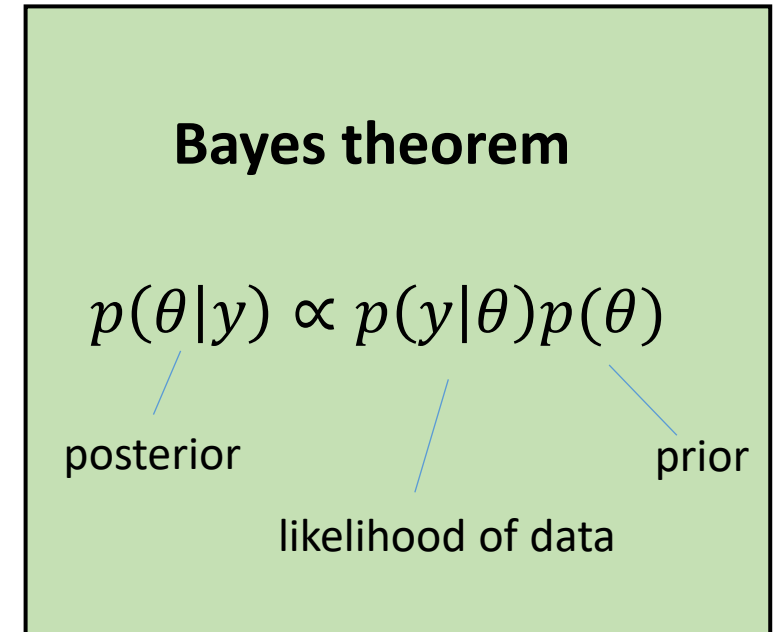
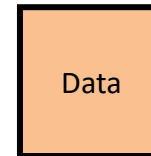
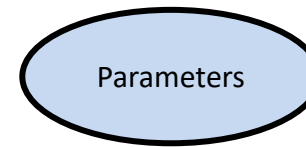
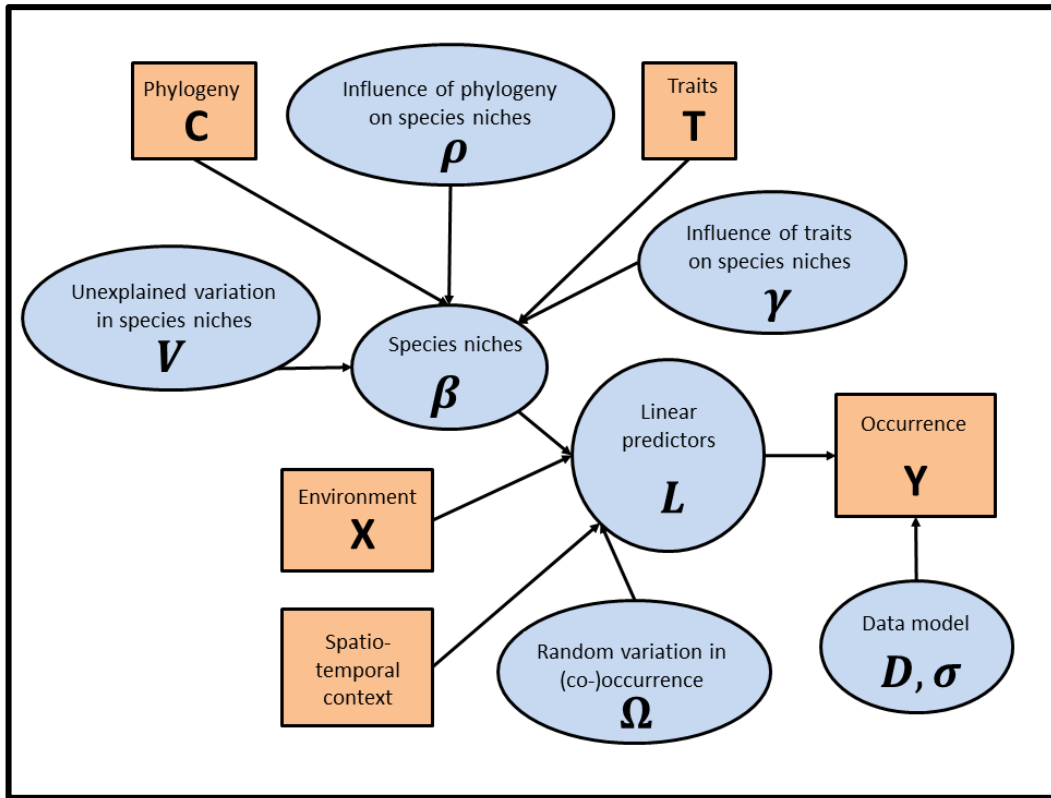


Random effects defined at hierarchically structured, spatially structured or temporally structured units

HMSC is a **multivariate hierarchical generalized linear mixed model** fitted with Bayesian inference

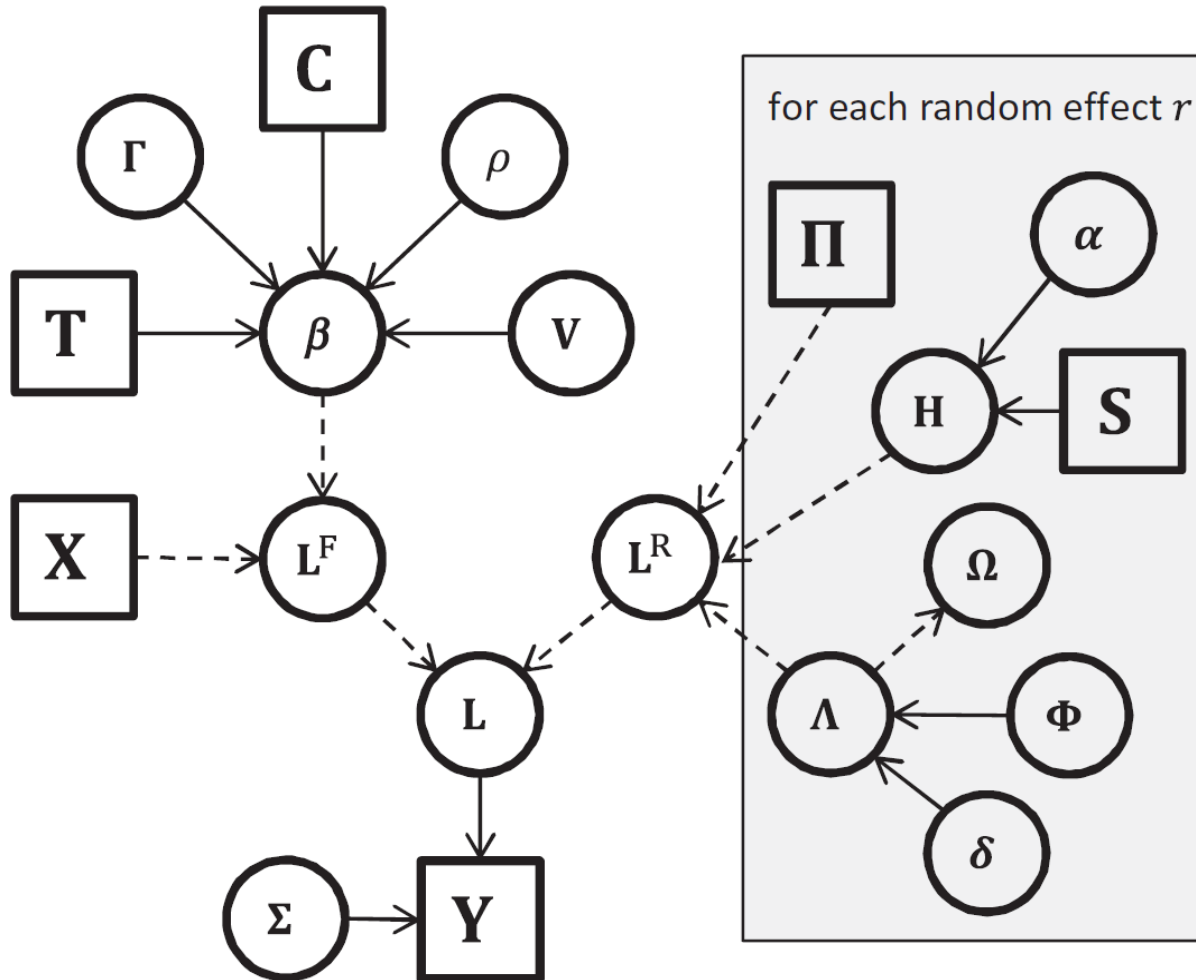


HMSC is a **multivariate hierarchical generalized linear mixed model** fitted with **Bayesian inference**





# HMSC as a directed acyclic graph



Fixed effects      Random effects

$$L_{ij} = L_{ij}^F + L_{ij}^R$$

# What data matrices HMSC takes as input and what are their dimensions?

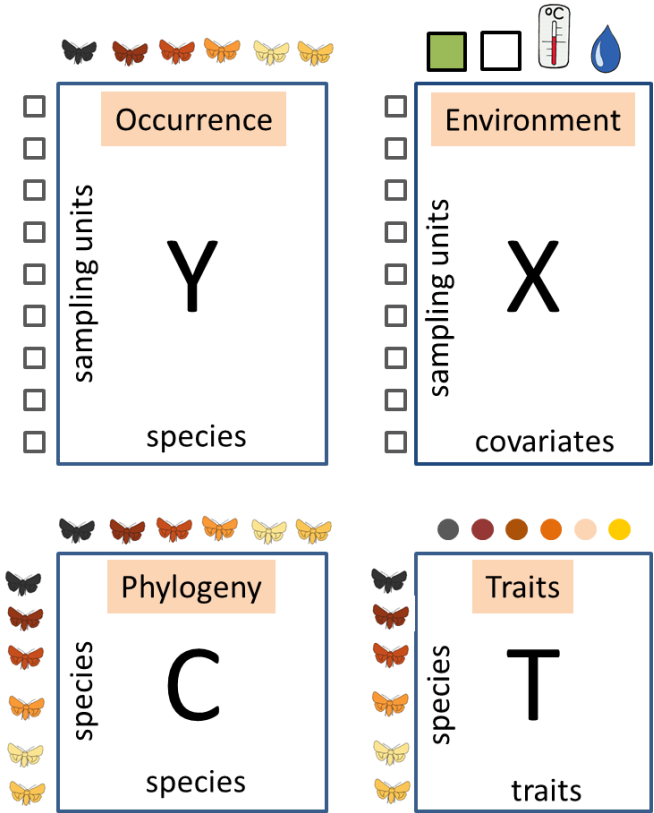
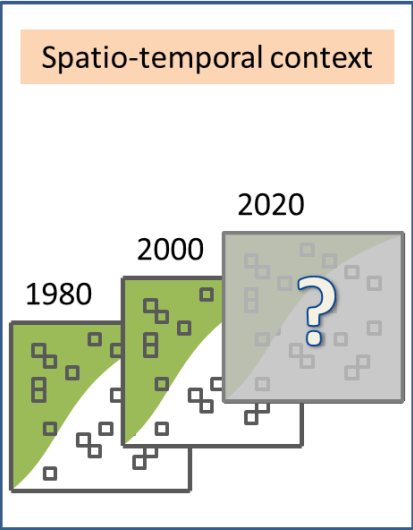


Table 4.2 *Indices and their ranges in the core HMSC model.*

Index and its range	Refers to
$i = 1, \dots, n$	Sampling unit
$j = 1, \dots, n_s$	Species
$k = 1, \dots, n_c$	Environmental covariate
$l = 1, \dots, n_t$	Species trait
$h = 1, \dots, n_f$	Latent factor
$u = 1, \dots, n_u$	Hierarchical unit
$q = 1, \dots, d$	Spatial coordinate in $\mathbb{R}^d$
$r = 1, \dots, n_r$	Random effect

Table 4.3 *Data matrices and their dimensions in the core HMSC model. The spatial coordinates are defined separately for each random effect  $r$ .*

Data matrix	Dimension	Refers to
<b>Y</b> , element $y_{ij}$	$n \times n_s$	Community data
<b>X</b> , element $x_{ik}$	$n \times n_c$	Environmental data
<b>T</b> , element $t_{jl}$	$n_s \times n_t$	Species trait data
<b>C</b> , element $c_{j,j2}$	$n_s \times n_s$	Phylogenetic data
<b><math>\Pi</math></b> , element $\pi_{iu}$	$n \times n_u$	Study design
<b>S</b> , element $s_{uq}$	$n_u \times d$	Spatial coordinates

# What parameters are estimated in an HMSC model?

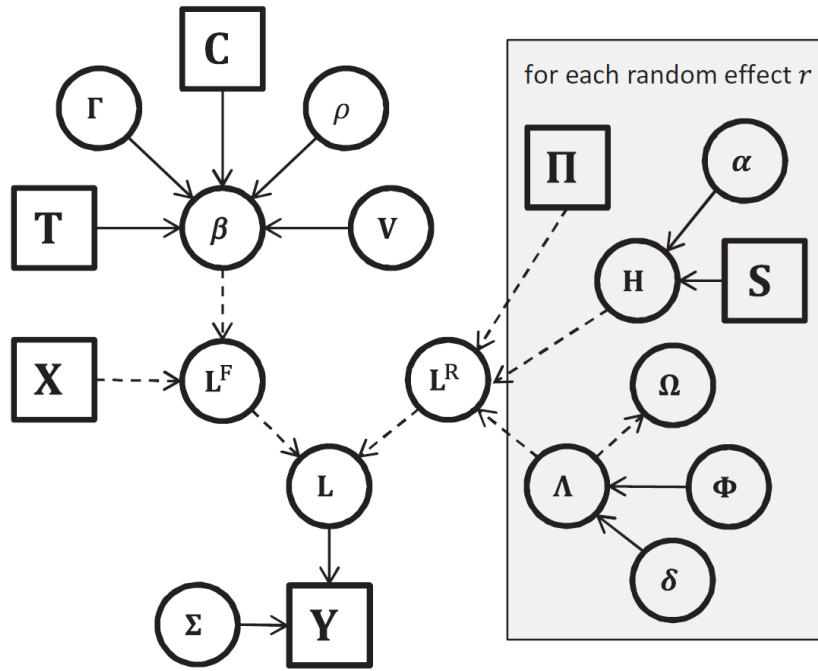
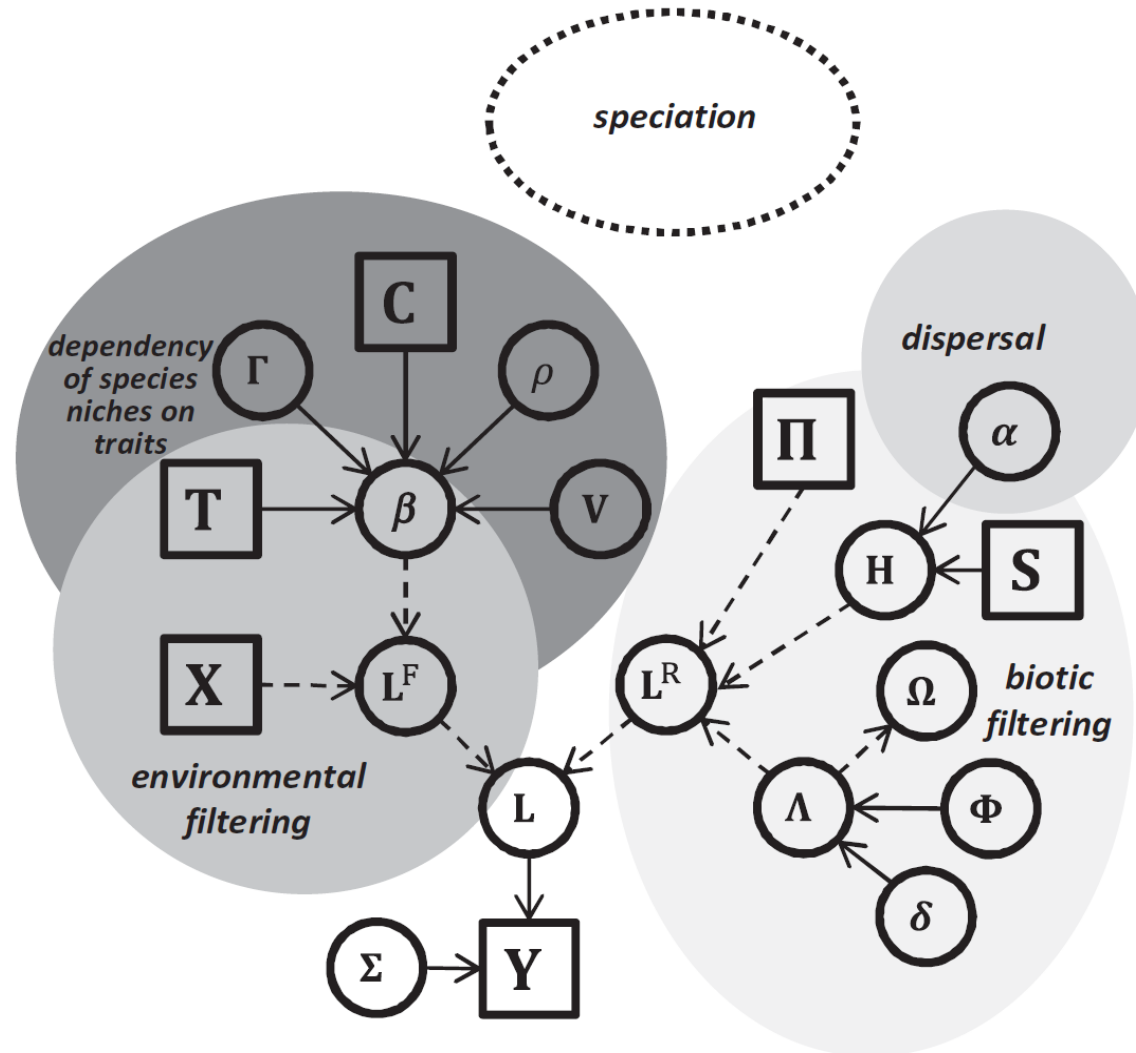


