Ordination with covariates

Bert van der Veen

Department of Mathematical Sciences, NTNU

Outline

- Studying gradients more deeply
- Constrained ordination
- Concurrent ordination

Outline

- Studying gradients more deeply
- Constrained ordination
- Concurrent ordination

Many bells and whistles to present

Questions so far?

Background

So far: only unconstrained ordination

- Which is fun, but not if you want to assess species-environment relationships
- Here we will focus on including covariates in the model
- Beneficial if you have **sparse** data and cannot estimate species-specific effects
- Will also cover: **residual** ordination, and **concurrent** ordination

Constrained ordination methods

- Redundancy analysis (RDA, Rao 1964)
- Canonical Correspondence Analysis (CCA, ter Braak 1985)

1. These methods filter the variation in the community based on covariates.

2. Although referred to as "ordination" they do not perform dimension reduction!

3. That makes them (essentially) vector GLMs

Constrained ordination

Goal: to determine if (how) environment affects community composition

Problem: many possible drivers (if not, multivariate GLM would do the trick)

- Why are sites different?
- Why do species co-occur (or not)?
- Which components of the environment are most important for the community?

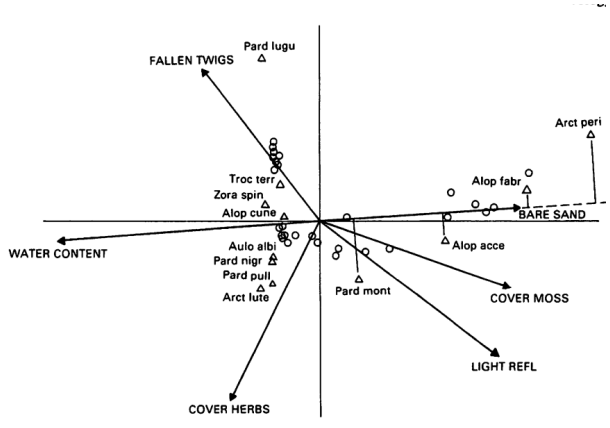


FIG. 1. The distribution of 12 species of hunting spiders caught in pitfall traps in a Dutch dune area.

Figure 1: ter Braak 1986

Constrained ordination

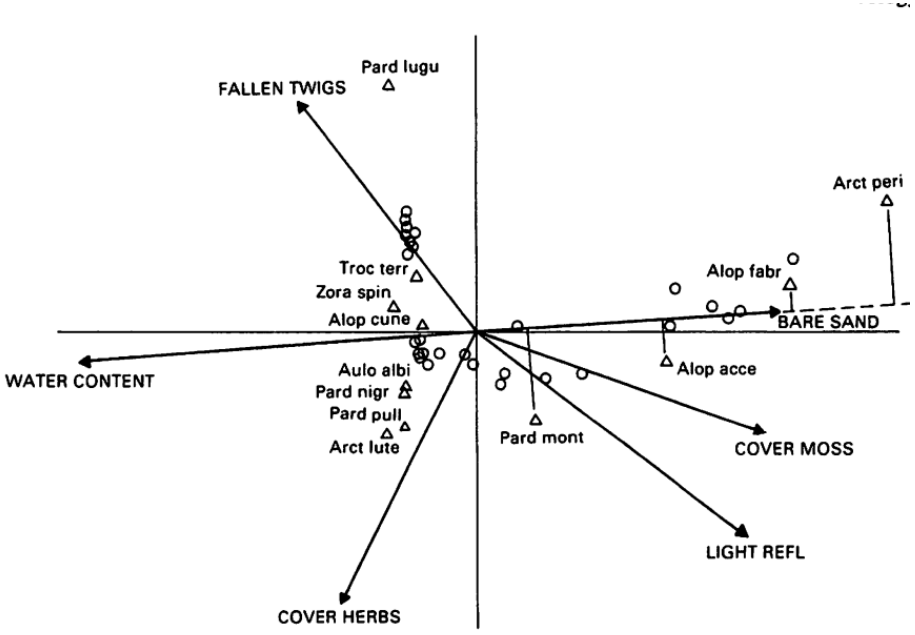


FIG. 1. The distribution of 12 species of hunting spiders caught in pitfall traps in a Dutch dune area.

Figure 2: ter Braak 1986

Now three quantities: so we call this a **triplot**. The arrows show the association of covariates to the axes.

