Generalised Linear (Mixed) Models for data of multiple species

Bert van der Veen

Department of Mathematical Sciences, NTNU

Species correlation

If we fit a GLM to data of multiple species, we assume **independence**

But, observations of the same species form groups. Co-occurring species have more similar observations than for other species

In GLMM language: **observations of species exhibit correlation**

1) Part of this can be explained by shared environmental responses
2) The other part remains

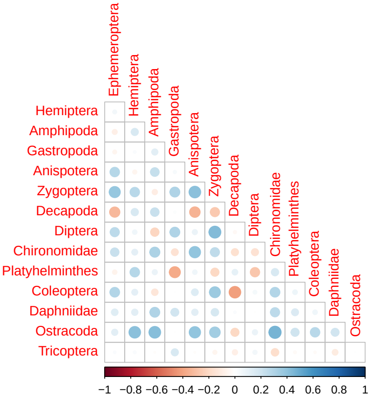
The previous model

```r
model4 <- gllvm::gllvm(y, X = X, formula = ~NO3, num.lv = 0,
                       family = "negative.binomial")
```

If we look at the correlations in the residuals of this model, we can see that.

Correlated residuals

```
corrplot::corrplot(cor(residuals(model4)$resi), type = "lower", diag = FAL
```



Background

If you have presence-absence data of a species, you fit a Species Distribution Model of the form:

$$y_{ij} = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} \tag{1}$$

- $\mathbf{x}_i$ is usually a bioclimatic variable
- Then you want to predict where a species may occur
- Potentially based on future climate scenarios
- But what if you have data on multiple species?
    - Fitting multiple SDMs requires *independence*

Joint Species Distribution Modeling

A decade ago, Joint Species Distribution Models (JSDM) were introduced to model binary data of multiple species

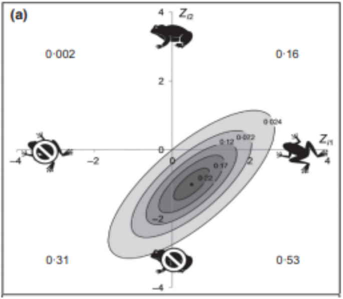
- Pollock et al. (2015): co-occurrence of frogs and trees
- Clark et al. (2015): co-occurrence of trees

The goal: to incorporate covariation of species

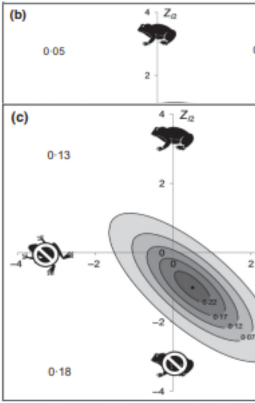
Species associations

What induces

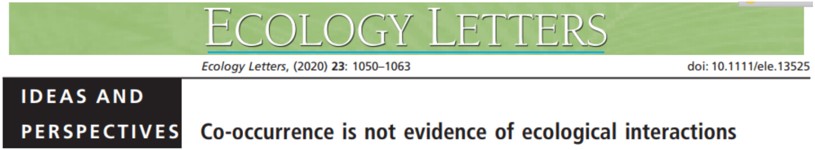
- Shared envir responses (a conditions)
- Biotic intera



1: Pollock e



Interactions and co-occurrence



ECOLOGY LETTERS

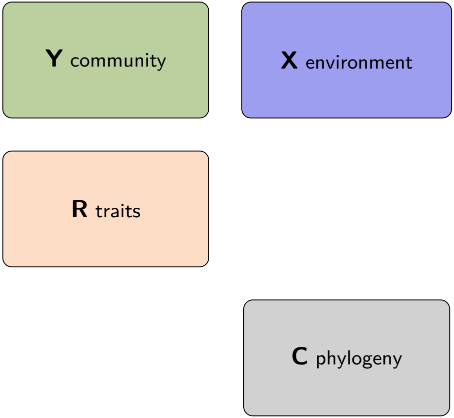
Ecology Letters, (2020) **23**: 1050–1063 doi: 10.1111/ele.13525

IDEAS AND PERSPECTIVES

Co-occurrence is not evidence of ecological interactions

Interactions induce correlation, but we cannot infer interactions from (non-temporal) co-occurrence data.

Typical questions in the framework



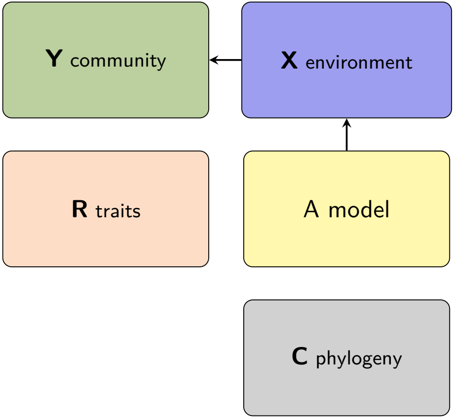
Y community

X environment

R traits

C phylogeny

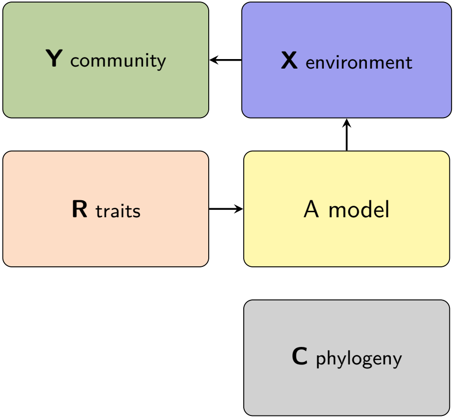
Typical questions in the framework



Q: How does the environment structure the community?