Table 4.4 *Parameters and their interpretations in the core HMSC model. The ‘Category’ column indicates whether the parameter is related to the fixed effect (F), random effect (R), or data model (D) part of HMSC. The parameters of the random effect part are defined separately for each random effect  $r$ .*

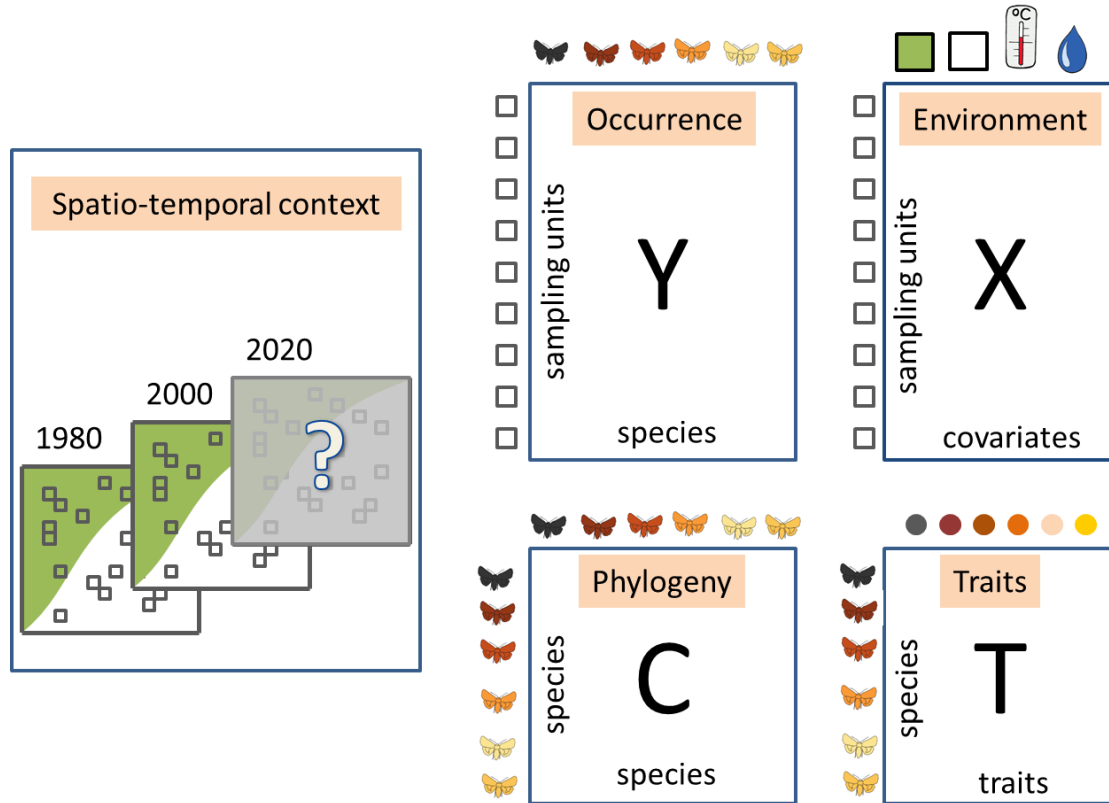
Category	Parameter	Type	Interpretation
F	$\mathbf{L}^F$ , element $L_{ij}^F$	$n \times n_s$ matrix	Linear predictor of fixed effects
F	$\mathbf{B}$ , element $\beta_{kj}$	$n_c \times n_s$ matrix	Species niches
F	$\mathbf{M}$ , element $\mu_{kj}$	$n_c \times n_s$ matrix	Expected species niches based on traits
F	$\rho$	scalar	Phylogenetic signal in species niches
F	$\mathbf{\Gamma}$ , element $\gamma_{kl}$	$n_c \times n_t$ matrix	Influence of traits on niches
F	$\mathbf{V}$ , element $V_{k_1 k_2}$	$n_c \times n_c$ matrix	Residual covariance of species niches
R	$\mathbf{L}^R$ , element $L_{ij}^R$	$n \times n_s$ matrix	Linear predictor of random effects
R	$\mathbf{H}$ , element $\eta_{uh}$	$n_u \times n_f$ matrix	Site loadings
R	$\boldsymbol{\alpha}$ , element $\alpha_h$	vector of length $n_f$	Spatial scale of site loadings
R	$\mathbf{\Lambda}$ , element $\lambda_{hj}$	$n_f \times n_s$ matrix	Species loadings
R	$\mathbf{\Omega}$ , element $\Omega_{j_1 j_2}$	$n_s \times n_s$ matrix	Species associations
R	$\mathbf{\Phi}$ , element $\phi_{hj}$	$n_f \times n_s$ matrix	Local shrinkage of species loadings
R	$\boldsymbol{\delta}$ , element $\delta_l$	vector of length $n_f$	Global shrinkage of species loadings
D	$\mathbf{L}$ , element $L_{ij}$	$n \times n_s$ matrix	Linear predictor
D	$\boldsymbol{\Sigma}$ , element $\sigma_j^2$	$n_s \times n_s$ diagonal matrix	Residual variance

# How does HMSC link to ecological theory?

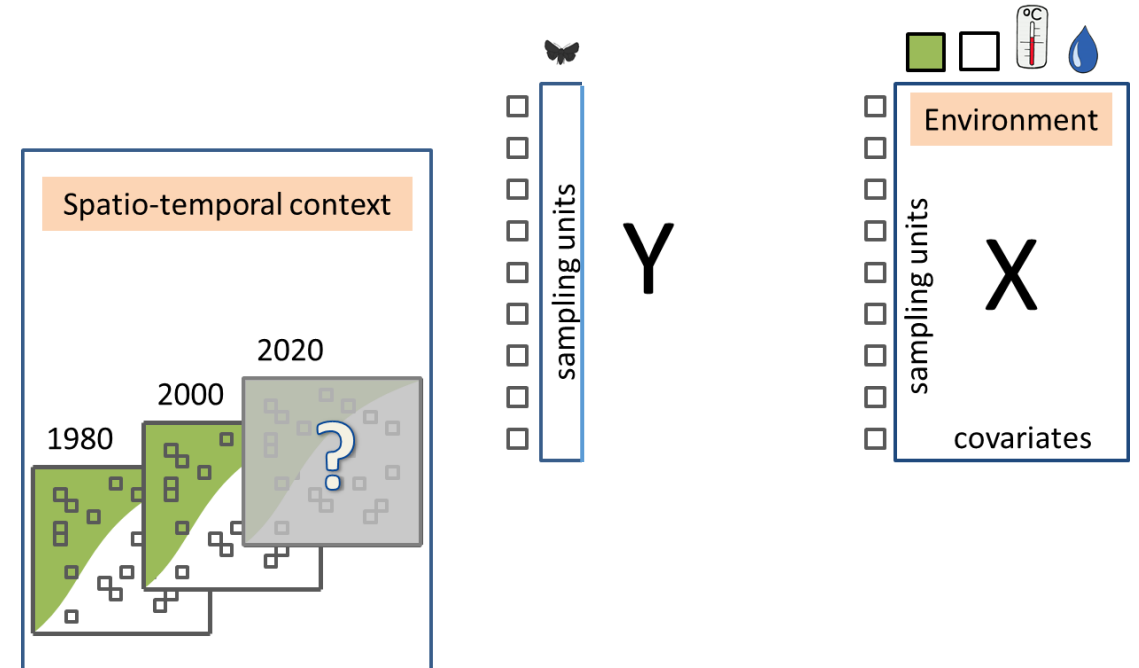


# **Closer look at the fixed effects**

## Full HMSC



## Single-species HMSC



Let's first look at the single-species case

## Back to basics: the linear model

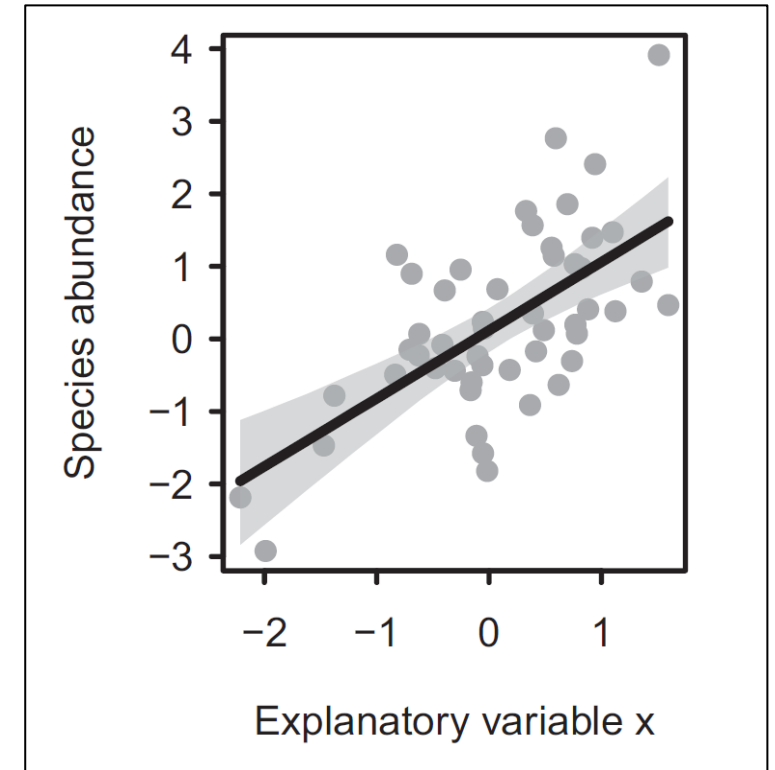
The linear model:  $y_i = \alpha + \beta x_i + \varepsilon_i$

Index for data points:  $i = 1, 2, 3, \dots, n$

Response (or dependent) variable:  $y$

Explanatory (or independent  
or predictor) variable:  $x$

Residual:  $\varepsilon_i \sim N(0, \sigma^2)$



Intercept:  $\alpha$

Slope:  $\beta$

## Several explanatory variables and the linear predictor

The linear model with two variables:  $y_i = \alpha + \beta_1 x_{i1} + \beta_1 x_{i2} + \varepsilon_i$

Can also be parameterized as:  $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$   
where  $x_{i1} = 1$  for all sampling units  $i$

Can be written more compactly as  $y_i = L_i + \varepsilon_i$  where

$$L_i = \sum_{k=1}^{n_c} \beta_k x_{ik}$$

is the linear predictor and  $n_c$  is the number of covariates (including the intercept)



## Continuous versus categorical predictors

In the basic linear model  $y_i = \alpha + \beta x_i + \varepsilon_i$   
 $x$  is a continuous explanatory variable (covariate)

Often  $x$  is a categorical explanatory variable (factor), e.g. habitat type classified as coniferous forest, broadleaved forest, or mixed forest.

This can be incorporated as:

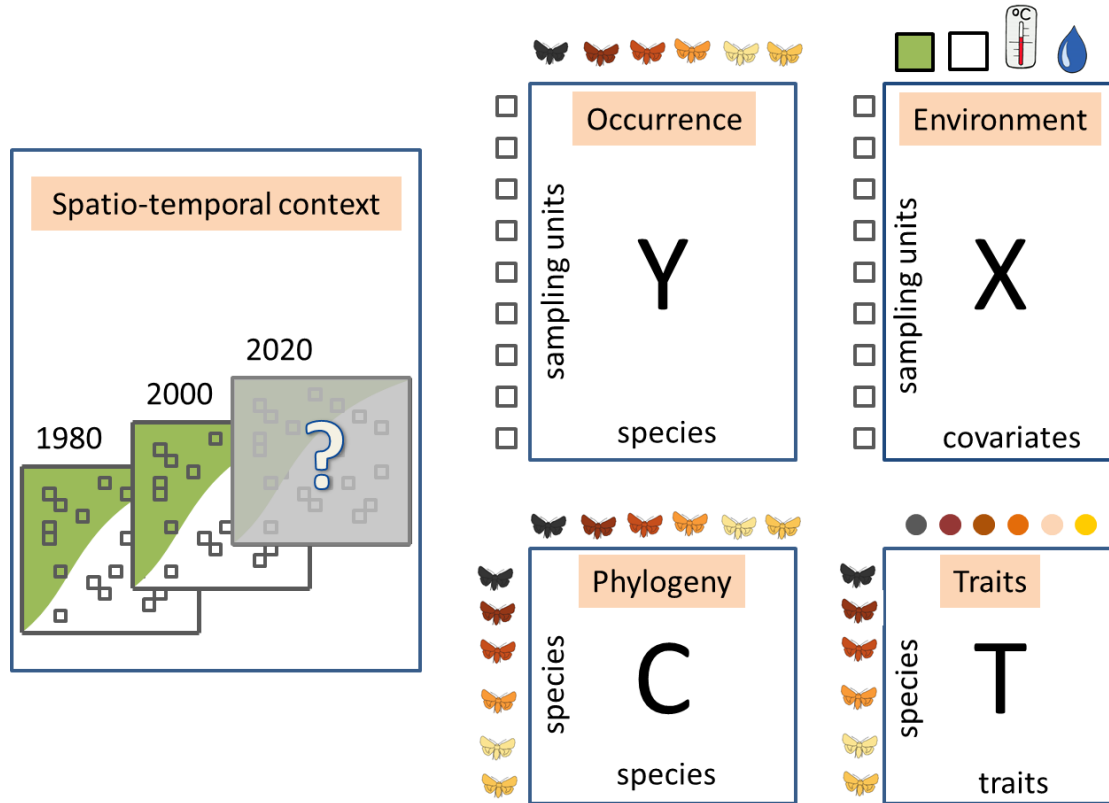
$$L_i = \sum_{k=1}^{n_c} \beta_k x_{ik}$$