Canonical Correspondence Analysis

- Although RDA was developed much earlier, CCA has been the leading constrained ordination method
- ter Braak (1986) developed CCA as a combination of ordination and regression
- Each axis is restricted (constrained) by covariate information
- CCA approximates Gaussian Ordination (i.e., to the unimodal model, Johnson and Altman, 1999)

Canonical Correspondence Analysis: arrows

The covariate coefficients **B** are referred to as **canonical** coefficients.

- `vegan` does not use these for plotting
- Instead it uses sample correlation coefficients as recommended by ter Braak (1986)
- The canonical coefficients can be "unstable" due to multicollinearity
- In `gllvm`, we do use **B** (more details later)

Model-based methods for constrained ordination

- RR-VGLMs (Yee et al. 1996,2003,2010,2015)
- Row-column interaction models (Hawinkel et al. 2019)
- GLLVMs (van der Veen et al. 2023)

Model-based constrained ordination had been around for a while, but not so often used (lack of user-friendly software).
It is a type of VGLM for sparse data; RR-VGLM.

Software for constrained ordination

In R e.g.

For constrained ordination:

- `vegan` - classical methods
- `VGAM` - only fixed effects
- `glmmTMB` - only random effects
- `gllvm` - easy to use, random or fixed effects

Constrained ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \tag{1}$$

So far, we have assumed that the latent variable is estimated by species information alone: $\mathbf{u}_i = \boldsymbol{\epsilon}_i$

Constrained ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \quad (1)$$

So far, we have assumed that the latent variable is estimated by species information alone: $\mathbf{u}_i = \boldsymbol{\epsilon}_i$

Constrained ordination instead assumes that we also have covariates in the ordination: $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv}$

Constrained ordination: the model

Plugging in $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv}$ we get:

$$\eta_{ij} = \beta_{0j} + \mathbf{x}_i^{lv,\top} \mathbf{B} \boldsymbol{\gamma}_j \tag{2}$$

From this we see that $\boldsymbol{\beta}_j \overset{d}{\approx} \mathbf{B}\boldsymbol{\gamma}_j$

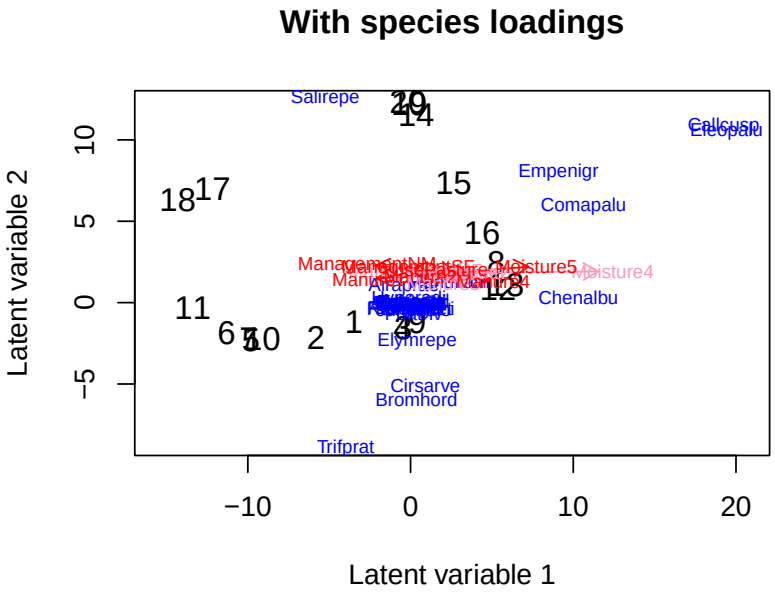
- These are the (reduced rank) approximated species-specific covariate coefficients
- We can extract these, and inspect them with statistical uncertainty
- So we use information across the whole community, to estimate species-specific responses