environmental filtering

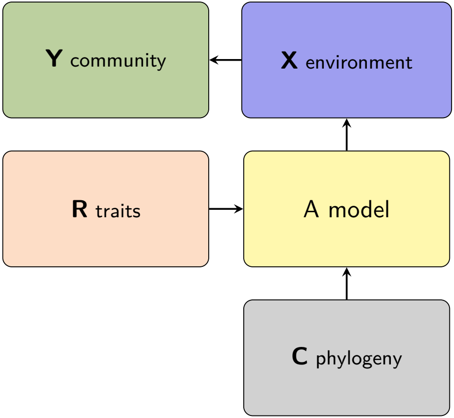
Typical questions in the framework



Q: How do traits affect species' responses to the environment?

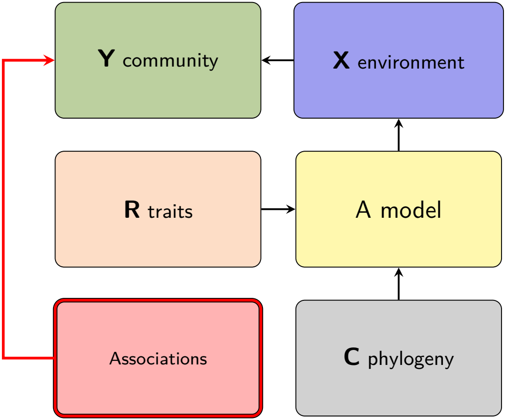
environmental filtering

Typical questions in the framework



Q: Do species with shared evolutionary history co-occur? (phylogenetic structuring) environmental filtering

Typical questions in the framework



Q: Do species co-occur **after** the environment has been considered? biotic filtering

Joint Species Distribution

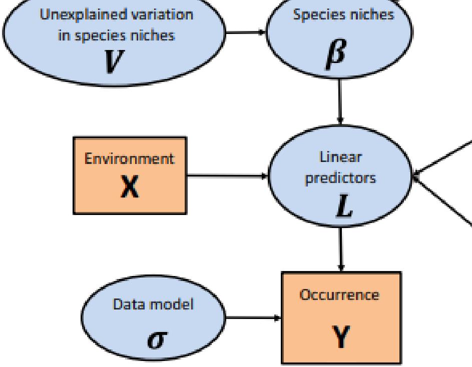


Figure 2: Figure from Ovaskainen et al. (2017)

Joint Species Distribution Modeling

The idea of JSDMs is to incorporate *species associations*

- Species may co-occur due to biotic interactions
- Due to similar environmental preferences
- Or because they have a similar history

Either how, it results in correlations between responses

Joint Species Distribution Model (JSDM)

- For community data, we want to incorporate correlation of species
- We have Multivariate data (in contrast to multivariable)

$$g\{\mathbb{E}(\mathbf{y}_i|\boldsymbol{\epsilon}_i)\} = \boldsymbol{\beta}_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \boldsymbol{\epsilon}_i \tag{2}$$

- we add $\boldsymbol{\epsilon}_i$ relative to the VGLM(M)
- This random effect takes care of the left-over (co)variation of species
- so we assume $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$
- $\boldsymbol{\Sigma}$ is the matrix of **species associations**

JSDM: the model

$$\eta_{ij} = \beta_{0j} + \ldots + \epsilon_{ij} \tag{3}$$

- $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$
- $\boldsymbol{\Sigma}$ is the matrix of *species associations*
- So we expect a positive values of species co-occur, and negative if they do not

JSDM: it is a mixed-effects model

The JSDM is “just” a mixed-effects model. So we can fit it with available software:

In `lme4`:

```
glmer(abundance ~ species + x:species + (0+species|sites), data = data)`
```

- There are $p(p+1)/2$ correlations between species
- This model becomes (very) large very quickly
- Will usually not fit
- So we need to do something smart!

Species associations

- Difficult to estimate: there are usually too many parameters
- Can only fit this way when there are (much) more sites than species
- The number of pairwise associations grows quadratically
  - 2 with 2 species, 6 for 4 species, 45 for 10 species, 4950 for 100

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & sp_{12} & \cdots & sp_{1j} \\ sp_{21} & 1 & \cdots & sp_{2j} \\ \vdots & & \ddots & \vdots \\ sp_{j1} & sp_{j2} & \cdots & 1 \end{bmatrix} \tag{4}$$

This very quickly becomes an issue for fitting models

Ordination to the rescue

- GLLVMs were introduced as a technical solution to this problem
- We represent the covariance matrix with fewer **dimensions**: $\boldsymbol{\Sigma} \approx \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$

“The factor analytic solution” because factor analysis (Spearman, 1904) is the precursor of GLLVMs

JSDM vs classical multivariate analysis

	Classic	JSDM
Focus	Low-dimensional space	Distributions
Goal	Inference	Prediction
Data type	Usually quantitative	Binary
Scale	Local	Regional
Covariates	Environmental	Bioclimatic
Presentation	Ordination diagram	Correlation plot/map
Audience	Community ecologists	Macro ecologists

JSDM vs classical multivariate analysis

	Classic	JSDM
Focus	Low-dimensional space	Distributions
Goal	Inference	Prediction
Data type	Usually quantitative	Binary
Scale	Local	Regional
Covariates	Environmental	Bioclimatic
Presentation	Ordination diagram	Correlation plot/map
Audience	Community ecologists	Macro ecologists

That is not to say JSDMs cannot be used for non-binary data, for inference, or for local scales

JSDM software implementations

There are many!