$x_{i1} = 1$  for all sampling units

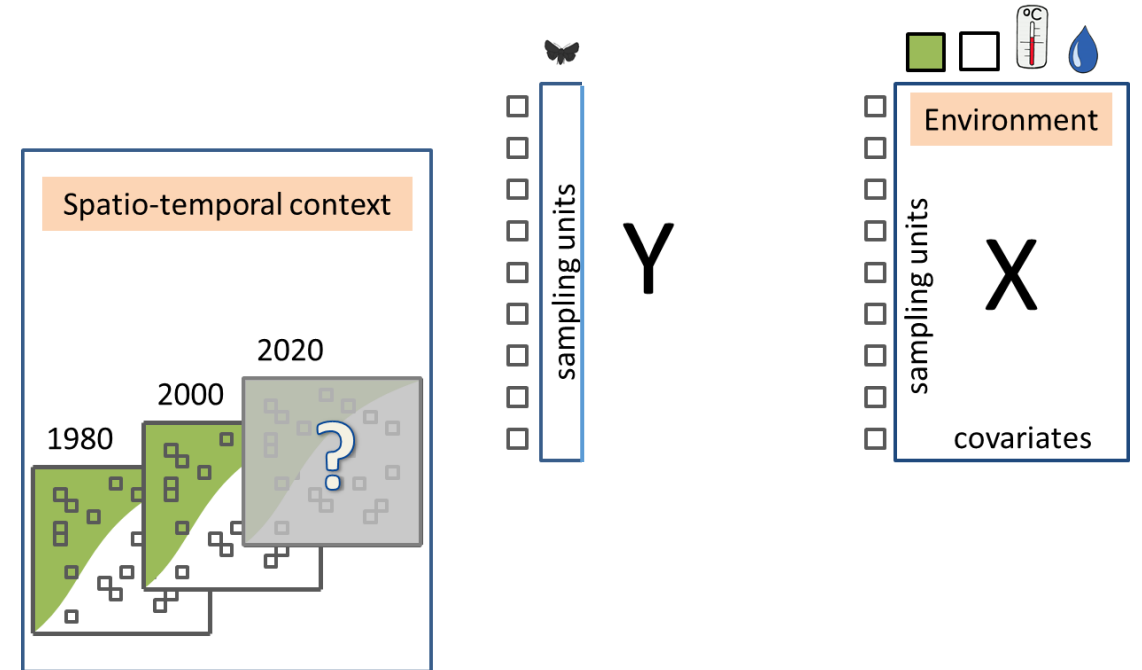
$x_{i2} = 1$  if  $i$  is in broadleaved forest, otherwise  $x_{i2} = 0$

$x_{i3} = 1$  if  $i$  is in mixed forest, otherwise  $x_{i3} = 0$

## Full HMSC

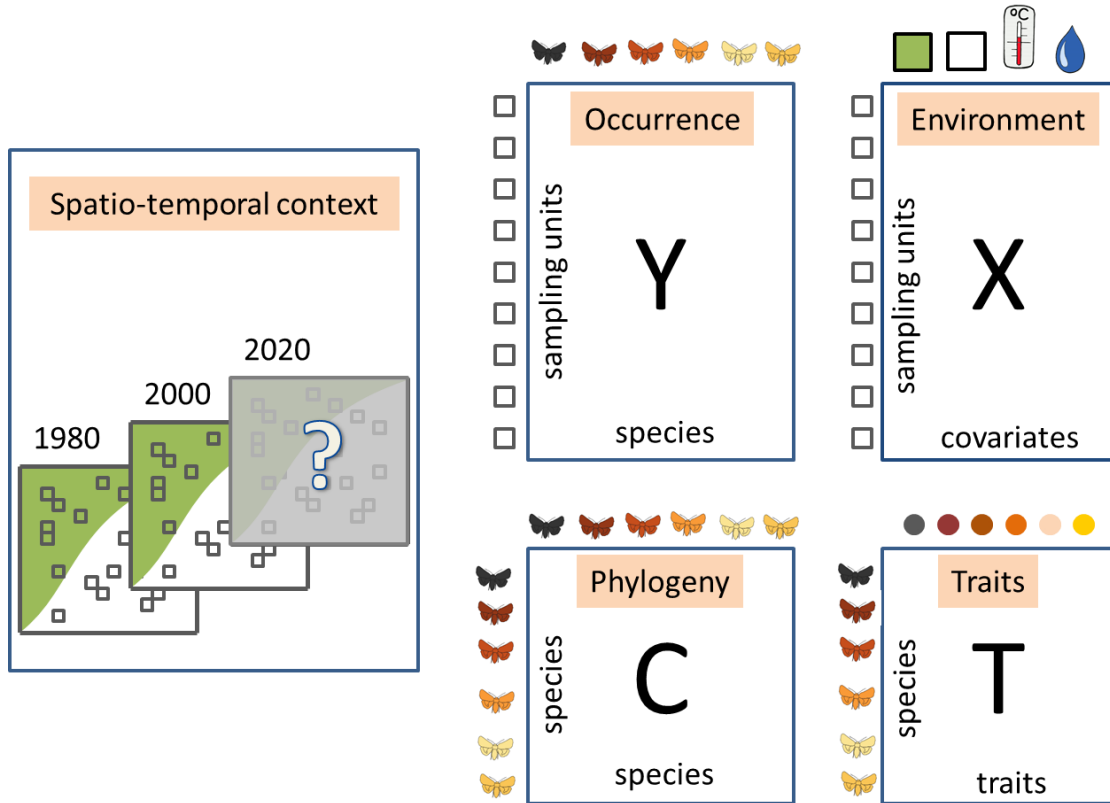


## Single-species HMSC



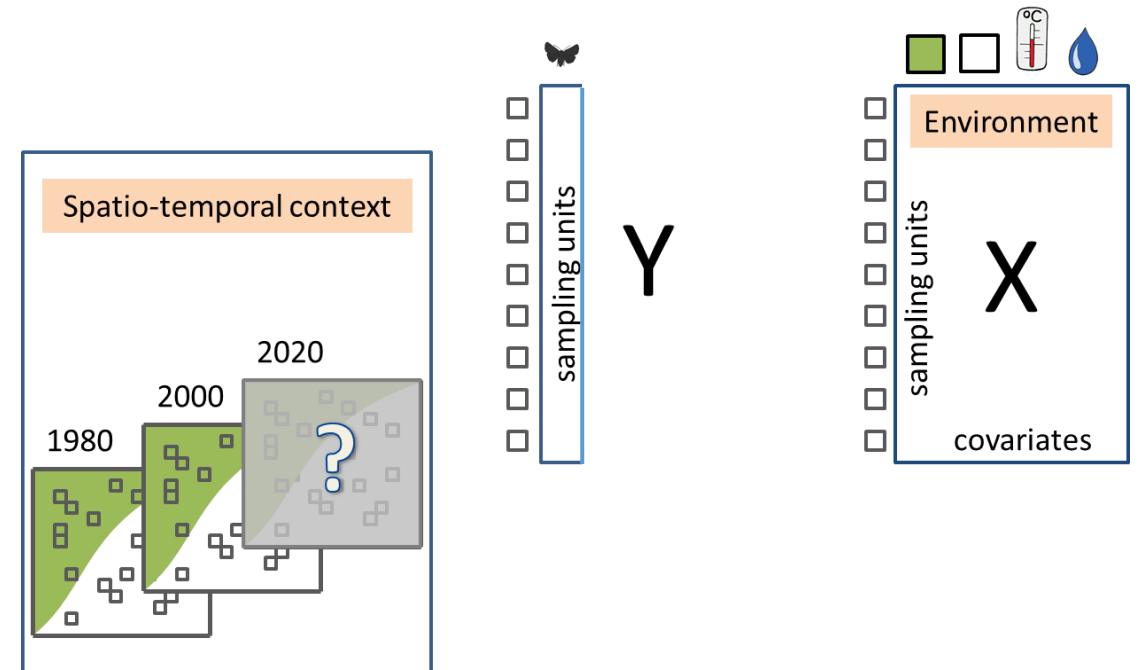
Let's now look at the multi-species case

## Full HMSC



$$L_{ij} = \sum_{k=1}^{n_c} \beta_{kj} x_{ik}$$

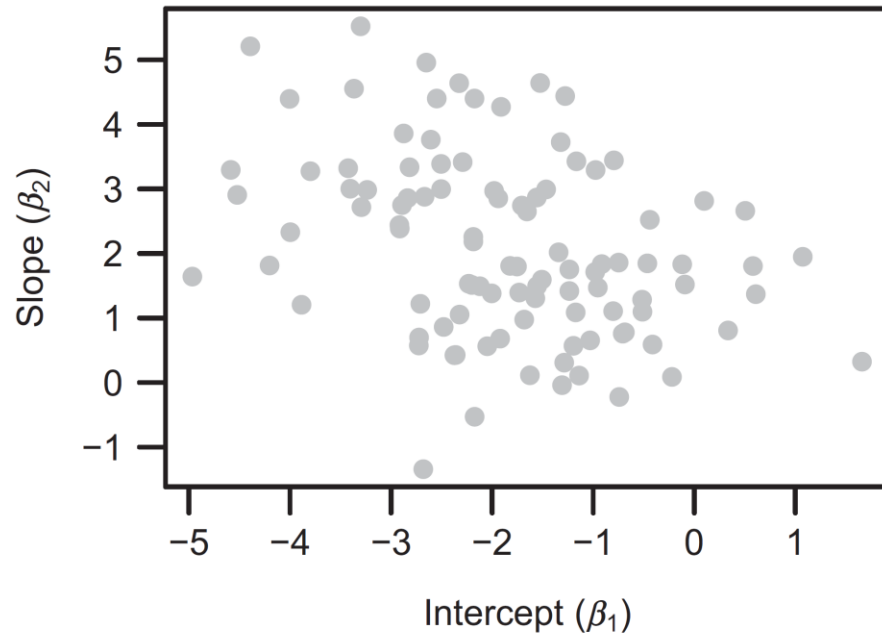
## Single-species HMSC



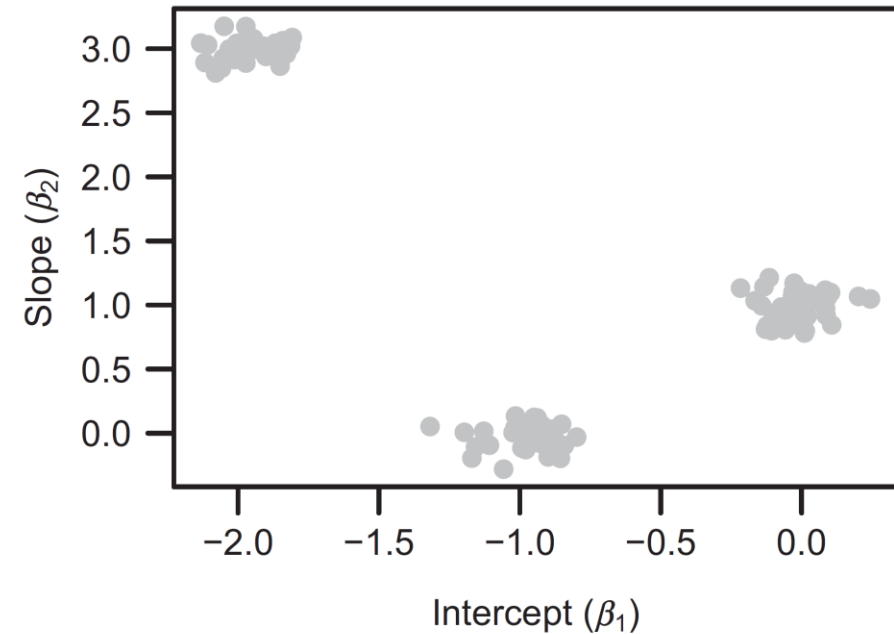
$$L_i = \sum_{k=1}^{n_c} \beta_k x_{ik}$$

# Variation in species niches among the species

(A)



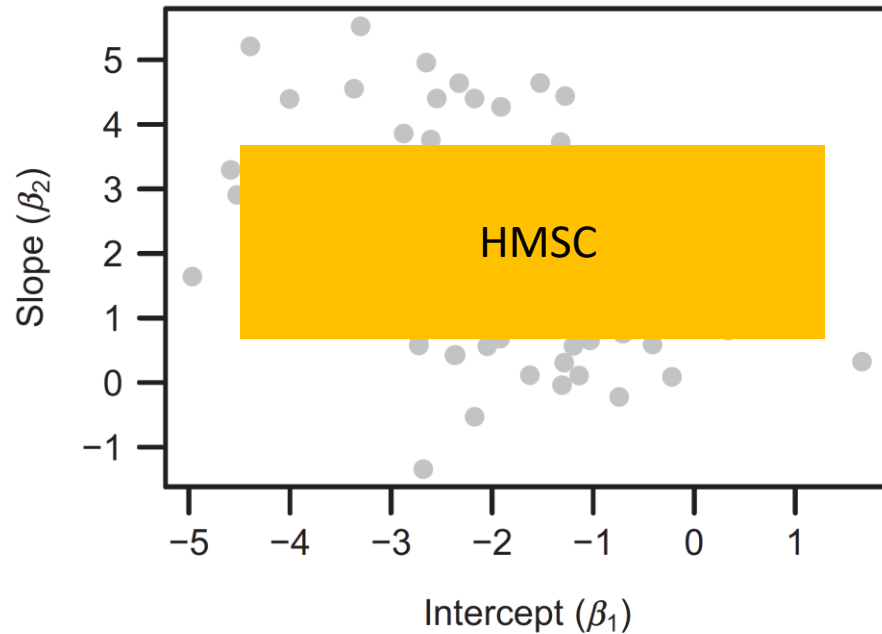
(B)



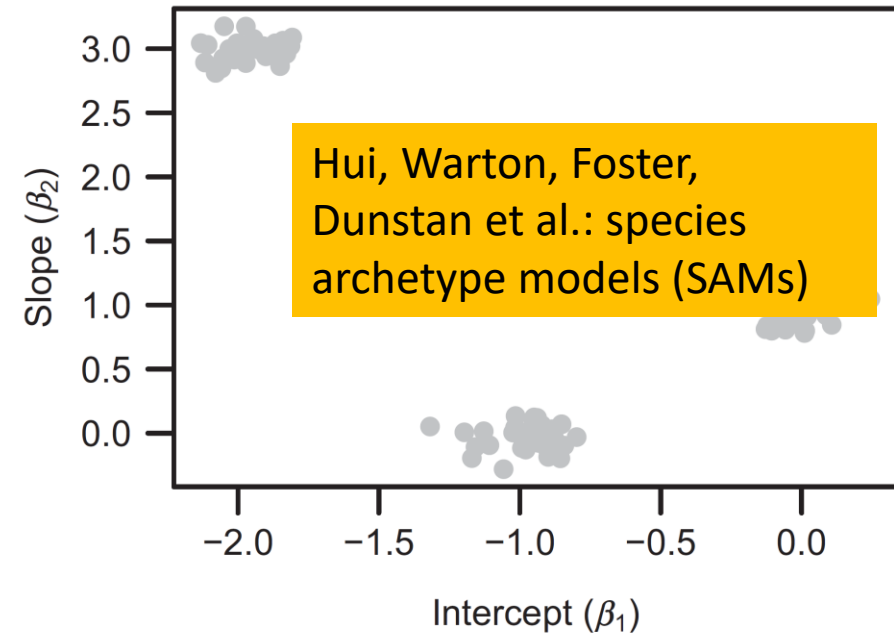
*Figure 6.1* Illustration of variation in species niches. In both panels, each dot corresponds to one species in a community of 100 species. In panel A, there is continuous variation among species niches, whereas in panel B the species niches form three clusters.