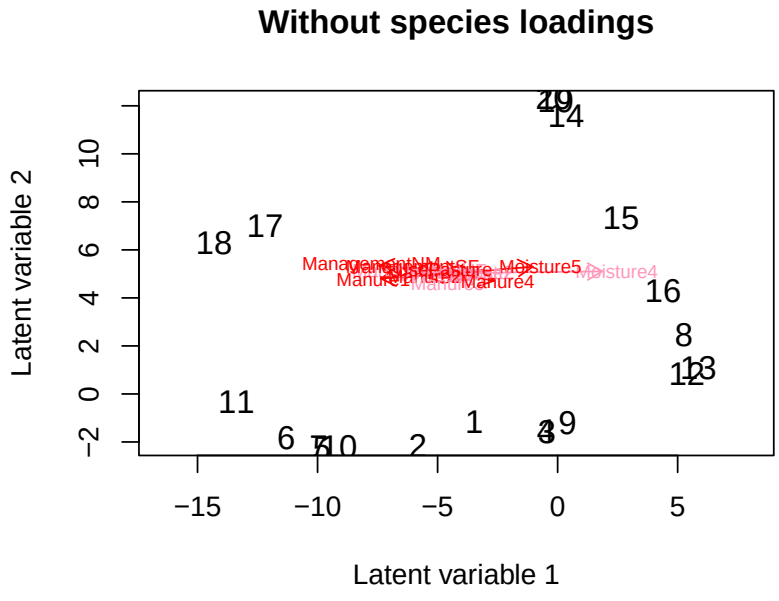
Example: Dune data

Same data as before, now we bring the covariate **in** the ordination, instead of taking information **out** of the ordination.

Example: fit the model

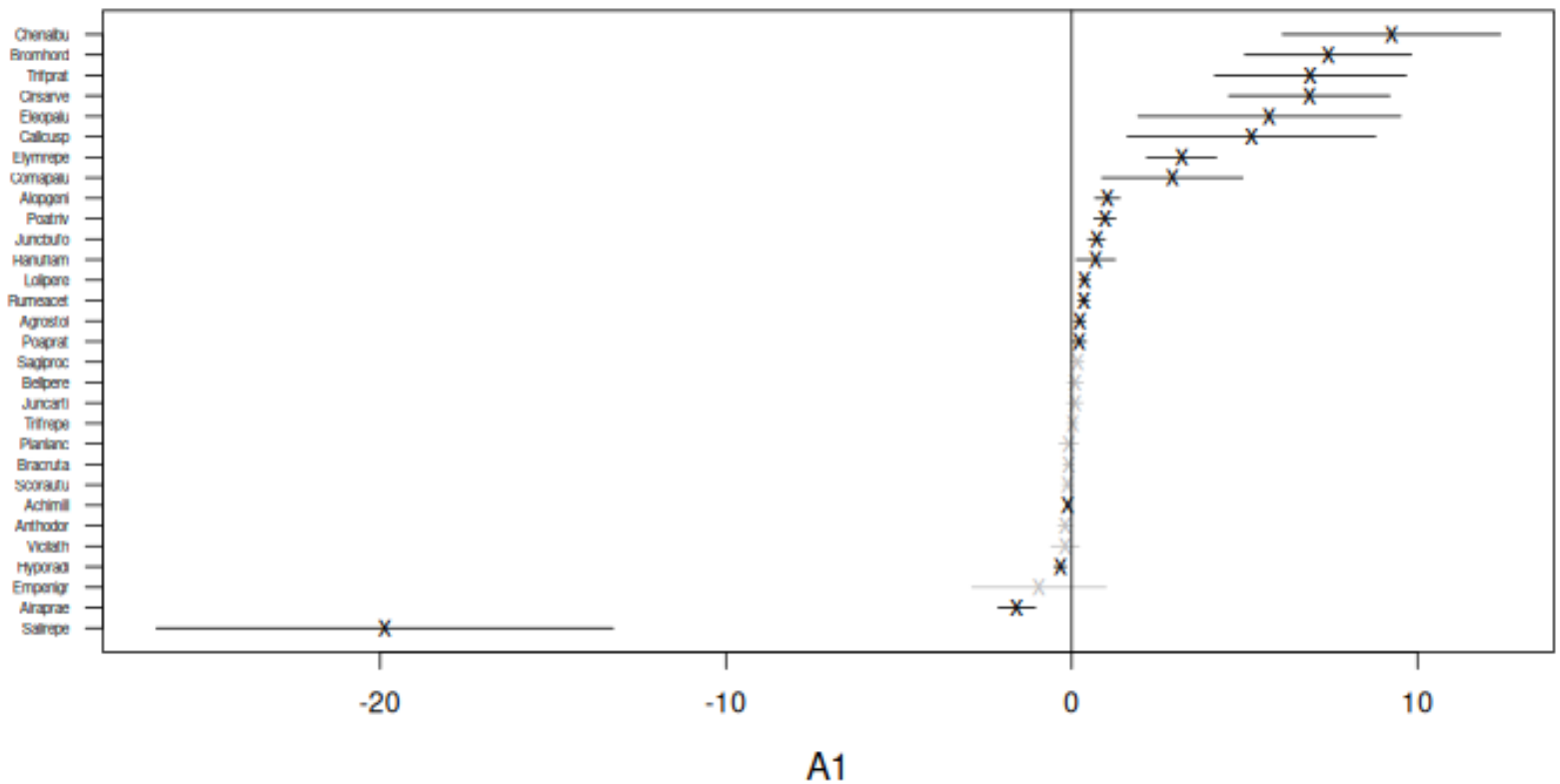
```
model1 <- gllvm(y = Y, X, num.RR = 2, family = "ordinal",
        lv.formula = ~A1 + Management + Moisture + Use + Manure)
```