JSDM software implementations

There are many!

- Boral (Bayesian, slow and somewhat outdated)
- sJSDM (Bayesian, relatively slow, but faster than Boral)
- Hmsc (Bayesian, generally slow, loads of functionality)
- ecoCopla (Frequentist, very fast but limited functionality)
- CBFM (Frequentist, geared towards spatio-temporal analysis)
- sjSDM (Frequentist, very fast but limited functionality, requires python)
- glmmTMB (Frequentist, fast and very versatile, not purpose-coded)
- gllvm (Frequentist, fast and very versatile, purpose-coded)

JSDM software implementations

There are many!

- Boral (Bayesian, slow and somewhat outdated)
- sJSDM (Bayesian, relatively slow, but faster than Boral)
- Hmsc (Bayesian, generally slow, loads of functionality)
- ecoCopla (Frequentist, very fast but limited functionality)
- CBFM (Frequentist, geared towards spatio-temporal analysis)
- sjSDM (Frequentist, very fast but limited functionality, requires python)
- glmmTMB (Frequentist, fast and very versatile, not purpose-coded)
- gllvm (Frequentist, fast and very versatile, purpose-coded)

Something practical

Presence-absence models in `gllvm` tend to be a bit finicky:

- The data has low information content: hard to find good starting values
- Try different starting values
- Try different approximation
- I.e., play around with the settings a little

Example with Eucalyptus data (Pollock et al. 2015)

```
Y <- read.csv("../data/eucalyptY.csv")[,-1]
X <- read.csv("../data/eucalyptX.csv")[,-1]
X <- data.frame(lapply(X, function(x)if(is.numeric(x)){scale(x)}else{as.factor(x)}))
knitr::kable(head(X), format="latex", booktabs = T)
```

IDENT	elev	PerSlope	aspect	Northness	Sdclass	Rockiness	VallyBotFlat	topowe
MGS1	0.1972540	1.0236658	-1.1686010	0.3254470	-1.551741	1.0315338	-0.5939667	-0.407019
MGS5	0.1972540	0.7778142	-1.1686010	0.3254470	-1.551741	1.4558834	-0.5939667	-0.407019
MGS4	0.1757452	1.0236658	-0.8419304	-0.5684498	-1.551741	1.3144335	-0.5939667	-0.163936
MGS3	0.1757452	0.4090369	-0.8419304	-0.5684498	-1.426354	0.4657344	-0.5939667	-0.476998
MGS2	0.1327276	0.5319627	-0.3583180	-1.5163833	-1.426354	-0.2415148	-0.5939667	-1.003677
MGN1	0.6381843	1.3924432	0.7306972	0.0144213	-1.426354	1.5973333	-0.5939667	-0.889502

- 20 species
- 458 sites
- Soil covariates and a few bioclimatic

Eucalyptus: fit a model

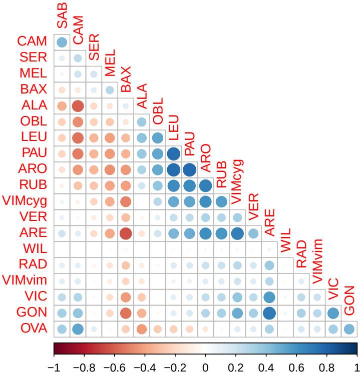
```r
jsdm1 <- gllvm::gllvm(Y, X = X, formula = ~ Sandiness + cvTemp,
  family = "binomial", num.lv = 2, method = "EVA", starting.val = "zero")
```

- `method`: controls the approximation (LA, VA (default), EVA)
- `starting.val`: how to generate initial values (multimodality)
- `n.init`: rerun multiple times

The `gllvm` R-package is fast, but might can take a minute.
Models should be refitted with ‘n.init‘.

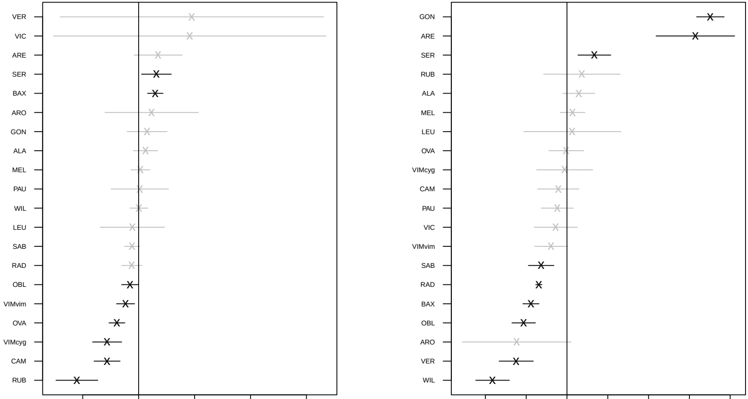
Eucalyptus: residual assocations

```
corrplot::corrplot(gllvm::getResidualCor(jsdm1), order = "AOE", type = "lower", diag = FALSE)
```



Eucalyptus: environment effects

```r
gllvm::coefplot(jsdm1)
```



Eucalyptus: explained variation

```
jsdm2 <- gllvm::gllvm(Y, family  = "binomial", num.lv = 2, method = "EVA",
                             starting.val = "zero")
gllvm::getResidualCov(jsdm1)$trace/gllvm::getResidualCov(jsdm2)$trace
```

```
## [1] 0.6172353
```