# Variation in species niches among the species

(A)



(B)

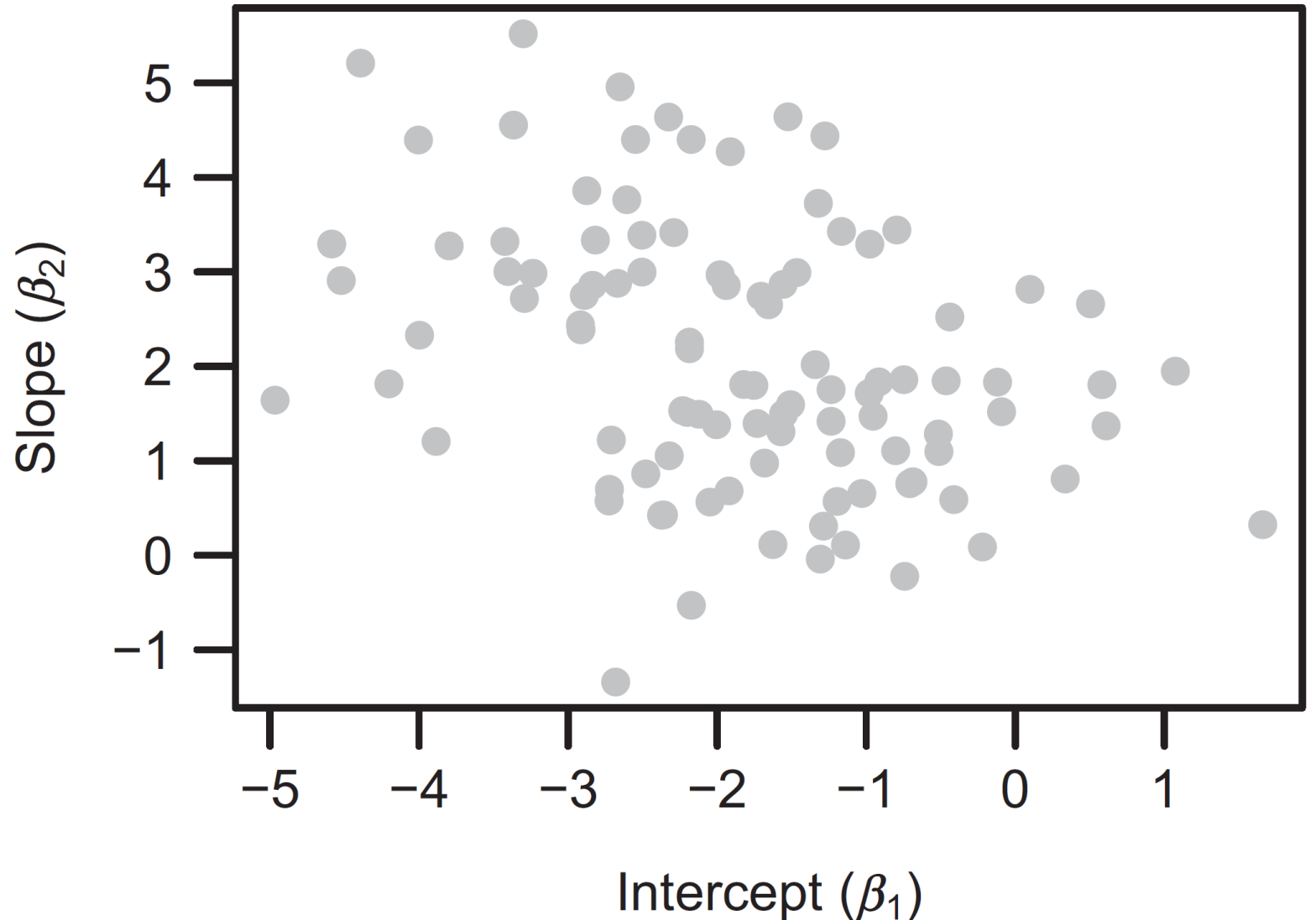


*Figure 6.1* Illustration of variation in species niches. In both panels, each dot corresponds to one species in a community of 100 species. In panel A, there is continuous variation among species niches, whereas in panel B the species niches form three clusters.

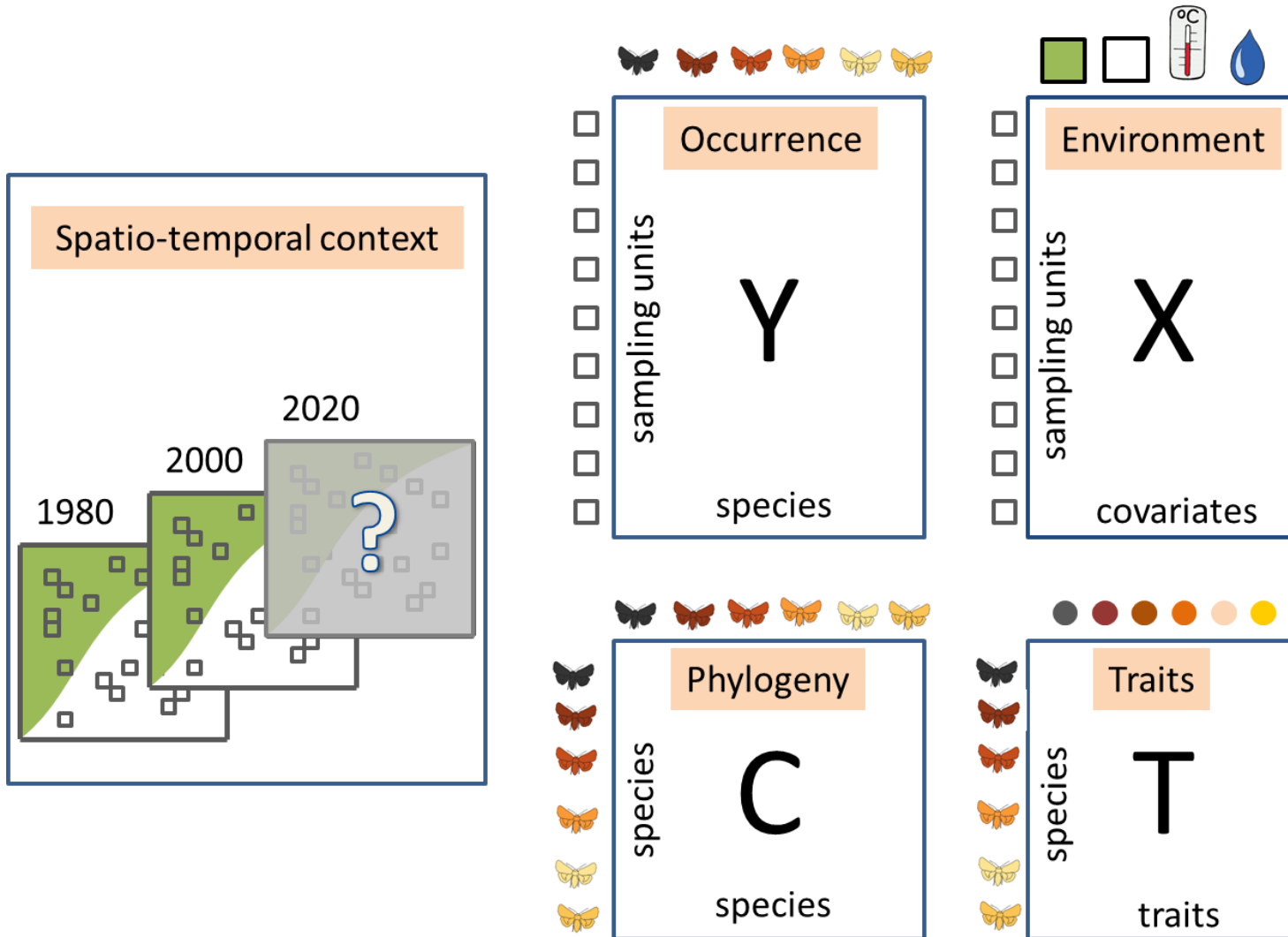
## Variation in species niches among the species

A simple statistical model for variation in species niches:

$$\boldsymbol{\beta}_{\cdot j} \sim N(\boldsymbol{\mu}, \mathbf{V})$$



# How to utilize data on species traits?



# Modelling the influence of species traits on their niches

Species-specific expected value:

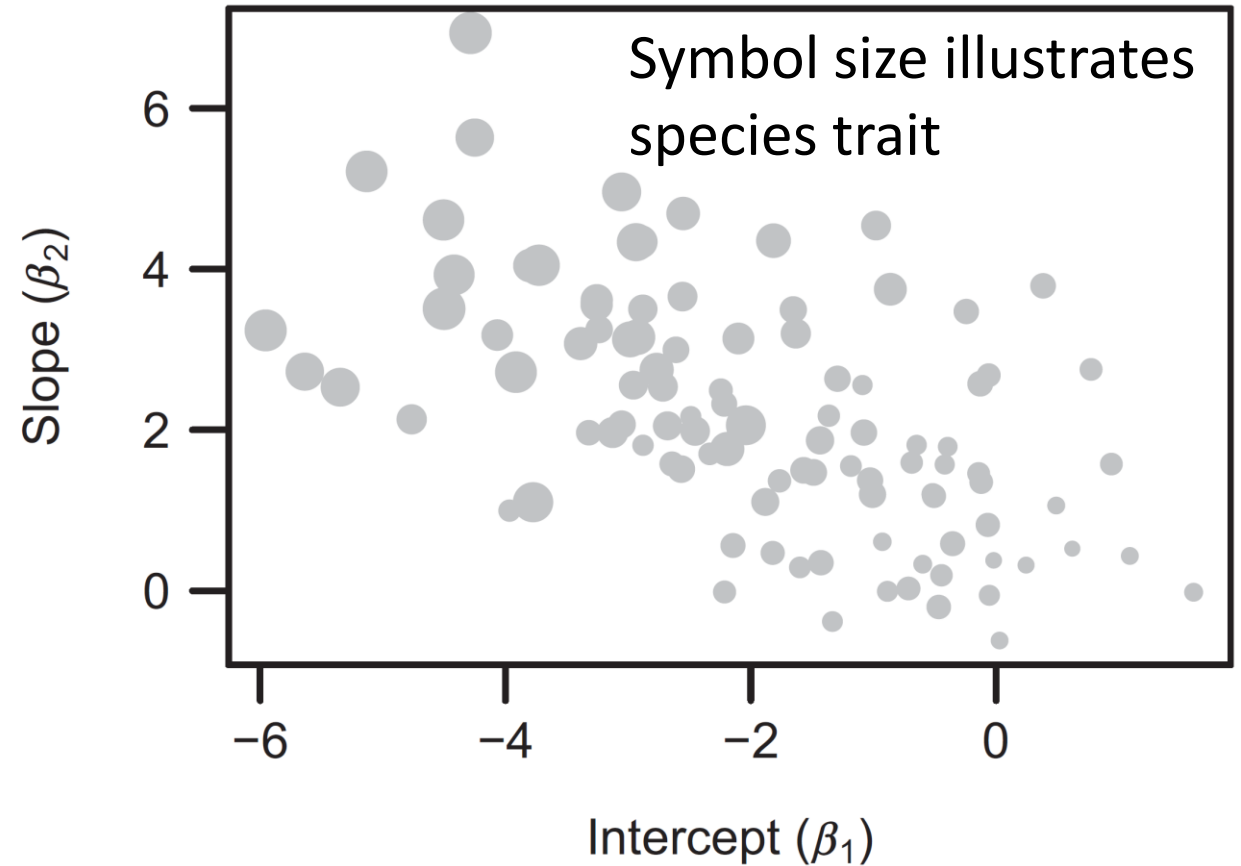
$$\boldsymbol{\beta}_{\cdot j} \sim N(\boldsymbol{\mu}_{\cdot j}, \mathbf{V})$$

Modelled as regression to species traits:

$$\mu_{kj} = \sum_{l=1}^{n_t} t_{jl} \gamma_{kl}$$

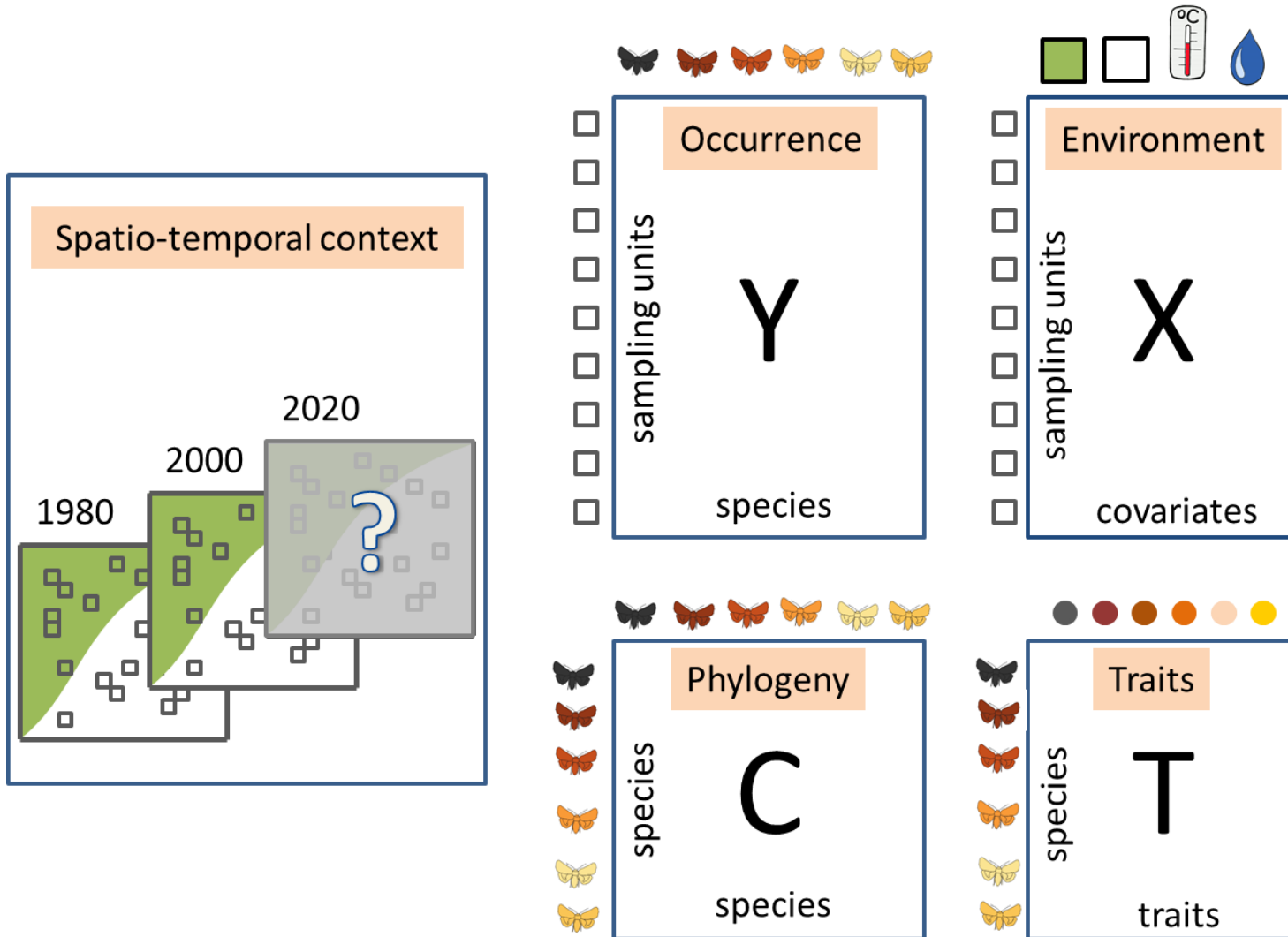
The trait  $l$  of species  $j$

The influence of trait  $l$  on how the species is expected to respond to covariate  $k$

