

Building HMSC step by step: biotic interactions

7 Joint Species Distribution Modelling: Biotic Interactions	142
7.1 Strategies for Estimating Biotic Interactions in Species Distribution Models	143
7.2 Occurrence and Co-occurrence Probabilities	144
7.3 Using Latent Variables to Model Co-occurrence	147
7.4 Accounting for the Spatio-temporal Context through Latent Variables	152
7.5 Covariate-Dependent Species Associations	156
7.6 A Cautionary Note about Interpreting Residual Associations as Biotic Interactions	159
7.7 Using Residual Species Associations for Making Improved Predictions	160
7.8 Simulated Case Studies with HMSC	165
7.9 Real Case Study with HMSC: Sequencing Data on Dead Wood-Inhabiting Fungi	172

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

Linear model: $y_{ij} = L_{ij} + \varepsilon_{ij}$

Presence-absence model: $y_{ij} \sim \text{Bernoulli}(\Phi(L_{ij}))$

Count model: $y_{ij} \sim \text{Poisson}(\exp(L_{ij} + \varepsilon_{ij}))$

Thus far:

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

Now we extend to:

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

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

Number of factors

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

Site loadings Species loadings

iid

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

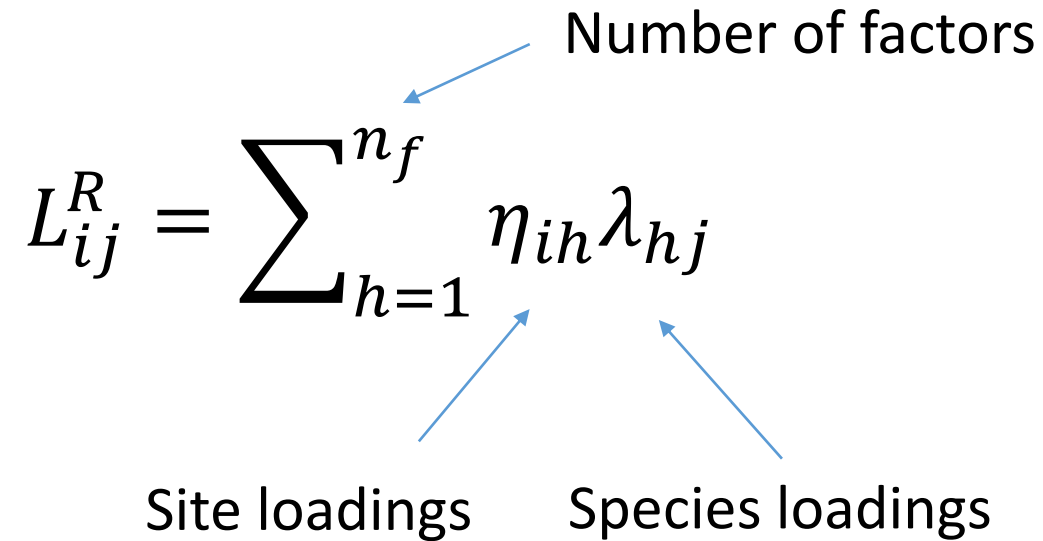
$$\lambda_{hj} \sim \text{Multiplicative gamma process shrinking prior (Bhattacharya and Dunson 2011)}$$

Prior distributions for site and species loadings

Number of factors

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

Site loadings Species loadings



iid

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

Multiplicative gamma process shrinking prior for λ_{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

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

Diagram illustrating the components of the linear predictor equation:

- n_f : Number of factors (indicated by an arrow from the text "Number of factors" to the superscript n_f)
- η_{ih} : Site loadings (indicated by an arrow from the text "Site loadings" to the term η_{ih})
- λ_{hj} : Species loadings (indicated by an arrow from the text "Species loadings" to the term λ_{hj})

iid

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

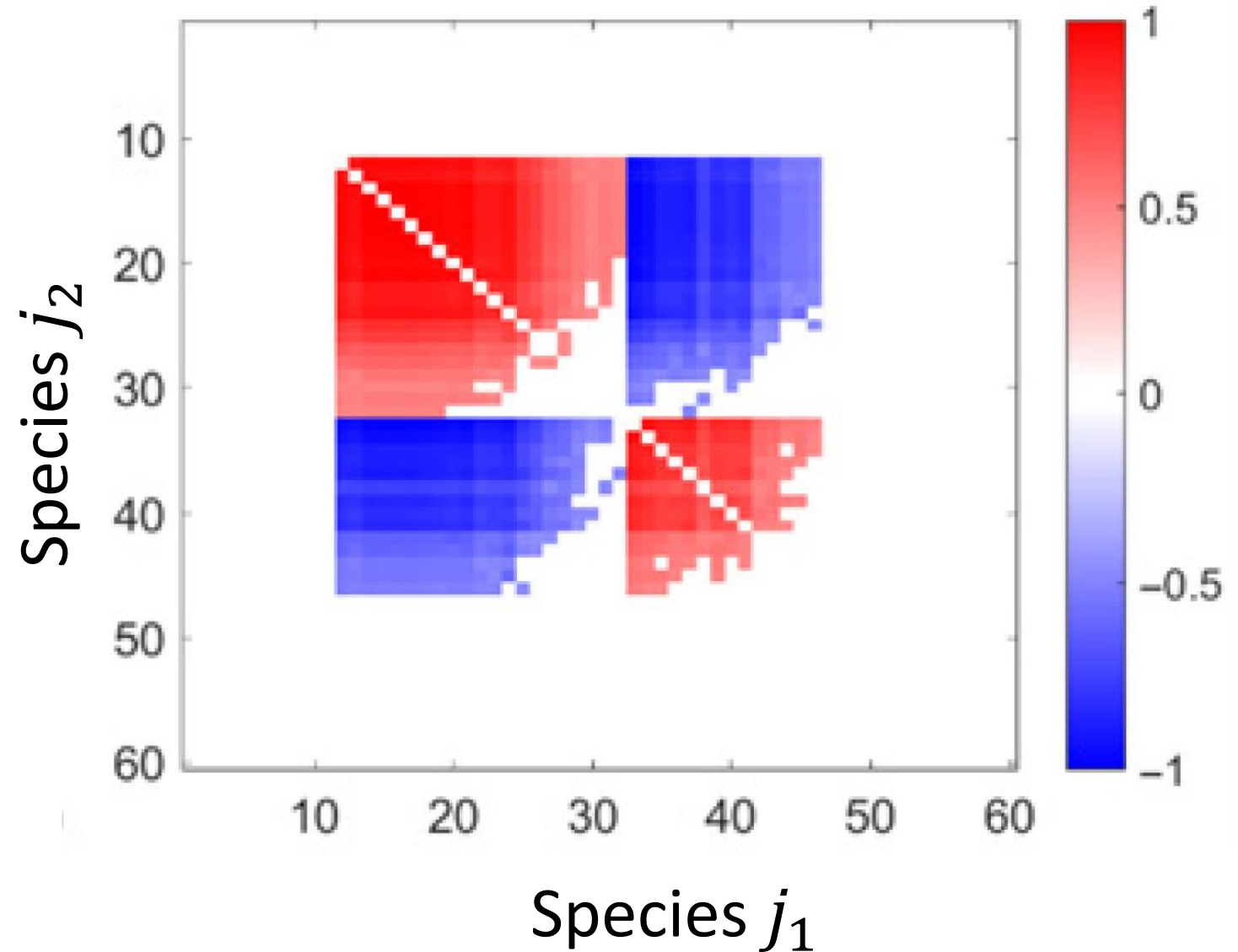
$$\Rightarrow \mathbf{L}_{i\cdot}^R \sim N(0, \mathbf{\Omega}) \quad \mathbf{\Omega} = \mathbf{\Lambda}^T \mathbf{\Lambda}$$

Association matrix at the correlation scale

R

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

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



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)