Eucalyptus: with random effects

```
jsdm3 <- gllvm::gllvm(Y, X = X, formula = ~(0+Sandiness + cvTemp|1),
                      family  = "binomial", num.lv = 2,
                      method = "EVA", starting.val = "zero")
```

Many random effects can take long. Speed up with:

- `sd.errors = FALSE` (turn off standard error calculation)
- `Ab.struct = "diagonal"` (simplify the approximation)
- `TMB::openmp(parallel::detectCores()-1, autopar = TRUE, DLL = "gllvm")` (parallel computation)
- `optim.method = "L-BFGS-B"`

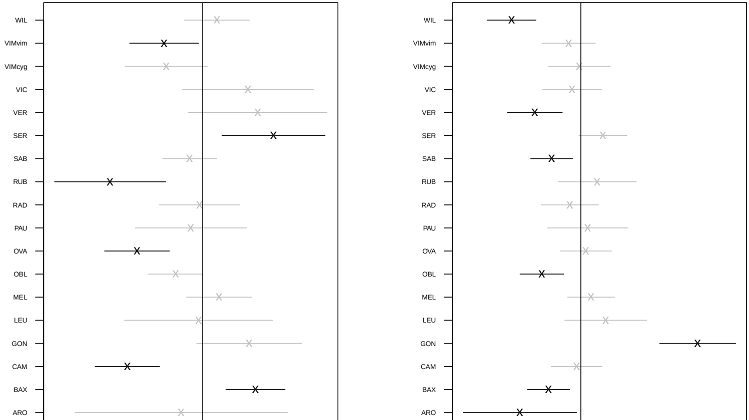
Eucalyptus: with random effects

```
summary(jsdm3)
```

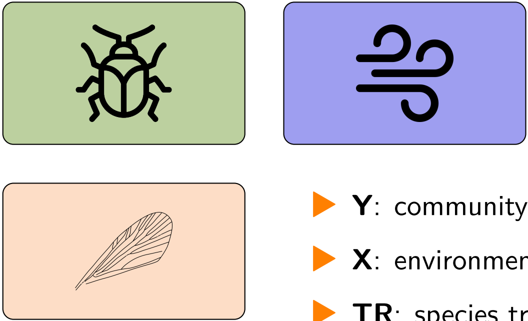
```
##
## Call:
## gllvm::gllvm(y = Y, X = X, formula = ~(0 + Sandiness + cvTemp |
##     1), family = "binomial", num.lv = 2, method = "EVA", starting.val =
##
## Family:  binomial
##
## AIC:  4486.426 AICc:  4487.341 BIC:  4942.273 LL:  -2179.2 df:  64
##
## Informed LVs:  0
## Constrained LVs:  0
## Unconstrained LVs:  2
##
## Formula:  ~(0 + Sandiness + cvTemp | 1)
```

Eucalyptus: with random effects

```
gllvm::randomCoefplot(jsdm3)
```

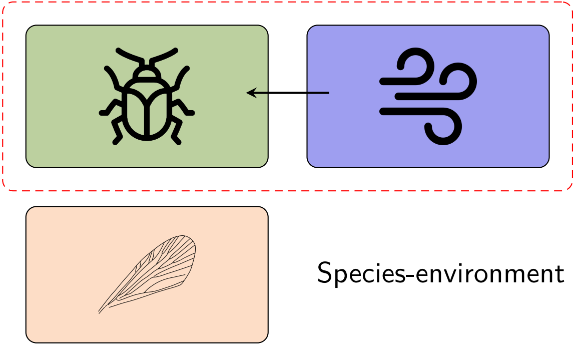


Fourth corner analysis



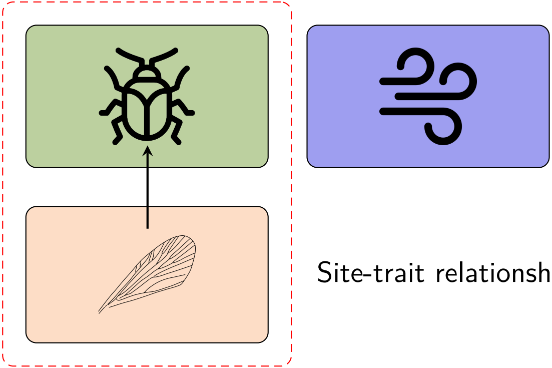
- **Y**: community data
- **X**: environmental variables
- **TR**: species traits

Fourth corner analysis



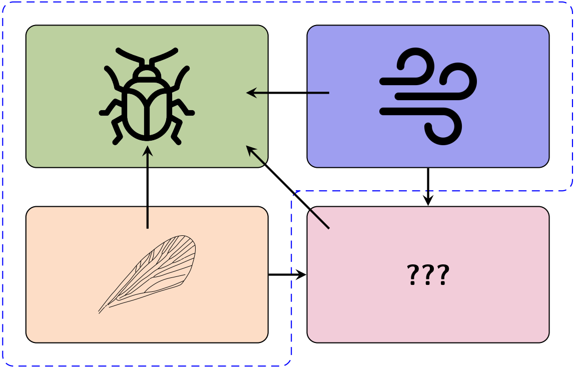
Species-environment relationship: the methods

Fourth corner analysis



Site-trait relationship: VGLM(M)