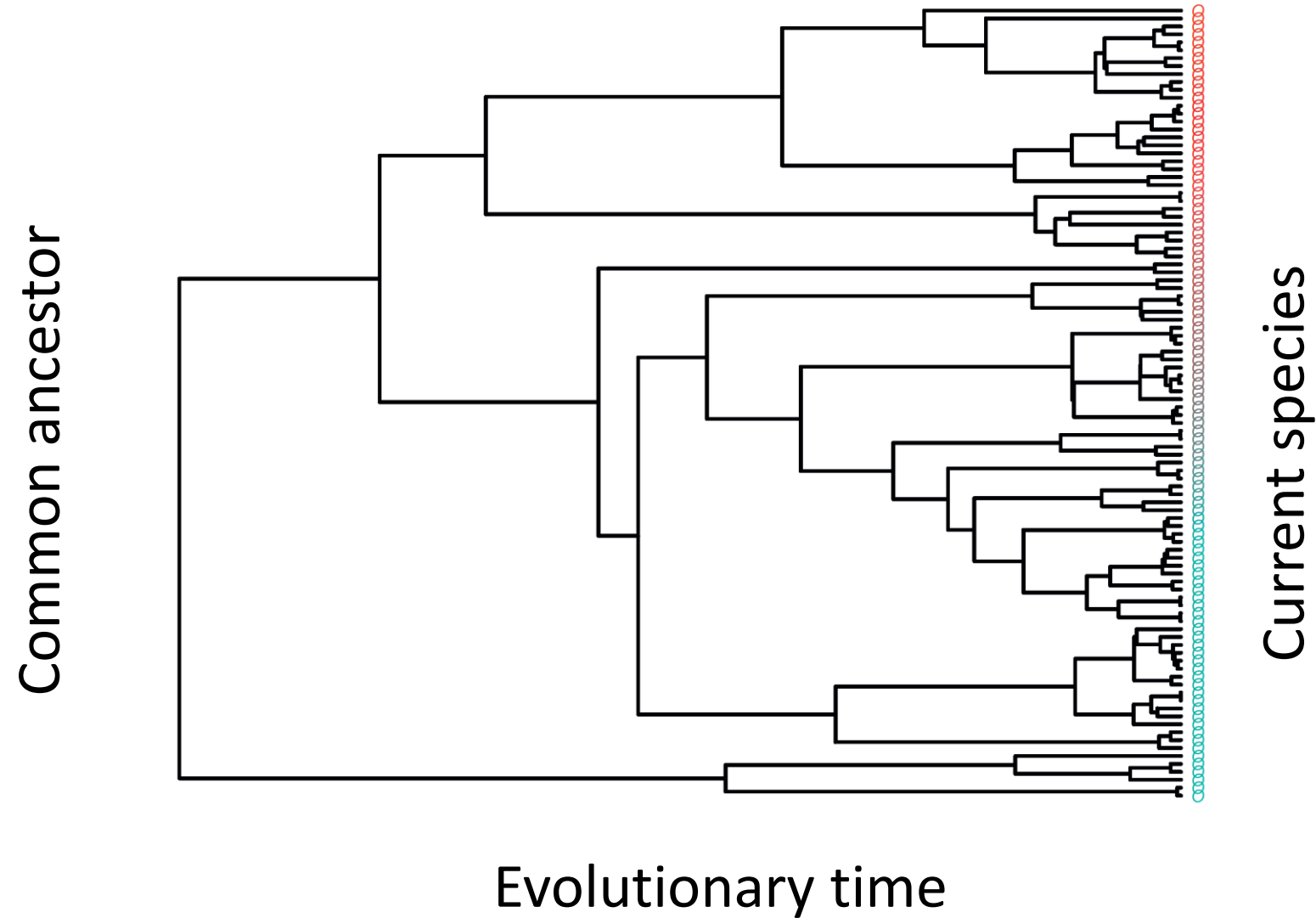




# How to utilize data on phylogenetic relationships?

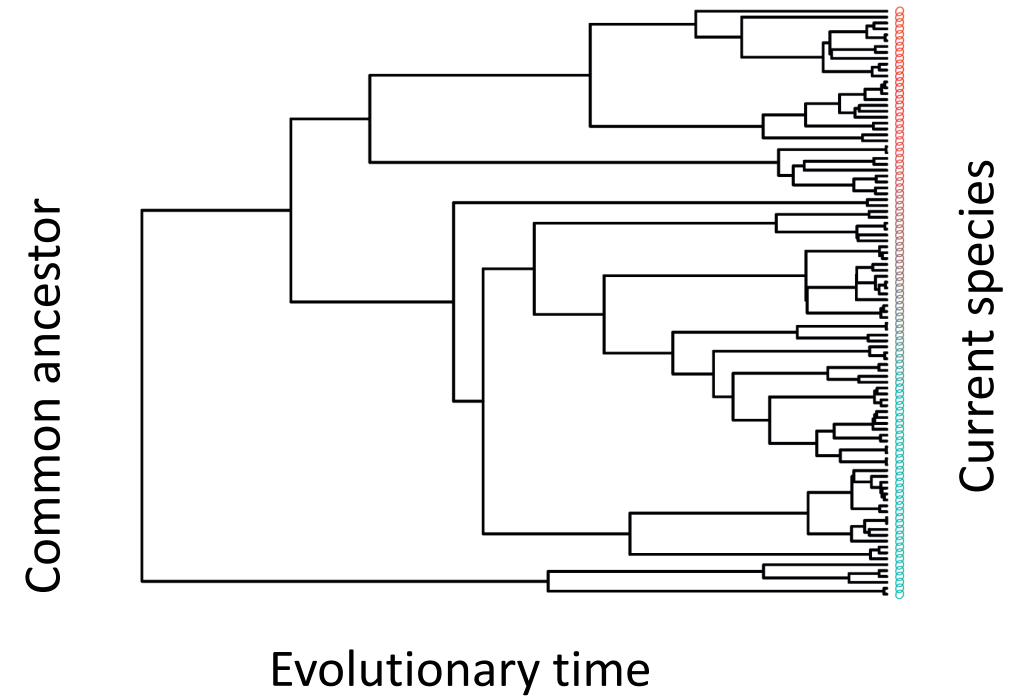
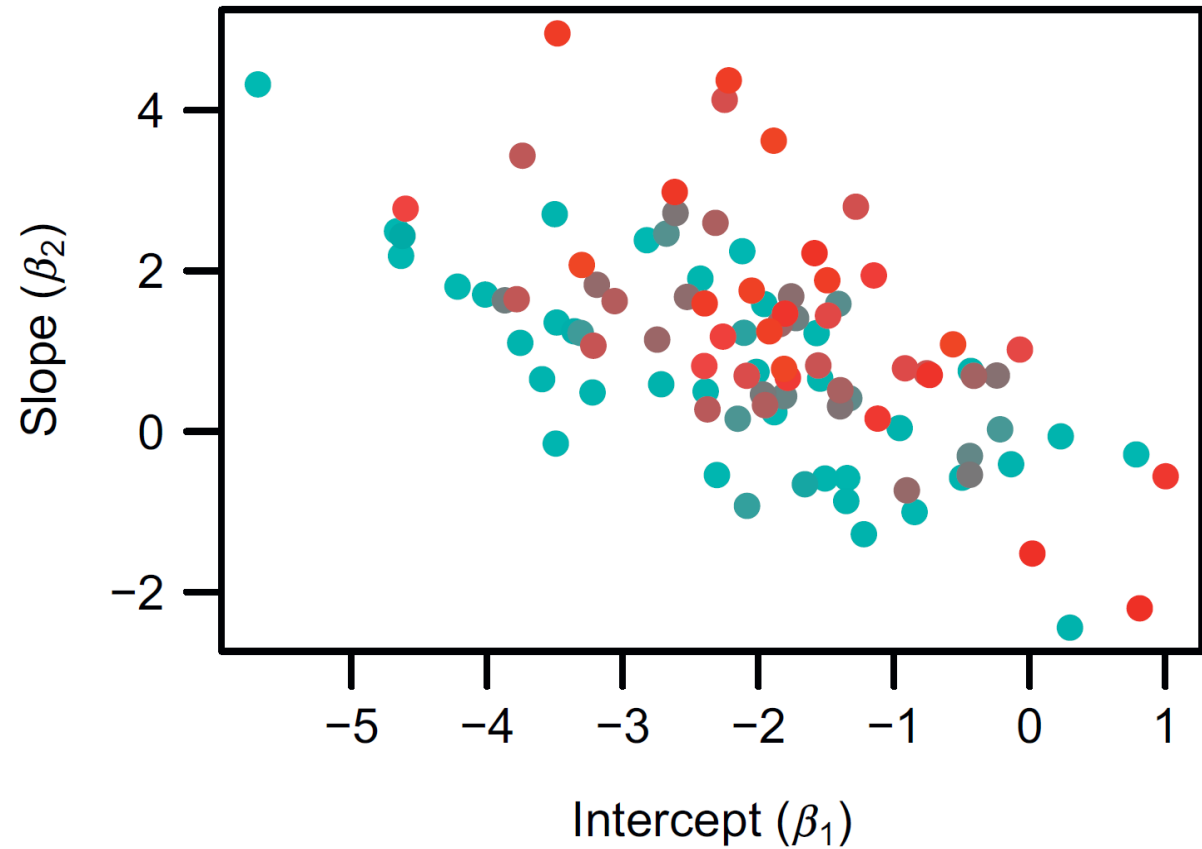


# Modelling the influence of phylogeny on species niches



# Modelling the influence of phylogeny on species niches

Illustration of systematic variation in species niches across the phylogeny



# Modelling the influence of phylogeny on species niches

The basic model in matrix notation:

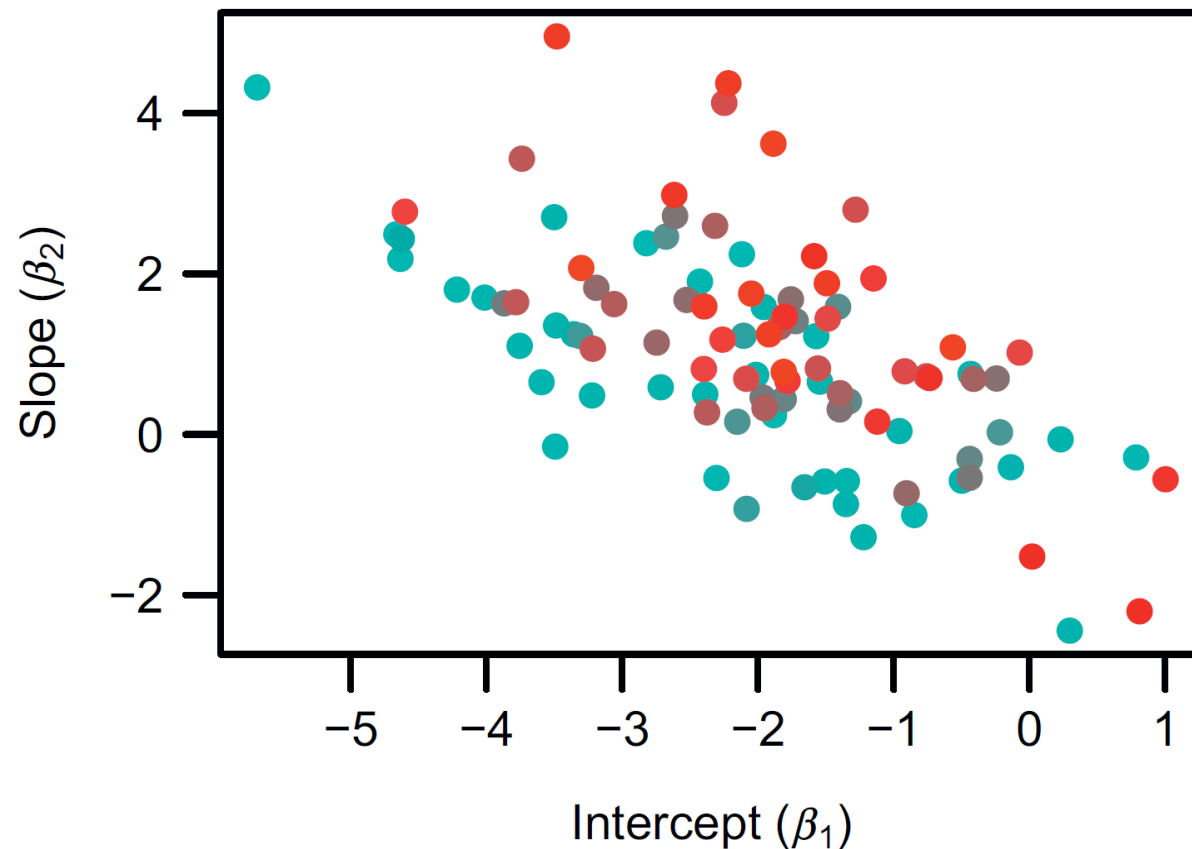
$$\boldsymbol{\beta}_{\cdot j} \sim N(\boldsymbol{\mu}, \mathbf{V})$$

$$\text{vec}(\mathbf{B}) \sim N(\text{vec}(\mathbf{M}), \mathbf{I} \otimes \mathbf{V})$$

Phylogenetically structured model  
in matrix notation:

$$\text{vec}(\mathbf{B}) \sim N(\text{vec}(\mathbf{M}), \mathbf{W} \otimes \mathbf{V})$$

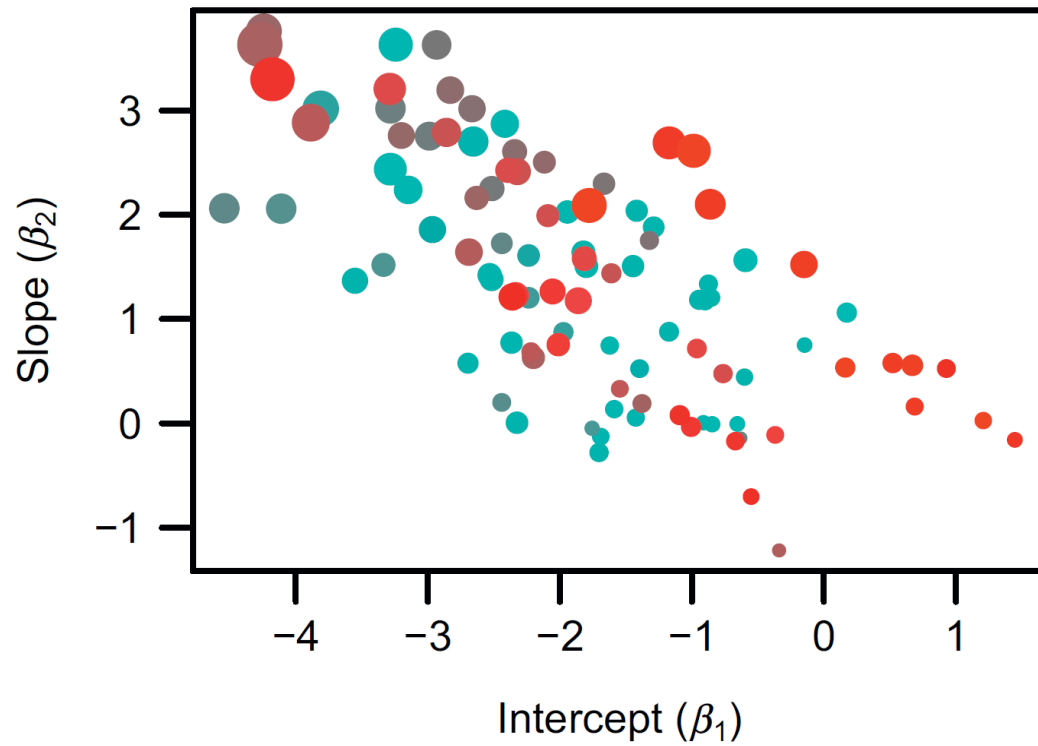
$$\mathbf{W} = \rho \mathbf{C} + (1 - \rho) \mathbf{I}$$



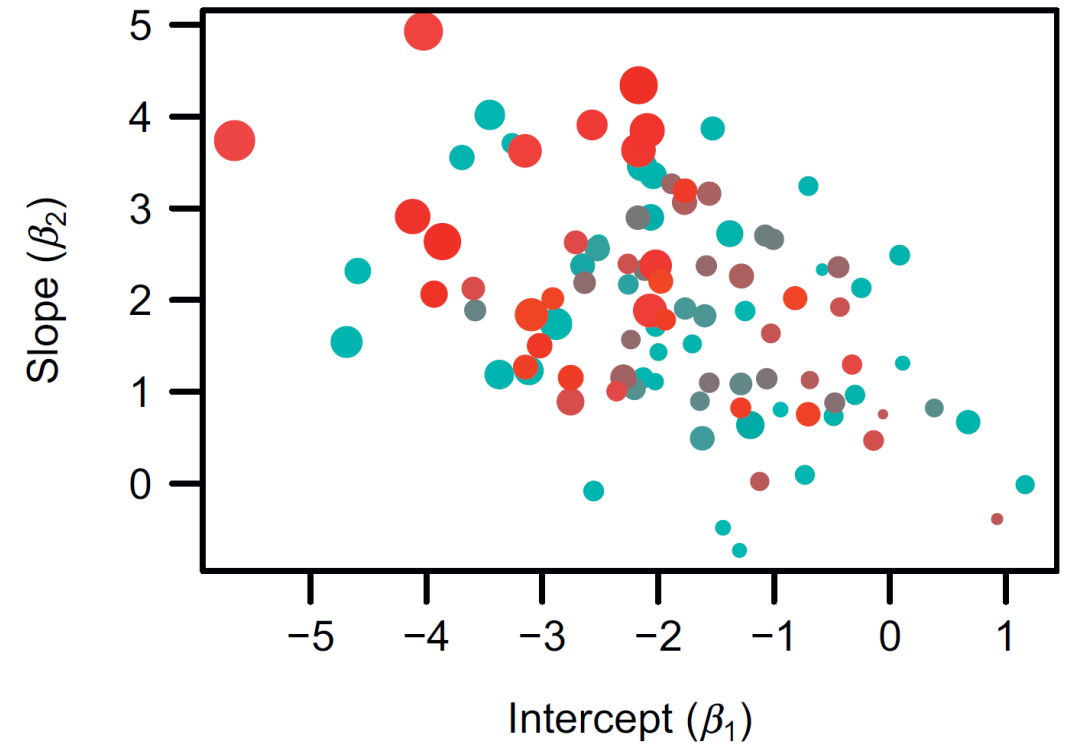
The parameter  $\rho$  measures the strength of phylogenetic signal in species niches