Study design

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

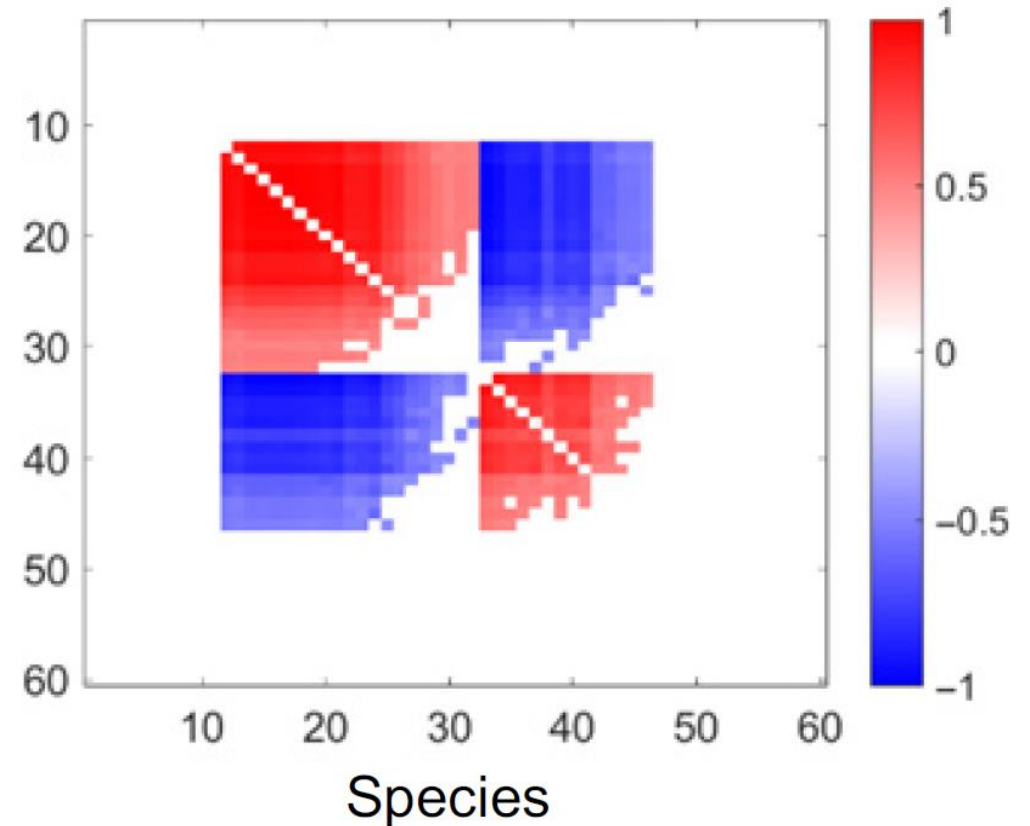


Retention aspens on a logging site.