Fourth corner analysis



Trait-based analysis

1. CWM + RDA *Doledec et al. (1996)*
2. Double constrained ordination *Lebreton et al. (1988), ter Braak et al. (2018)*
3. Fourth corner (LV) Models *Brown et al. (2014), Ovaskainen at al. (2017), Niku et al. (2021)*

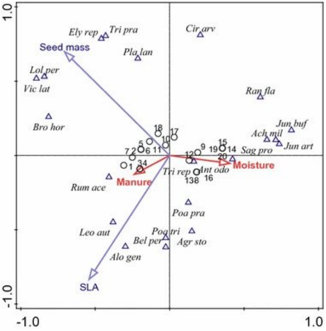
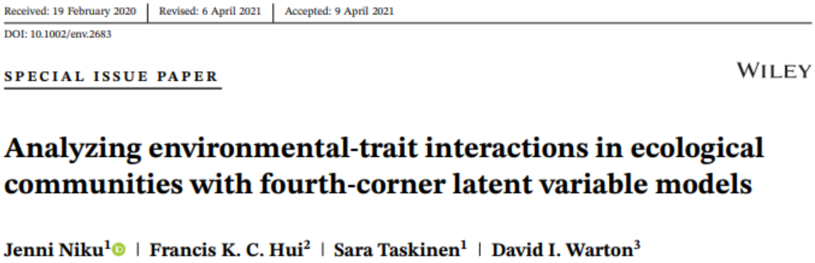


Figure 3: Quadriplot
ter Braak et al. (2018)

Fourth-corner LVMs

A kind of LVM (JSDM) that also includes traits



Received: 19 February 2020 | Revised: 6 April 2021 | Accepted: 9 April 2021

DOI: 10.1002/env.2683

SPECIAL ISSUE PAPER

WILEY

Analyzing environmental-trait interactions in ecological communities with fourth-corner latent variable models

Jenni Niku[1] | Francis K. C. Hui[2] | Sara Taskinen[1] | David I. Warton[3]

Fourth-corner LVMs

The model is very similar to before:

$$\eta_{ij} = \beta_{0j} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \epsilon_{ij} \tag{5}$$

Fourth-corner LVMs

The model is very similar to before:

$$\eta_{ij} = \beta_{0j} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \epsilon_{ij} \tag{5}$$

but now, we are hierarchically modelling species' effects $\boldsymbol{\beta}_j$

Fourth-corner LVMs

Traits

$$\eta_{ij} = \beta_{0j} + \mathbf{x}_i^\top(\boldsymbol{\beta}_x + \mathbf{b}_j) + \mathbf{tr}_j^\top \mathbf{B}_{xtr}\mathbf{x}_i \tag{6}$$

- $\boldsymbol{\beta}_x$ species-common effects
- $\mathbf{b}_j$ species-specific effects
- $\mathbf{B}_{xtr}$ 4th-corner coefficients

Fourth-corner LVMs

$$\boldsymbol{\beta}_j = \boldsymbol{\beta}_x + \mathbf{B}_{xtr}\textbf{tr}_j + \textbf{b}_j \tag{7}$$

We are modeling our species coefficients of covariates

- With an species-common effect: $\boldsymbol{\beta}_x$
- Slopes for traits $\mathbf{B}_{xtr}$: these describe how our species' environment responses depend on traits
    - Can be used nicely for hypothesis testing
- An error term $\textbf{b}_j$ to capture everything left

Fourth-corner LVMs

- With a 4th corner LVM we can examine trait-environment relationships
- Figure out **why** species prefer particular conditions
  - In contrast to “just” which conditions they prefer
- While still incorporating **other reasons** for co-occurrence

Example with Eucalyptus data

```
TR <- read.csv("../data/eucalyptTR.csv");row.names(TR) <- TR$TAXON
```

Example with Eucalyptus data

```
jsdm4 <- gllvm::gllvm(Y, X, TR = TR,
              formula = ~ Sandiness + cvTemp + (Sandiness + cvTemp):
(MedianSLA + MaxHeight.m. + MedianSeedMass.mg.),
              randomX = ~Sandiness + cvTemp,
              family = "binomial", method = "EVA", starting.val = "zero")
```

The formula formulation follows the equation: the first two terms are β_x, the next terms represent $\mathbf{B}_{xtr}$, and b_{kj} last