Example: extracting species-specific coefficients with 95% CI

```
coefplot(model1, which.Xcoef = "A1")
```



Constrained ordination

- Species effects can be retrieved for any covariate
- Extreme results occur, usually due to insufficient data
- GLLVMs picks up on extreme clustering -very- well

Constrained ordination

The first implementation of CO that can be combined with random effects

- Random site effects (outside ordination)
- Random canonical coefficients (more in a few slides)

Common misconception

Post-hoc relating unconstrained ordination axes to environmental covariates is **not** equivalent to a constrained ordination

Also it is bad practice: please do not do it. Instead **adjust your model**.

Hybrid ordination

- Incorporate both constrained and unconstrained ordination
- But without explicit connection
- Default in `vegan` and you can also do it in `gllvm` (use both `num.RR` and `num.lv`)

Concurrent ordination

- In practice, constrained and unconstrained ordination are often combined into an analysis
- Variation not due to the environment is discarded, while potentially of large importance
- *Concurrent ordination* is a new type of ordination method that combines unconstrained and constrained ordination

Concurrent ordination

Concurrent:'existing or happening at the same time' (Oxford's dictionary)

Concurrent ordination

Concurrent:'existing or happening at the same time' (Oxford's dictionary)

1. Suggested in van der Veen et al. (2023)
2. Performs both unconstrained and constrained ordination **simultaneously**

3. Ordination axes have **measured** and **unmeasured** components
4. Covariates **inform** rather than **constrain**
5. Separates out drivers of community composition

Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \tag{3}$$

The model is flexible, $\mathbf{u}_i$ can be all kinds of things.

Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \tag{3}$$

The model is flexible, $\mathbf{u}_i$ can be all kinds of things.

1. $\mathbf{u}_i = \boldsymbol{\epsilon}_i$, **unconstrained**

Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \tag{3}$$

The model is flexible, $\mathbf{u}_i$ can be all kinds of things.

1. $\mathbf{u}_i = \boldsymbol{\epsilon}_i$, **unconstrained**
2. $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv}$, **constrained**

Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \tag{3}$$

The model is flexible, $\mathbf{u}_i$ can be all kinds of things.

1. $\mathbf{u}_i = \boldsymbol{\epsilon}_i$, **unconstrained**
2. $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv}$, **constrained**
3. $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv} + \boldsymbol{\epsilon}_i$, **concurrent**

Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \tag{3}$$

The model is flexible, $\mathbf{u}_i$ can be all kinds of things.

1. $\mathbf{u}_i = \boldsymbol{\epsilon}_i$, **unconstrained**
2. $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv}$, **constrained**
3. $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv} + \boldsymbol{\epsilon}_i$, **concurrent**

Often unconstrained and concurrent ordinations are similar

Concurrent ordination: site scores

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j \tag{3}$$

The model is flexible, $\mathbf{u}_i$ can be all kinds of things.

1. $\mathbf{u}_i = \boldsymbol{\epsilon}_i$, residual
2. $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv}$, marginal
3. $\mathbf{u}_i = \mathbf{B}^\top \mathbf{x}_i^{lv} + \boldsymbol{\epsilon}_i$, conditional

Often unconstrained and concurrent ordinations are similar

Concurrent ordination: the model

Essentially a linear mixed-effects model of the latent variable $\mathbf{u}_i$
(hiearchcally)

Concurrent ordination: the model

Essentially a linear mixed-effects model of the latent variable $\mathbf{u}_i$
(hiearchcally)

We can use the LV-level error variance for a measure of explained variation.