Multiple random effects in the same model

Site



Multiple random effects in the same model

Number of random effects

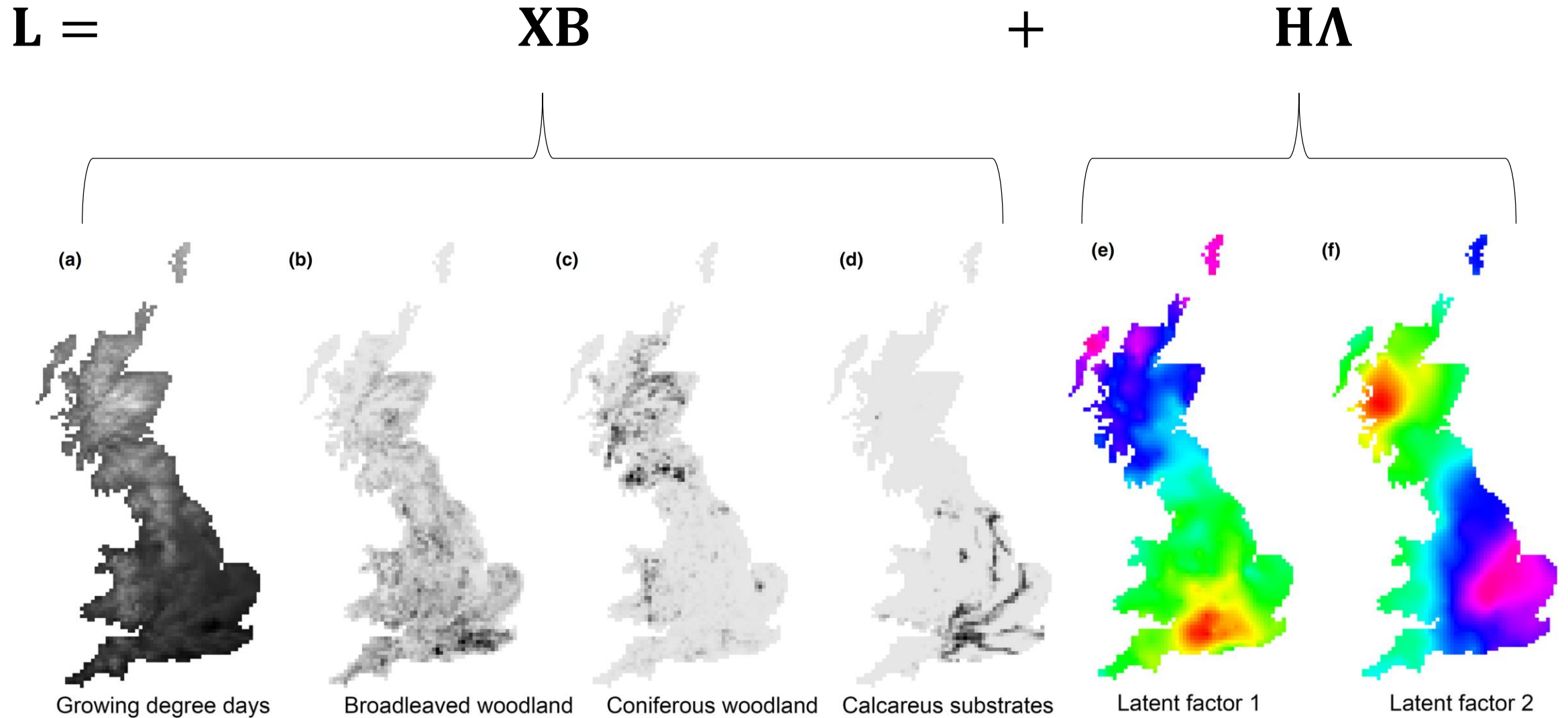

$$L_{ij}^R = \sum_{r=1}^{n_r} L_{ij}^{r,R}$$

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



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

Prior distribution for spatial site loadings

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

HmscRandomLevel(units=plots)

Un-structured site loadings:

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

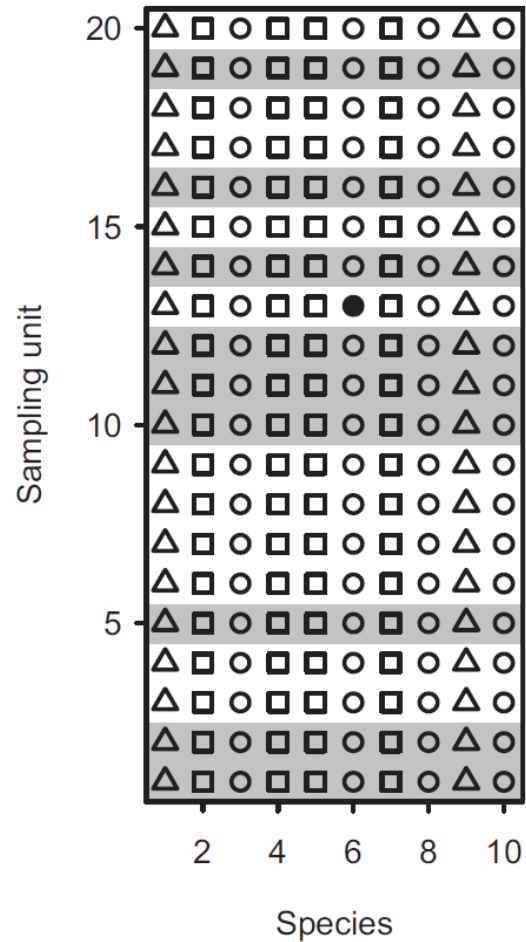
HmscRandomLevel(sData=xy)