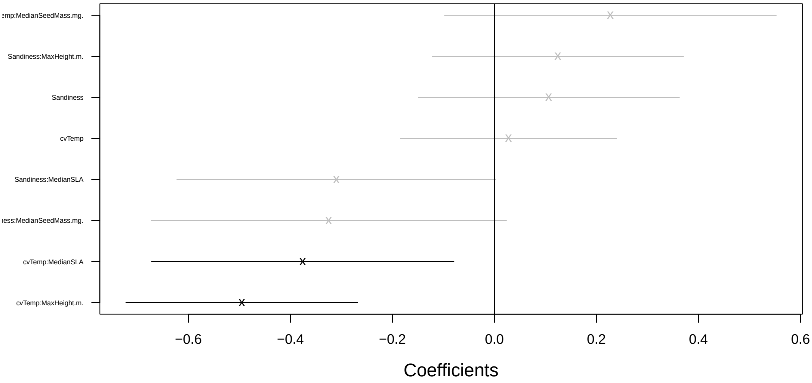
Example with Eucalyptus data

```
summary(jsdm4)
```

```
##
## Call:
## gllvm::gllvm(y = Y, X = X, TR = TR, formula = ~Sandiness + cvTemp +
##     (Sandiness + cvTemp):(MedianSLA + MaxHeight.m. + MedianSeedMass.mg.),
##     family = "binomial", method = "EVA", randomX = ~Sandiness +
##         cvTemp, starting.val = "zero")
##
## Family:  binomial
##
## AIC:  4401.794 AICc:  4402.888 BIC:  4900.377 LL:  -2130.9 df:  70
##
## Informed LVs:  0
## Constrained LVs:  0
## Unconstrained LVs:  2
##
## Formula:  ~Sandiness+cvTemp+Sandiness:MedianSLA+Sandiness:MaxHeight.m.+Sandiness:MedianSeedMass.mg.+cvTemp:MedianSLA+cvTemp:MaxHeight.m.+cvTemp:MedianSeedMass.mg.
## LV formula:  ~ 0
## Row effect:  ~ 1
##
## Random effects:
##  Name      Variance Std.Dev Corr
##  Sandiness 0.0575   0.2398
##  cvTemp    0.3502   0.5918  0.0000
##
## Coefficients predictors:
##                              Estimate Std. Error z value Pr(>|z|)
## Sandiness                     0.02195    0.06392   0.343   0.7313
## cvTemp                        0.02825    0.13767   0.205   0.8374
## Sandiness:MedianSLA          -0.08825    0.08374  -1.054   0.2919
## Sandiness:MaxHeight.m.        0.07646    0.06793   1.126   0.2604
## Sandiness:MedianSeedMass.mg. -0.21936    0.08903  -2.464   0.0137 *
## cvTemp:MedianSLA             -0.08018    0.17758  -0.452   0.6516
## cvTemp:MaxHeight.m.          -0.25183    0.14613  -1.723   0.0848 .
## cvTemp:MedianSeedMass.mg.    -0.13758    0.18615  -0.739   0.4599
```

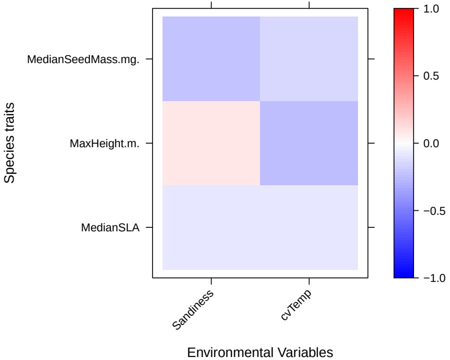
Example with Eucalyptus data

```
gllvm::coefplot(jsdm4)
```



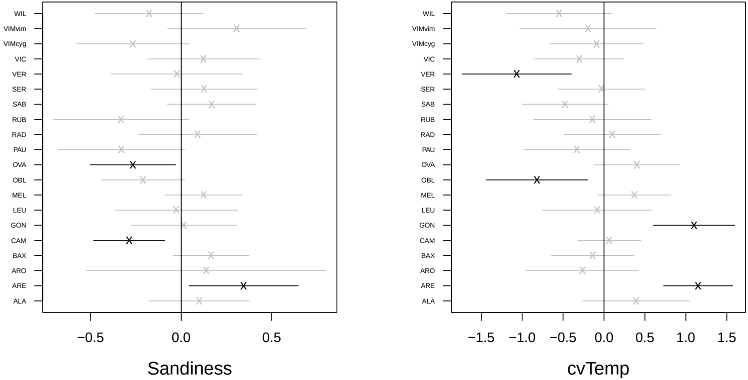
Example with Eucalyptus data

```
plot.4th <- lattice::levelplot(as.matrix(jsdm4$fourth.corner), xlab = "Environmental Variables",
                    ylab = "Species traits", col.regions = colort(100), cex.lab = 1.3,
                    at = seq(-1, 1, length = 100), scales = list(x = list(rot = 45)))
plot.4th
```

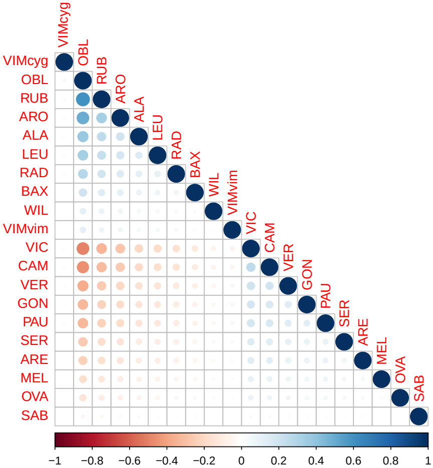


Example with Eucalyptus data

```
gllvm::randomCoefplot(jsdm4)
```



Example with Eucalyptus data



Example with Eucalyptus data: hypothesis testing

```r
anova(jsdm4, jsdm3)
```

```
## Model  1 :  y ~ NULL
## Model  2 :  ~ Sandiness + cvTemp + (Sandiness + cvTemp):(MedianSLA + Ma
```

```
##   Resid.Df        D Df.diff P.value
## 1     9096  0.00000       0
## 2     9090 96.63163       6       0
```

We accept the alternative hypothesis: species responses to the environment are structured by traits

Phylogenetic random effects

So far, we have been discussing **unstructured** species associations

But what if we have information to provide?