# Modelling the joint influence of species traits and phylogeny on species niches

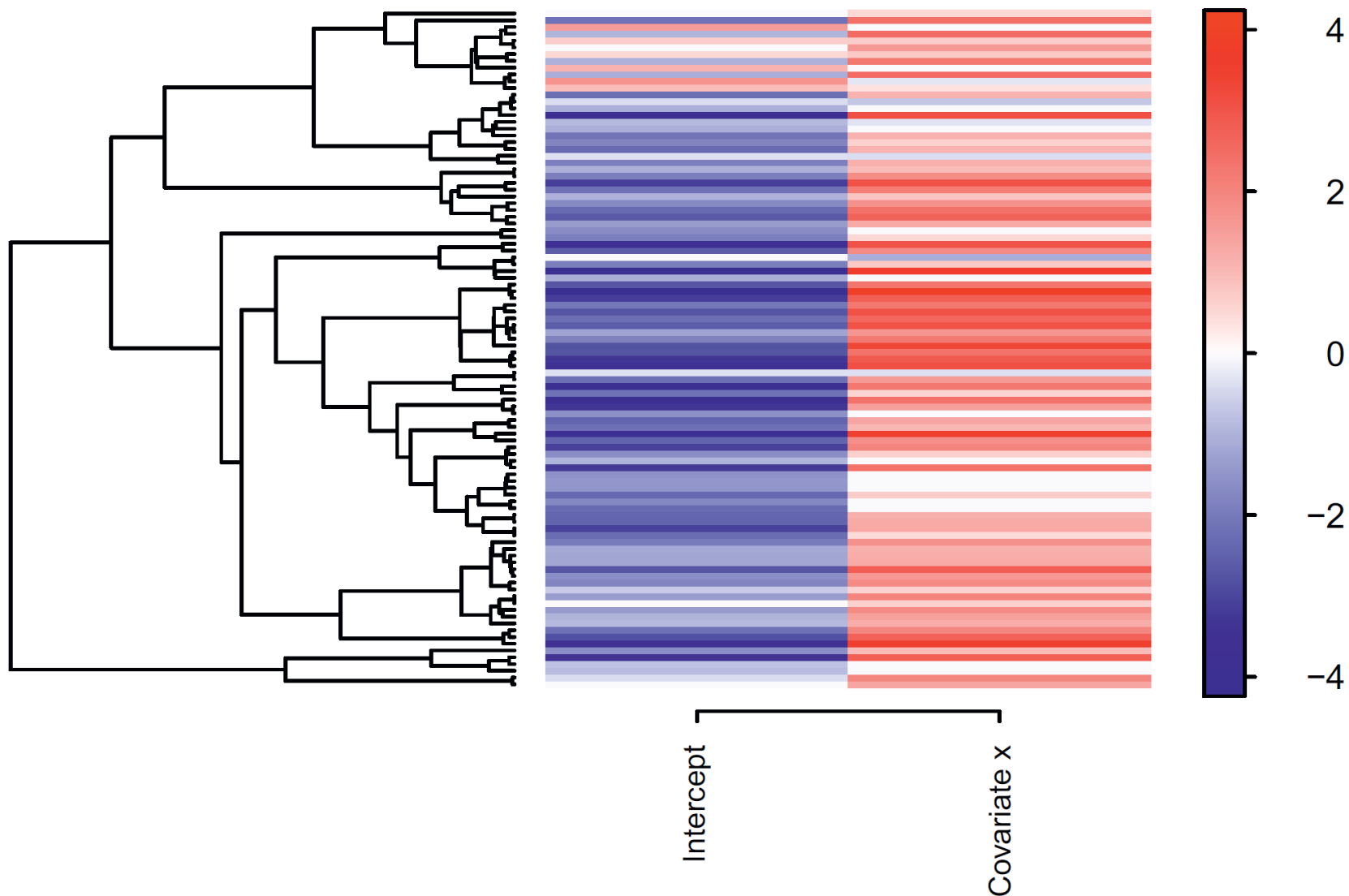
(A)



(B)

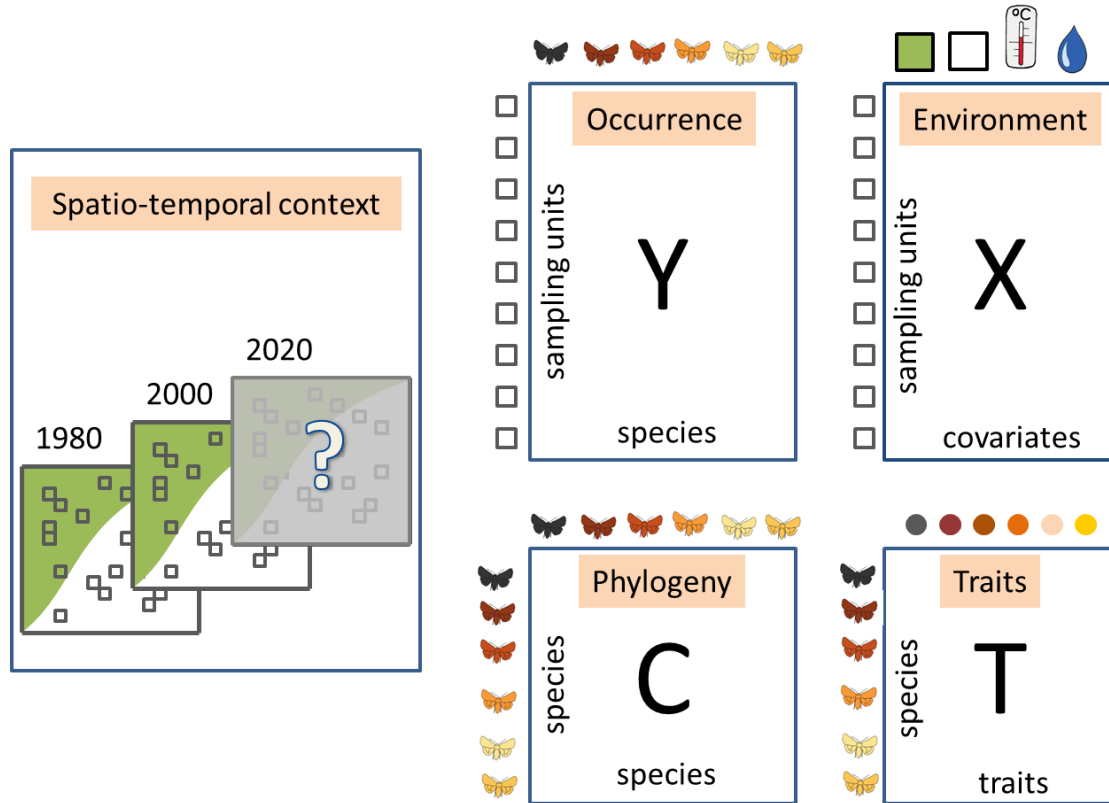


# Distribution of species niches, as estimated with HMSC (BetaPlot)

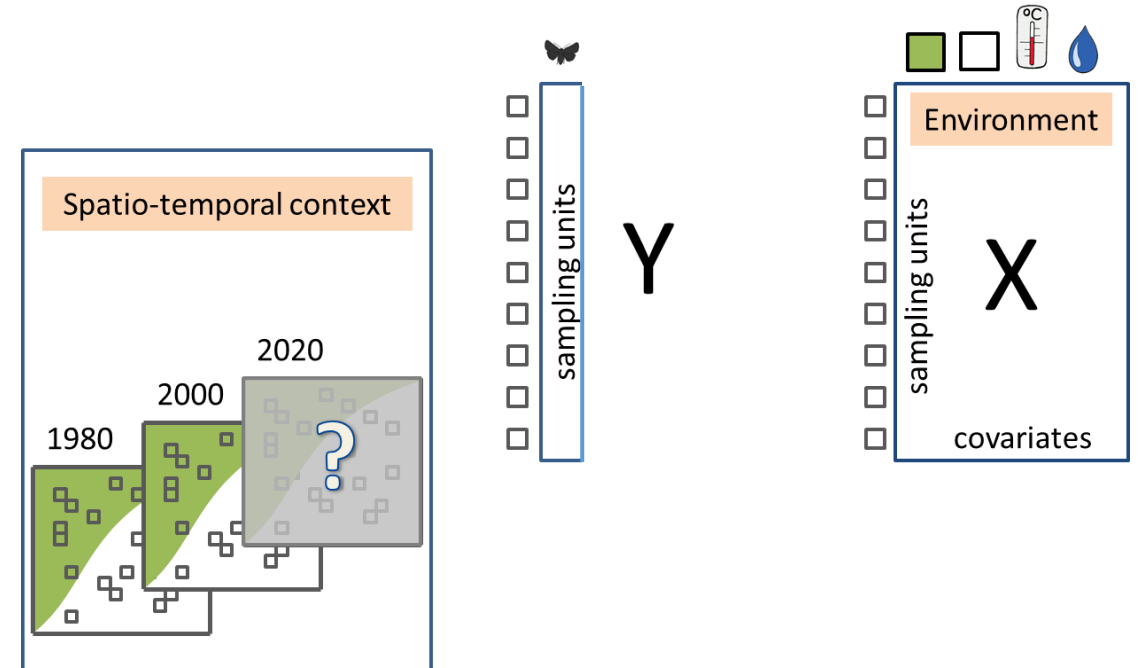


# **Closer look at the random effects**

## Full HMSC



## Single-species HMSC

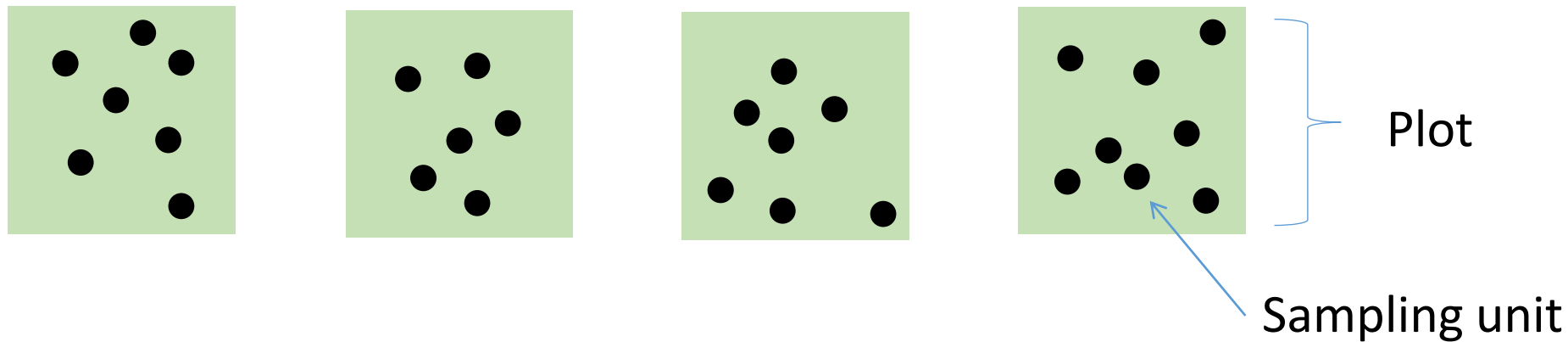


Let's first look at the single-species case



## Mixed models: fixed effects and random effects

**Hierarchical  
study design:**



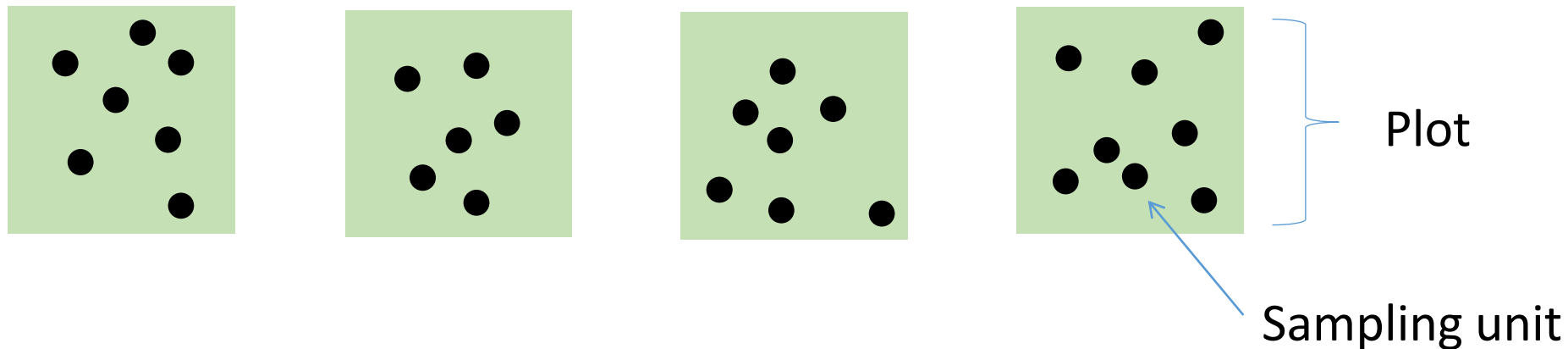
Linear model with fixed effects only:  $y_i = L_i + \varepsilon_i$

$$L_i = \sum_{k=1}^{n_c} \beta_k x_{ik}$$

$$\begin{array}{c} \text{iid} \\ \varepsilon_i \sim N(0, \sigma^2) \end{array}$$

## Mixed models: fixed effects and random effects

**Hierarchical  
study design:**



Linear model with fixed and random effects:  $y_i = L_i + a_{p(i)} + \varepsilon_i$

$$L_i = \sum_{k=1}^{n_c} \beta_k x_{ik}$$

$$\text{iid} \\ a_p \sim N(0, \sigma_P^2)$$

$$\text{iid} \\ \varepsilon_i \sim N(0, \sigma^2)$$

# Mixed models: fixed effects and random effects

Spatial study  
design:

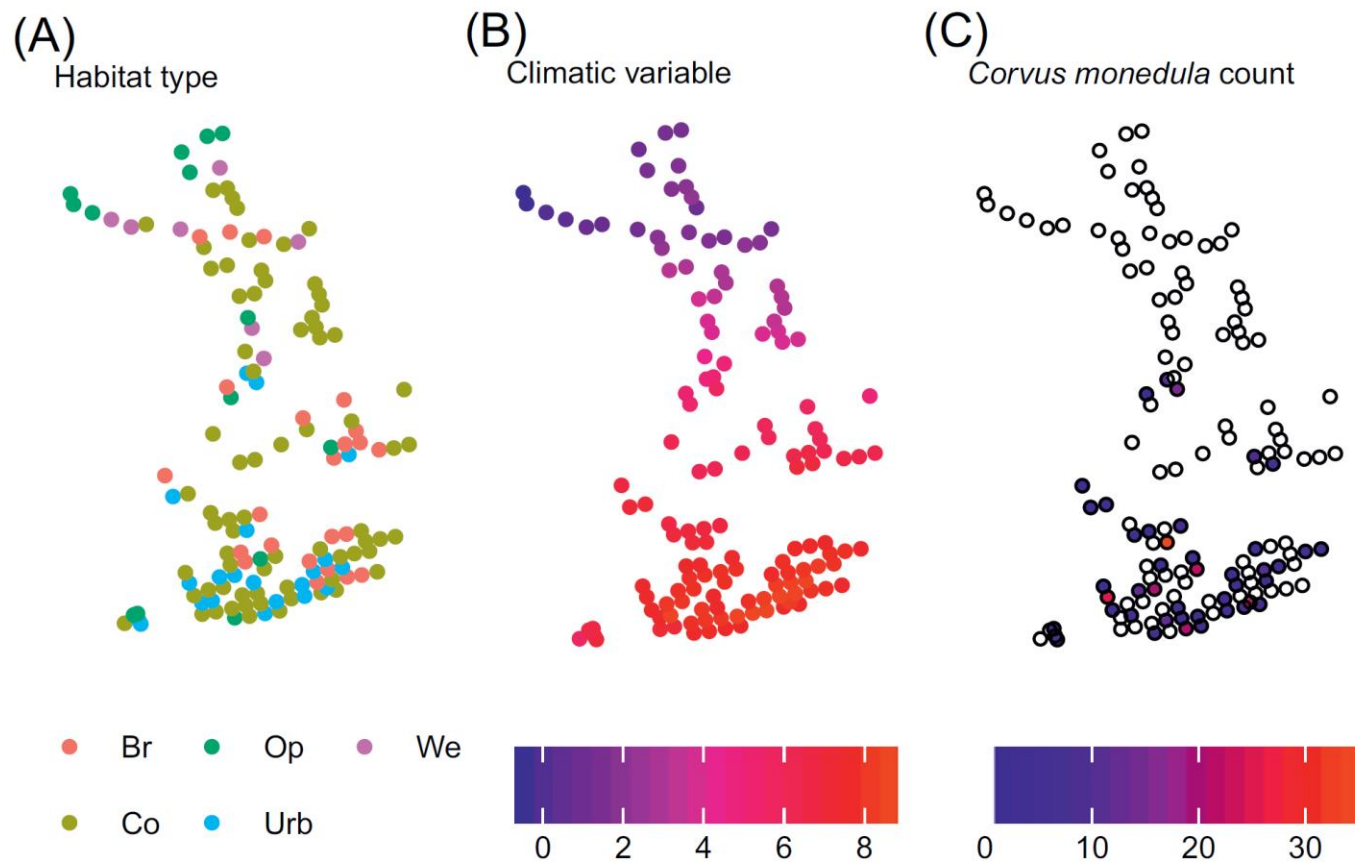


Figure 5.9 An illustration of environmental and species data used in this example. The panels show spatial variation in habitat type (A), climatic conditions (B), and the counts of the target species across Finland (C).