Spatial site loadings:

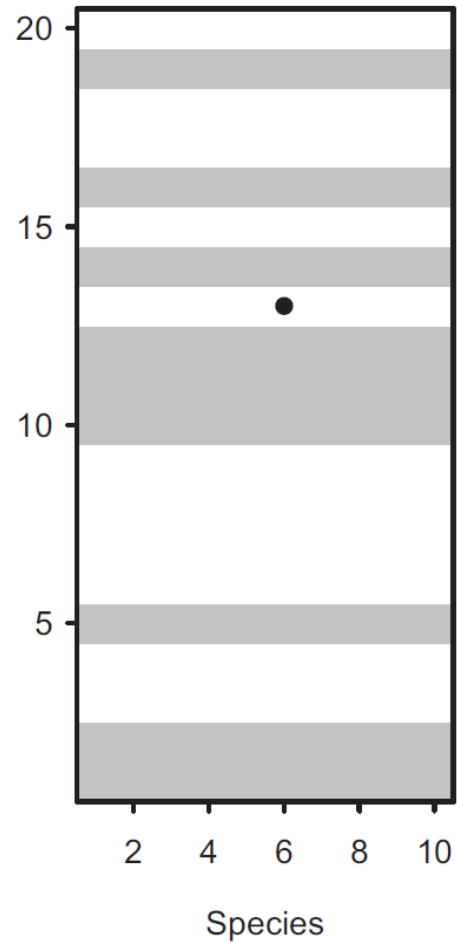
$$\boldsymbol{\eta}_{\cdot h} \sim N(0, \boldsymbol{\Sigma}_h), \quad \Sigma_{h, i_1 i_2} = \exp(-d_{i_1 i_2} / \alpha_h)$$
$$\Rightarrow \text{Cov}[L_{i_1 j_1}^R, L_{i_2 j_2}^R] = \sum_{h=1}^{n_f} \lambda_{h j_1} \lambda_{h j_2} \exp(-d_{i_1 i_2} / \alpha_h)$$

Conditional cross-validation

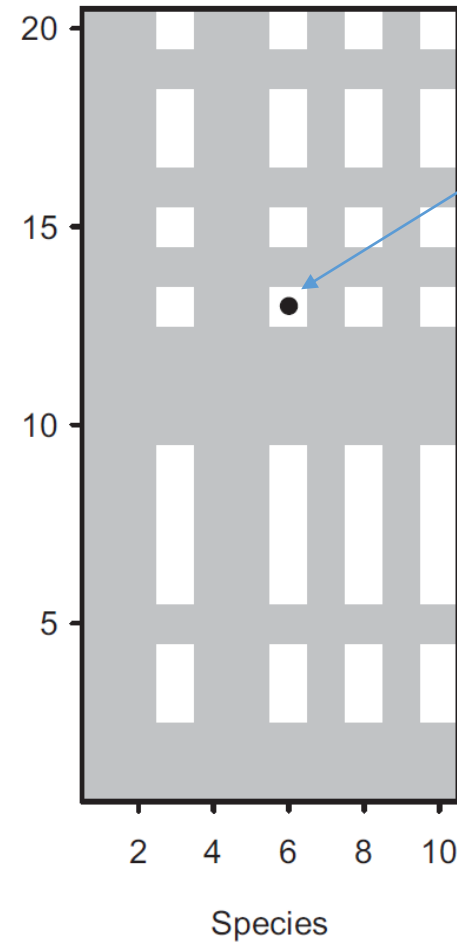
Splitting the data to folds across sampling units and species



Cross-validation



Conditional cross-validation



Focal data point to be predicted

- Data used for model fitting
- Data not used for model fitting