Phylogenetic random effects

- We can structure species random effects $\mathbf{b}_j$ by Phylogeny
- We assume that more closely related species have similar responses to the environment

The Phylogeny provides more information and makes for more accurate estimation

(and we can predict for species without data)

Phylogenetic random effects

Here I will omit traits and LVs for brevity. So our model is:

$$\boldsymbol{\eta} = \mathbf{1}\beta_{0j}^{\top} + \mathbf{XB} \tag{8}$$

- $\mathbf{B}$ are the random effects for covariates
- We assume $\mathbf{B} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_m \otimes \mathbf{\Sigma}_r)$
- $\mathbf{\Sigma}_r$ covariance matrix of random effects
- $\mathbf{\Sigma}_m$ correlation matrix due to phylogeny

We assume that all our random effects are structured by the Phylogeny

Phylogenetic random effects

$$\mathbf{\Sigma}_m = \mathbf{C}\rho + (1-\rho)\mathbf{I} \tag{9}$$

1) **C** is a correlation matrix due to the Phylogeny (`ape::vcv(., corr = TRUE)`)
2) $0 \leq \rho \leq 1$ is Pagel's λ: the Phylogenetic signal parameter

Phylogenetic random effects

$$\boldsymbol{\Sigma}_m = \mathbf{C}\rho + (1-\rho)\mathbf{I} \tag{9}$$

1) **C** is a correlation matrix due to the Phylogeny (`ape::vcv(., corr = TRUE)`)
2) $0 \leq \rho \leq 1$ is Pagel's λ: the Phylogenetic signal parameter

This model only generates positive species associations.

Phylogenetic signal

- 1: Fully phylogenetically structured responses
- 0: Normal ("iid") random effects

When it is 0, it does not mean there is nothing going on.

Absence of phylogenetic signal:

- Scale mismatch
- Evolution moves very fast
- Too little information
- Traits are phylogenetically structured
- There are other (flexible) terms in the model
- Model misspecification

Presence of phylogenetic signal:

- Related species have similar "traits" (environmental response)
- Occupy similar environments

Model limitation

This phylogenetic model assumes traits evolve following the Brownian motion model of evolution. This can only generate positive associations.

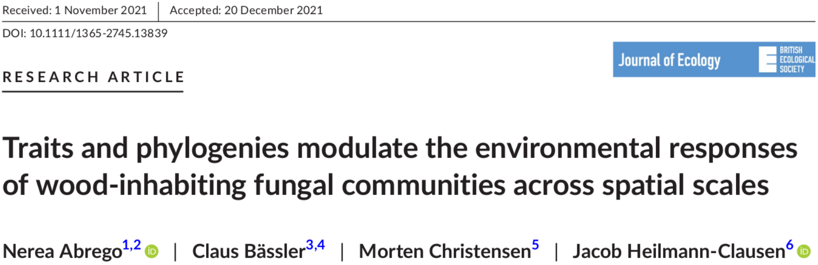
But, competitive exclusion tells us that species evolve to differentiate resource.

- Similar species can (stably) co-occur if they utilize a different resource
- Similar species that utilize the same resource should not (stably) co-occur

The latter results in negative correlations, but no corresponding model for trait evolution has been developed

- unless species do not stably co-occur and/or evolution is still ongoing

Example with fungi data (Abrego 2021)



Received: 1 November 2021 | Accepted: 20 December 2021

DOI: 10.1111/1365-2745.13839

Journal of Ecology BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Traits and phylogenies modulate the environmental responses of wood-inhabiting fungal communities across spatial scales

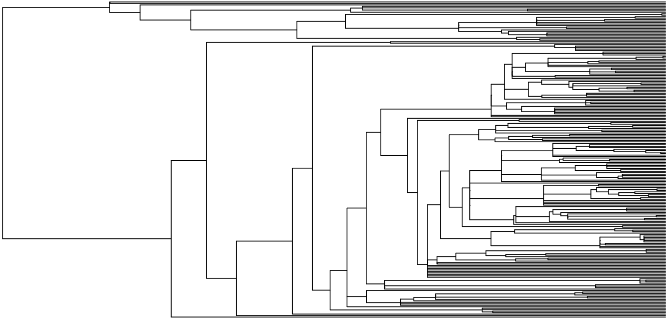
Nerea Abrego[1,2] | Claus Bässler[3,4] | Morten Christensen[5] | Jacob Heilmann-Clausen[6]

Example with fungi data

```r
Y = read.csv("../data/fungiY.csv",)[,-1]
X = read.csv("../data/fungiX.csv")[,-1]
tree = ape::read.tree("../data/fungiTree.txt")
```

- 215 species (after cleaning)
- 1666 sites
- 19 covariates of various kinds

Example with fungi data



Example with fungi data

Phylogenetic models in `gllvm` use a **nearest neighbour approximation**

- We need to set the number of tips to consider on the tree
- The ordering of species matters!

```
covMat <- ape::vcv(tree)
e <- eigen(covMat)
distMat <- ape::cophenetic.phylo(tree)
ord <- gllvm::findOrder(covMat = covMat, distMat = distMat,
species <- colnames(covMat)[ord]
Y <- Y[, species]
covMat <- covMat[species, species]
distMat <- distMat[species, species]
```