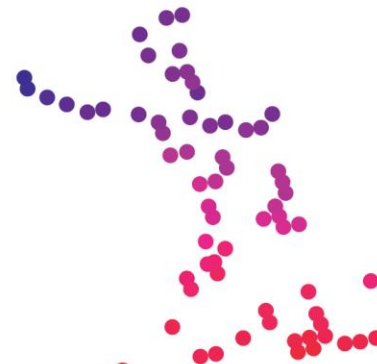
# Mixed models: fixed effects and random effects

Spatial study  
design:

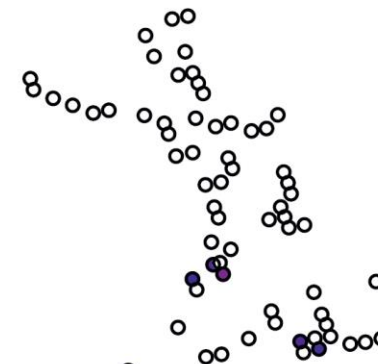
(A)  
Habitat type



(B)  
Climatic variable



(C)  
*Corvus monedula* count



Linear model without spatial structure:  $y_i = L_i + \varepsilon_i$

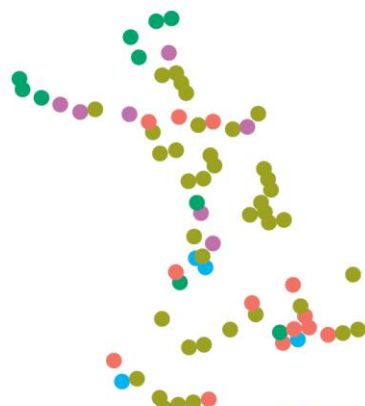
$$L_i = \sum_{k=1}^{n_c} \beta_k x_{ik}$$

$$\begin{array}{c} \text{iid} \\ \varepsilon_i \sim N(0, \sigma^2) \end{array}$$

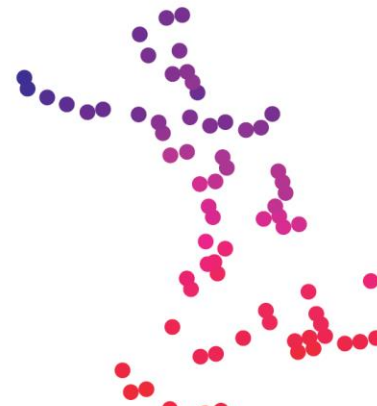
# Mixed models: fixed effects and random effects

Spatial study  
design:

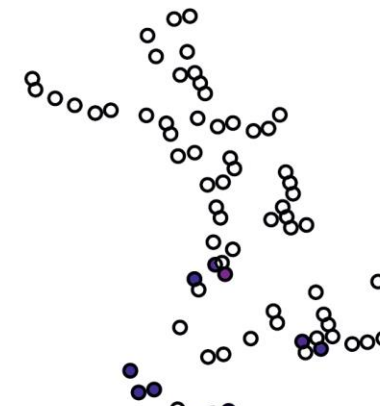
(A)  
Habitat type



(B)  
Climatic variable



(C)  
*Corvus monedula* count



Linear model with spatial structure:

$$y_i = L_i + a_i + \varepsilon_i$$

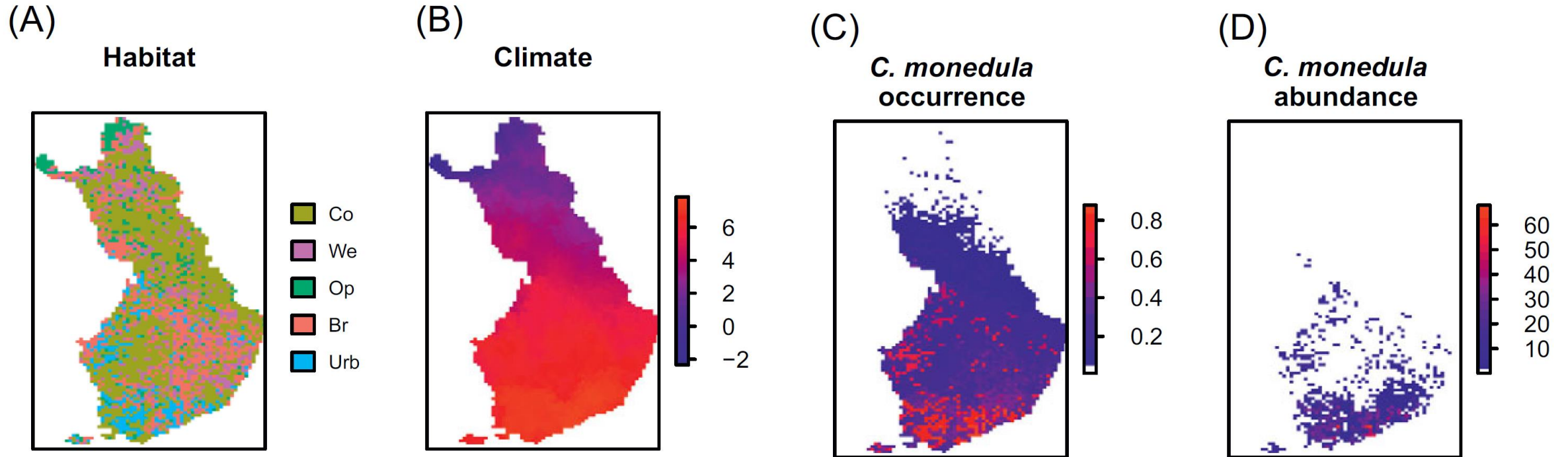
$$L_i = \sum_{k=1}^{n_c} \beta_k x_{ik}$$

$$a_i \sim N(0, \sigma_S^2)$$

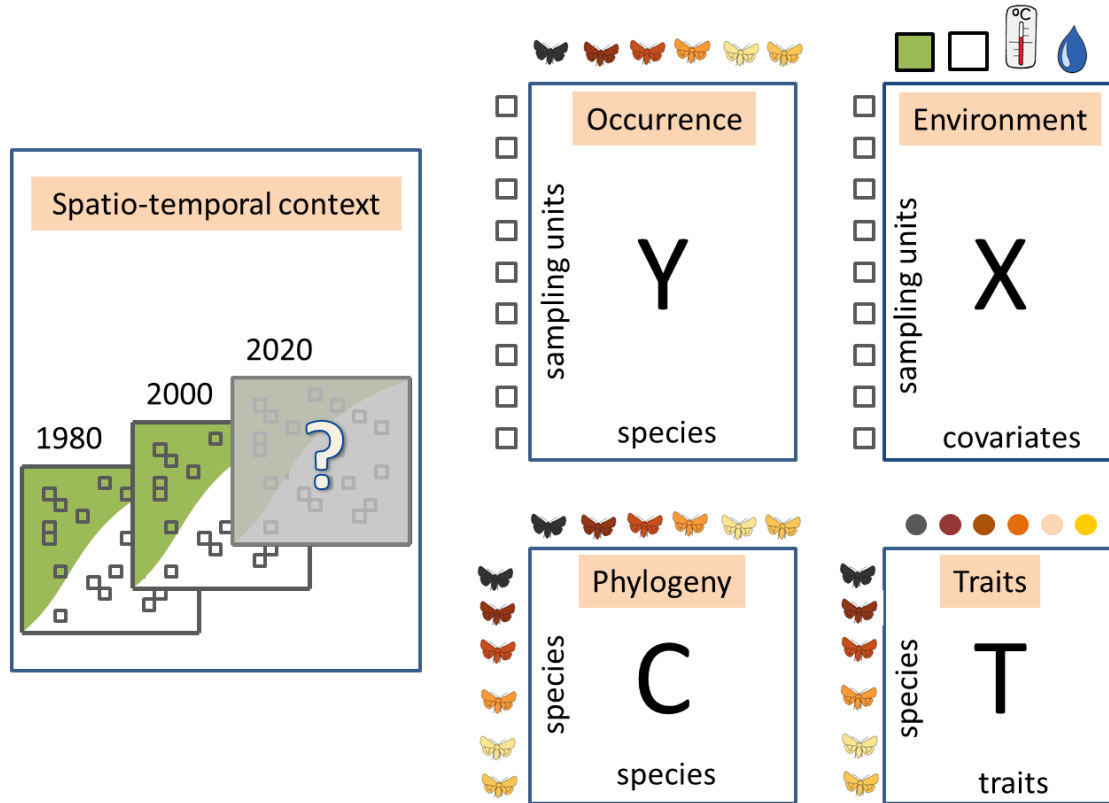
$$\text{Cov}(a_i, a_j) = \sigma_S^2 \exp(-d_{ij}/\alpha)$$

$$\begin{array}{c} \text{iid} \\ \varepsilon_i \sim N(0, \sigma^2) \end{array}$$

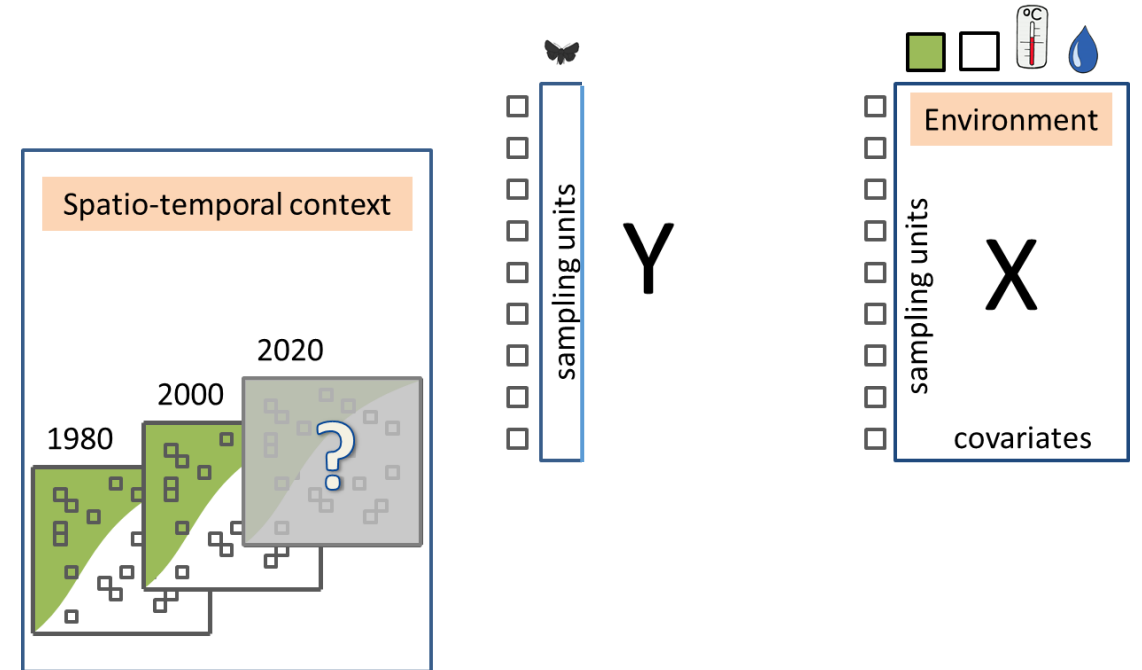
# Fitting a spatial model enables using spatial information when generating predictions



## Full HMSC



## Single-species HMSC



Let's now look at the multi-species case



# Occurrence and co-occurrence probabilities

Capercaillie  
(*Tetrao urogallus*)



Source: Wikimedia

White-backed woodpecker  
(*Dendrocopos leucotos*)



Source: Wikimedia



## Occurrence probabilities



$$p_1 = 0.5$$



$$p_2 = 0.5$$

## Co-occurrence probabilities



+



$$q_{11} = \dots$$



+



$$q_{10} = \dots$$



+



$$q_{01} = \dots$$



+



$$q_{00} = \dots$$

## Occurrence probabilities



$$p_1 = 0.5$$



$$p_2 = 0.5$$

## Co-occurrence probabilities



+



$$q_{11} = 0.25$$



+



$$q_{10} = 0.25$$



+



$$q_{01} = 0.25$$



+



$$q_{00} = 0.25$$

No association

## Occurrence probabilities



$$p_1 = 0.5$$



$$p_2 = 0.5$$

## Co-occurrence probabilities



+



$$q_{11} = 0$$



+



$$q_{10} = 0.5$$



+



$$q_{01} = 0.5$$



+



$$q_{00} = 0.0$$

Negative association



## Occurrence probabilities



$$p_1 = 0.5$$



$$p_2 = 0.5$$

## Co-occurrence probabilities



+



$$q_{11} = 0.5$$



+



$$q_{10} = 0$$



+



$$q_{01} = 0$$



+



$$q_{00} = 0.5$$

Positive association

## Occurrence probabilities



$$p_1 = 0.5$$



$$p_2 = 0.5$$

## Co-occurrence probabilities



+



$$q_{11} = 0.1$$



+



$$q_{10} = 0.4$$



+



$$q_{01} = 0.4$$



+



$$q_{00} = 0.1$$

Negative association

## HMSC with fixed and random effects

$$L_{ij} = L_{ij}^F + L_{ij}^R$$

**Fixed effects:**

$$L_{ij}^F = \sum_{k=1}^{n_c} x_{ik} \beta_{kj}$$

**Random effects:**

$$L_{ij}^R = \sum_{h=1}^{n_f} \eta_{ih} \lambda_{hj}$$

Site loadings

Species loadings

## HMSC with fixed and random effects (in matrix notation)

$$L_{ij} = L_{ij}^F + L_{ij}^R$$

$$\mathbf{L} = \mathbf{L}^F + \mathbf{L}^R$$

**Fixed effects:**

$$L_{ij}^F = \sum_{k=1}^{n_c} x_{ik} \beta_{kj}$$

$$\mathbf{L}^F = \mathbf{X}\mathbf{B}$$

**Random effects:**

$$L_{ij}^R = \sum_{h=1}^{n_f} \eta_{ih} \lambda_{hj}$$

$$\mathbf{L}^R = \mathbf{H}\mathbf{\Lambda}$$

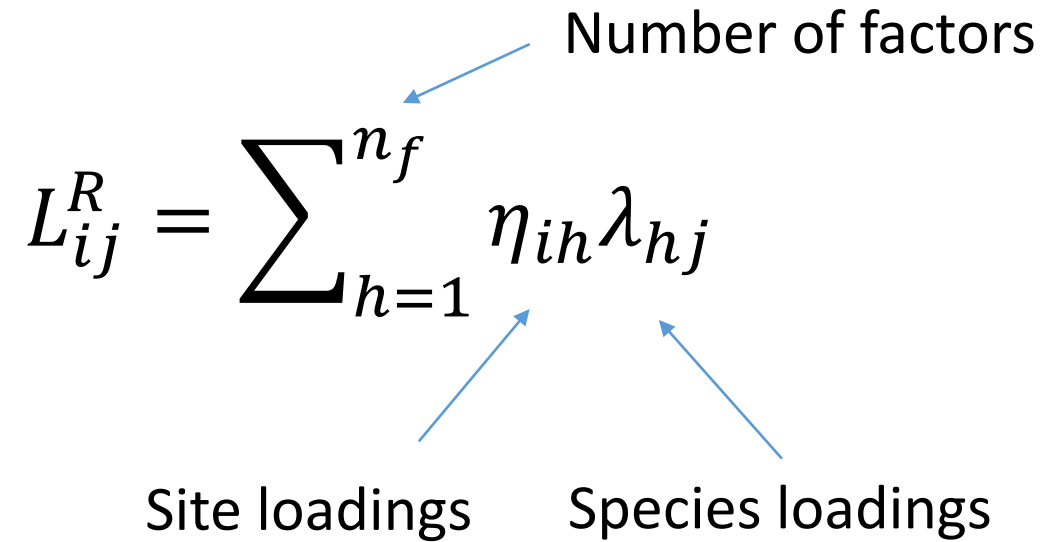
# Prior distributions for site and species loadings

$$L_{ij}^R = \sum_{h=1}^{n_f} \eta_{ih} \lambda_{hj}$$

Number of factors

Site loadings

Species loadings



iid

$$\eta_{ih} \sim N(0,1)$$

$$\lambda_{hj} \sim \text{Multiplicative gamma process shrinking prior (Bhattacharya and Dunson 2011)}$$