Ordering species

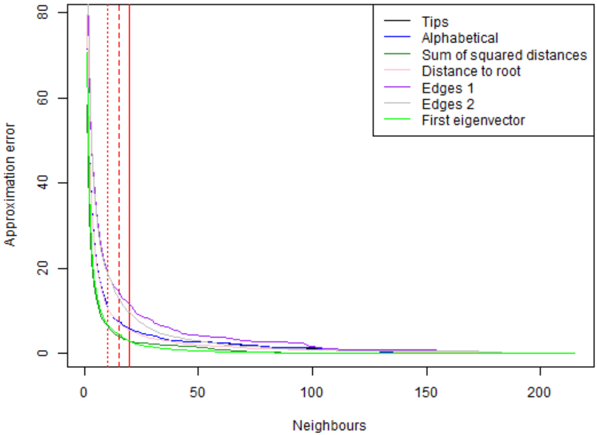


Figure 4: See vignette 7

Example with fungi data

```r
TMB::openmp(parallel::detectCores()-1, autopar = TRUE, DLL = "gllvm")
jsdm5 <- gllvm::gllvm(y = Y, X=X, family = "binomial", num.lv = 0, beta0com = TRUE,
      row.eff = ~(1 | REGION/RESERVE), studyDesign = X[,c("REGION","RESERVE")],
      formula = ~(DBH.CM+AVERDP+I(AVERDP^2)+CONNECT10+TEMPR+PRECIP+log.AREA|1),
      colMat = list(covMat, dist = distMat), nn.colMat = 15, max.iter = 10e3, optim.method = "L-BFGS-B")
```

This takes a while to fit, but is really incredibly fast (it is just a complex model)

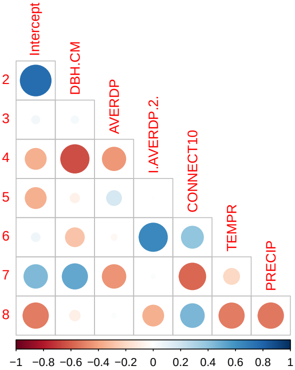
Example with fungi data

```
summary(jsdm5)
```

```
##
## Call:
## gllvm::gllvm(y = Y, X = X, formula = ~(DBH.CM + AVERDP + I(AVERDP^2) +
##     CONNECT10 + TEMPR + PRECIP + log.AREA | 1), family = "binomial",
##     num.lv = 0, studyDesign = X[, c("REGION", "RESERVE")], colMat = list(covMat,
##         dist = distMat), row.eff = ~(1 | REGION/RESERVE), beta0com = TRUE,
##     nn.colMat = 15, max.iter = 10000, optim.method = "L-BFGS-B")
##
## Family:  binomial
##
## AIC:  103171.8 AICc:  103171.8 BIC:  103678.9 LL:  -51539 df:  47
##
## Informed LVs:  0
## Constrained LVs:  0
## Unconstrained LVs:  0
##
## Formula:  ~(DBH.CM + AVERDP + I(AVERDP^2) + CONNECT10 + TEMPR + PRECIP + log.AREA | 1)
## LV formula:  ~ 0
## Row effect:  ~(1 | REGION/RESERVE)
##
## Random effects:
##  Name        Signal Variance Std.Dev Corr
##  Intercept   0.6037 1.0495   1.0244
##  DBH.CM      0.6037 0.0051   0.0715   0.7642
##  AVERDP      0.6037 0.1796   0.4238   0.0529  0.0458
##  I.AVERDP.2. 0.6037 0.0066   0.0815  -0.3550 -0.6454 -0.4397
##  CONNECT10   0.6037 0.0401   0.2003  -0.3544 -0.0711  0.1790 -0.0091
##  TEMPR       0.6037 0.0689   0.2625   0.0626 -0.2879 -0.0321  0.6438  0.3917
##  PRECIP      0.6037 0.0440   0.2098   0.4461  0.5139 -0.4465  0.0148 -0.5701
##  log.AREA    0.6037 0.0140   0.1184  -0.5196 -0.0889  0.0173 -0.3518  0.4538
##
##
##
##
```

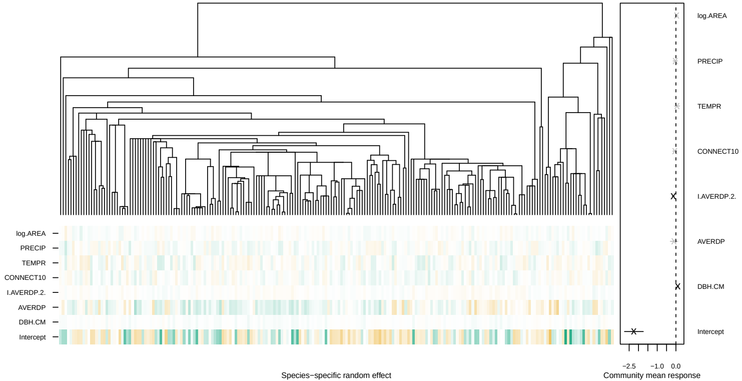
Example with fungi data

```
corrplot::corrplot(cov2cor(jsdm5$params$sigmaB), type = "lower", diag = FA
```



Example with fungi data

```
gllvm::phyloplot(jsdm5, tree)
```



Conclusion:

There is phylogenetic structuring; species environmental responses are more similar if they have a shared evolutionary history.

Of course, this might be covariate dependent (`colMat.rho.struct = "term"`)

Summary

- JSDMs is a framework for analysing species co-occurrence data
- Focussed on prediction, but also suitable for inference
- We can also fit models with non-binary data (e.g., counts or biomass)
- The GLLVM framework is used here to implement JSDM efficiently
- We can incorporate random effects
- Phylogenetically structure species' effects
- Above all: we incorporate correlation of species