# Prior distributions for site and species loadings

$$L_{ij}^R = \sum_{h=1}^{n_f} \eta_{ih} \lambda_{hj}$$

Number of factors

Site loadings

Species loadings

iid

$$\eta_{ih} \sim N(0, 1)$$

Multiplicative gamma process shrinking prior for  $\lambda_{hj}$ :

$$\lambda_{hj} \mid \phi_{hj}, \delta \sim N\left(0, \phi_{hj}^{-1} \tau_h^{-1}\right), \tau_h = \prod_{l=1}^h \delta_l$$

$$\phi_{hj} \mid v \sim \text{Ga}(v/2, v/2)$$

$$\delta_1 \mid a, b \sim \text{Ga}(a_1, b_1), \delta_l \mid a, b \sim \text{Ga}(a_2, b_2) \text{ for } l \geq 2$$

# Covariance between linear predictors

$$L_{ij}^R = \sum_{h=1}^{n_f} \eta_{ih} \lambda_{hj}$$

Number of factors

Site loadings

Species loadings

iid

$$\eta_{ih} \sim N(0,1) \Rightarrow \text{Cov}[L_{i_1 j_1}^R, L_{i_2 j_2}^R] = \sum_{h=1}^{n_f} \lambda_{h j_1} \lambda_{h j_2} \delta_{i_1 i_2}$$

## Covariance between linear predictors

$$L_{ij}^R = \sum_{h=1}^{n_f} \eta_{ih} \lambda_{hj}$$

Number of factors

Site loadings

Species loadings

iid

$$\eta_{ih} \sim N(0,1) \Rightarrow \text{Cov}[L_{i_1 j_1}^R, L_{i_2 j_2}^R] = \sum_{h=1}^{n_f} \lambda_{h j_1} \lambda_{h j_2} \delta_{i_1 i_2}$$

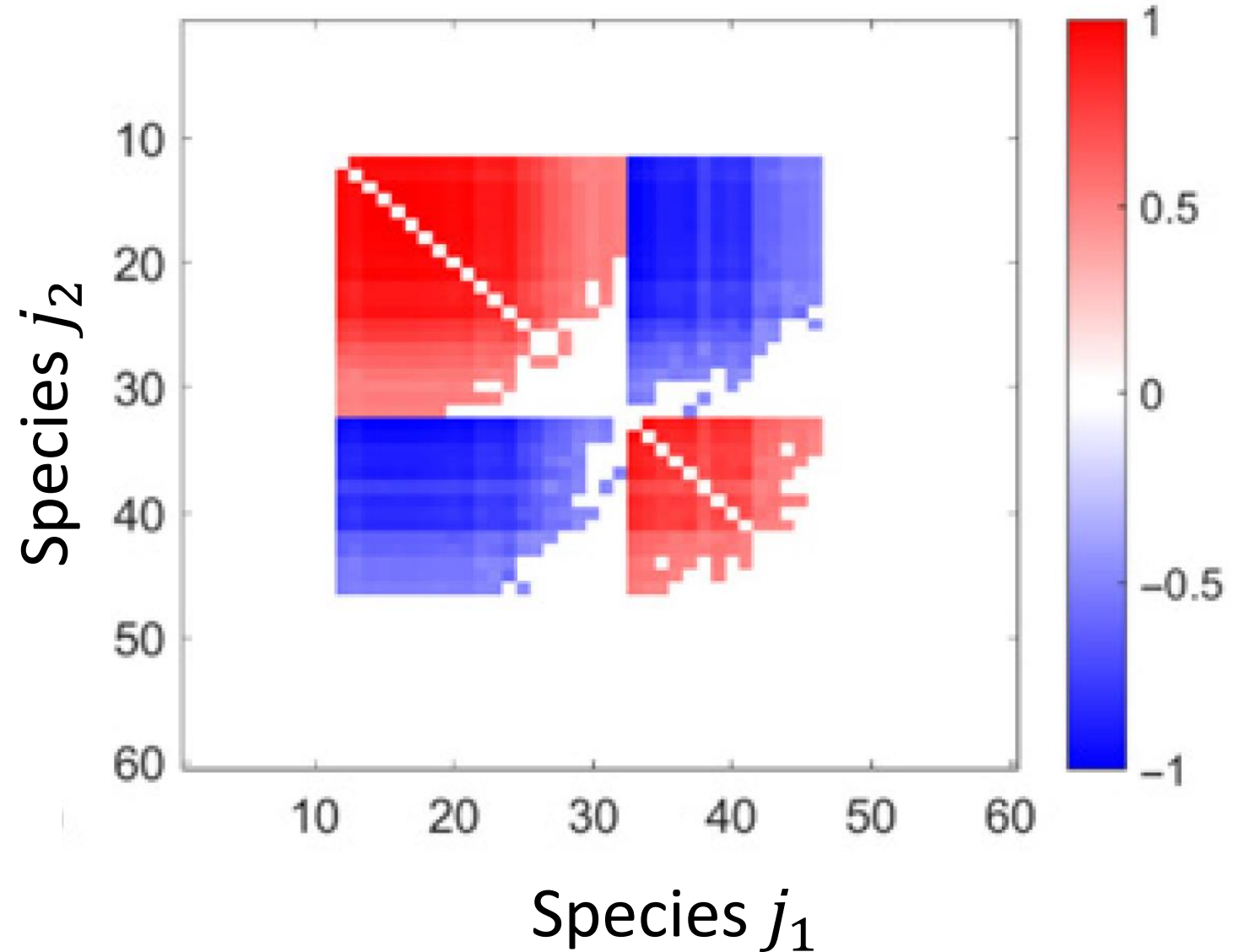
$$\Rightarrow \mathbf{L}_{i\cdot}^R \sim N(0, \mathbf{\Omega}) \qquad \mathbf{\Omega} = \mathbf{\Lambda}^T \mathbf{\Lambda}$$

# Association matrix at the correlation scale

**R**

$$\mathbf{R} = \text{scale}(\mathbf{\Omega})$$

$$R_{j_1 j_2} = \frac{\Omega_{j_1 j_2}}{\sqrt{\Omega_{j_1 j_1} \Omega_{j_2 j_2}}}$$



# Unstructured & structured (spatial) site loadings

$$L_{ij}^R = \sum_{h=1}^{n_f} \eta_{ih} \lambda_{hj}$$

HmscRandomLevel(units=plots)

**Un-structured site loadings:**  $\eta_{ih} \stackrel{\text{iid}}{\sim} N(0,1) \Rightarrow \text{Cov}[L_{i_1 j_1}^R, L_{i_2 j_2}^R] = \sum_{h=1}^{n_f} \lambda_{h j_1} \lambda_{h j_2} \delta_{i_1 i_2}$

HmscRandomLevel(sData=xy)

**Spatial site loadings:**  $\boldsymbol{\eta}_{\cdot h} \sim N(0, \boldsymbol{\Sigma}_h), \quad \Sigma_{h, i_1 i_2} = \exp(-d_{i_1 i_2} / \alpha_h)$

$$\Rightarrow \text{Cov}[L_{i_1 j_1}^R, L_{i_2 j_2}^R] = \sum_{h=1}^{n_f} \lambda_{h j_1} \lambda_{h j_2} \exp(-d_{i_1 i_2} / \alpha_h)$$

# Multiple random effects in the same model

## Study design

Presence-absence of 60 bryophyte species surveyed on 204 aspen trees within 14 natural forest sites and 14 logging sites.



Retention aspens on a logging site.



*Radula complanata*



*Neckera pennata*

Ovaskainen et al. 2017 (ELE), Oldén et al. (2014)

# Multiple random effects in the same model

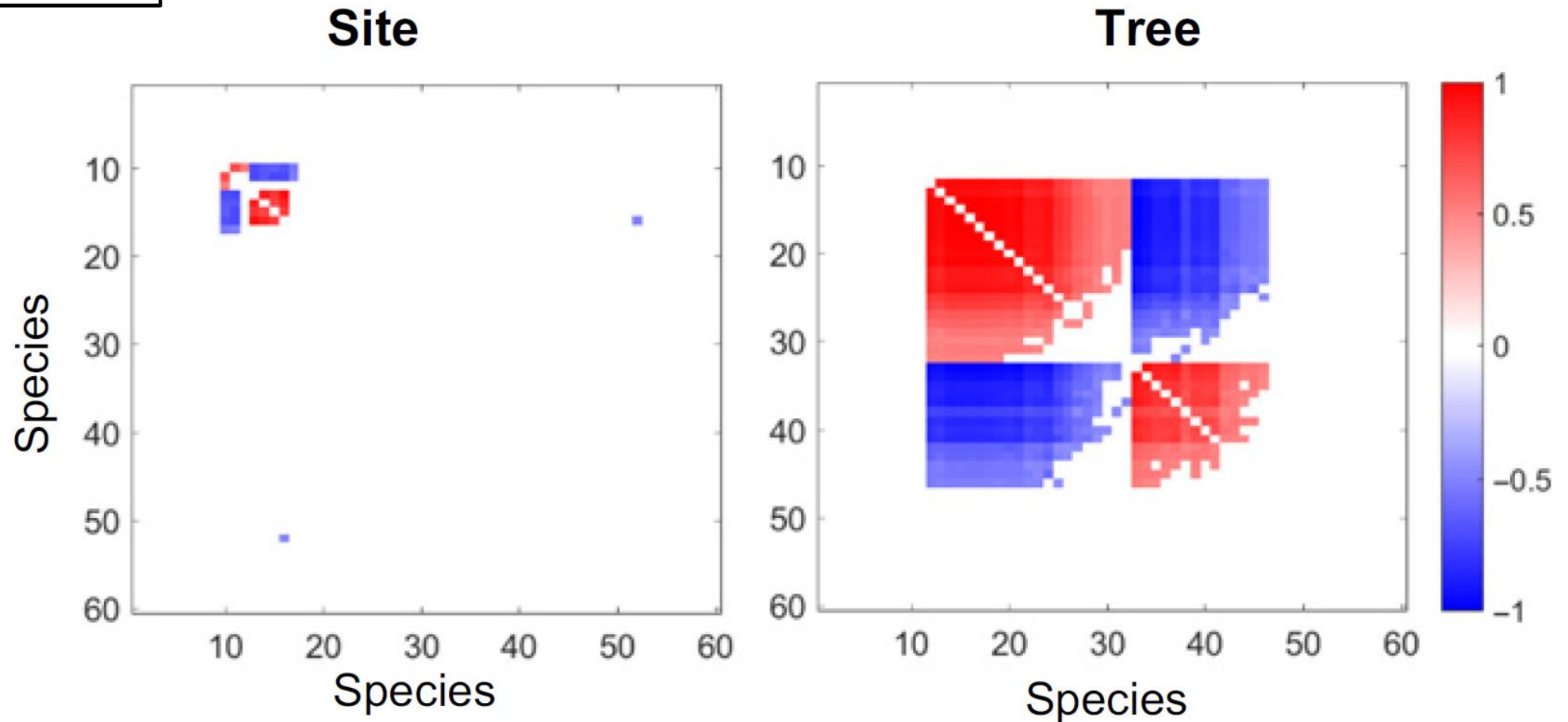
## Study design

Presence-absence of 60 bryophyte species surveyed on 204 aspen trees within 14 natural forest sites and 14 logging sites.




Retention aspens on a logging site.

## Species associations



# Multiple random effects in the same model

Number of random effects


$$L_{ij}^R = \sum_{r=1}^{n_r} L_{ij}^{r,R}$$

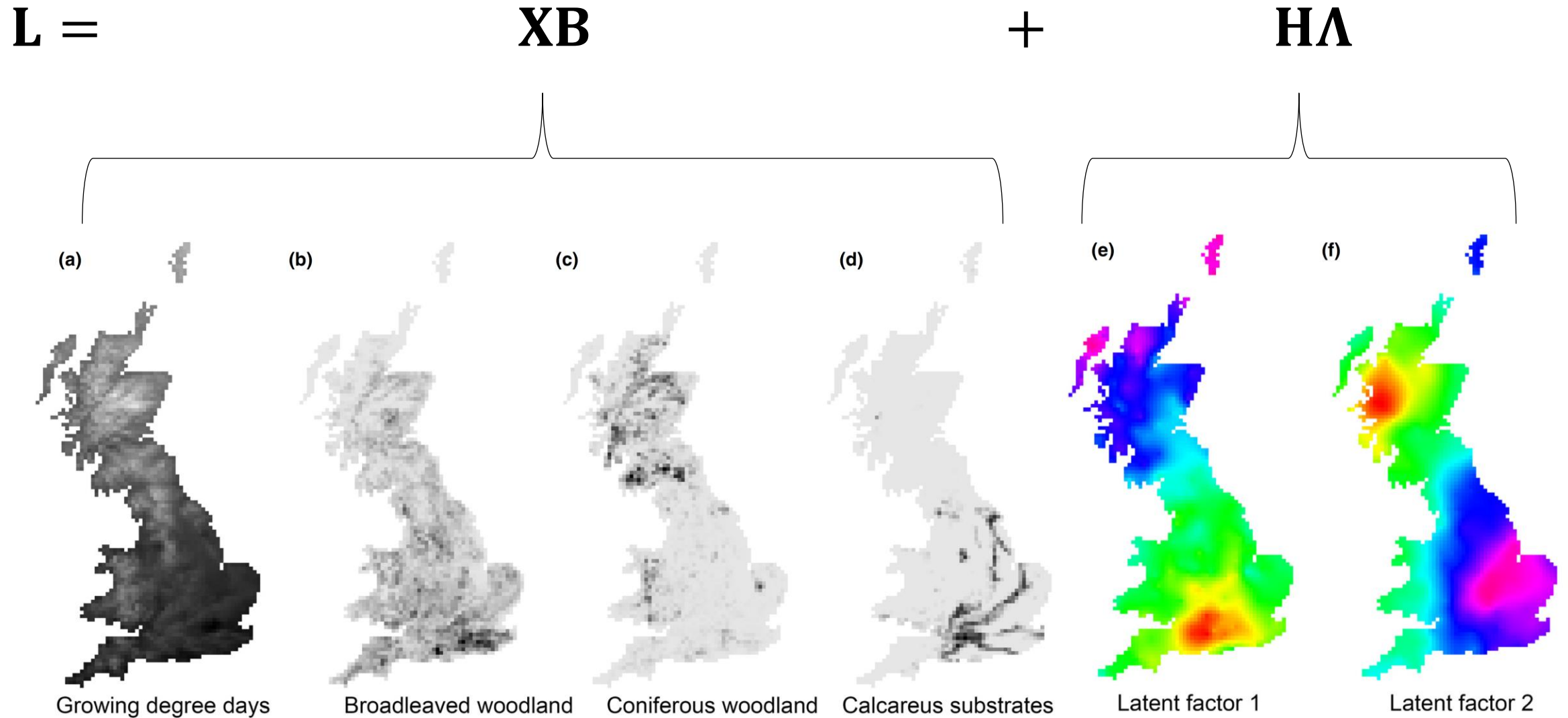
$$L_{ij}^{r,R} = \sum_{h=1}^{n_f^r} \eta_{u^r(i)h}^r \lambda_{hj}^r$$



Units (e.g. plots) of random effect  $r$



# Spatial random effects / site loadings as hidden environmental covariates



Ovaskainen et al. 2016 (MEE): butterflies in UK

# Fitting a multivariate spatial model enables using spatial information OF THE SAME AND OTHER SPECIES when generating predictions