Concurrent ordination with `gllvm`

Example 2: Wadden data

Wadden sea data Dewenter et al. (2023)

- Abundance (counts) or Biomass of macrozoobenthos
- Covariates
- Transects at islands (Norderney, Spiekeroog, Wangerooge)

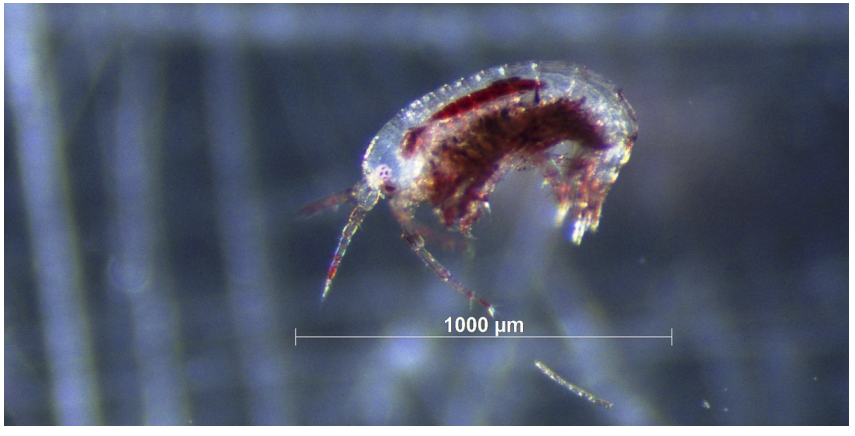


Figure 3: nioz.nl

```
Y <- read.csv("../data/waddenY2.csv")[,-c(1:2)]
Y <- Y[, colSums(ifelse(Y==0,0,1))>2]
X <- read.csv("../data/waddenX.csv")
X[,unlist(lapply(X,is.numeric))] <- scale(X[,unlist(lapply(X,is.numeric))])
```

Example 2: Fit the model

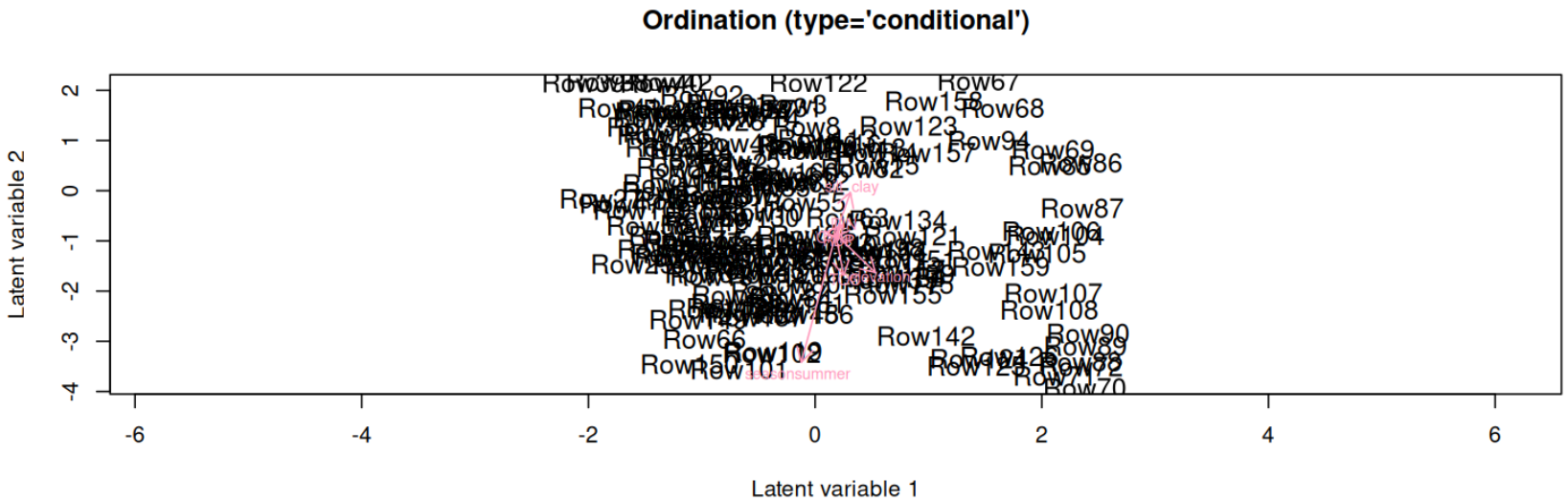
```
model3 <- gllvm(y = Y, X, num.lv.c = 2,
                family = "tweedie", Power = NULL,
          lv.formula = ~elevation+TOC+DIN+RDP+Chl.a+silt_clay+season,
          disp.formula = rep(1,ncol(Y)), n.init = 3, starting.val = "zero")
coef(model3, parm="Cancoef")
```

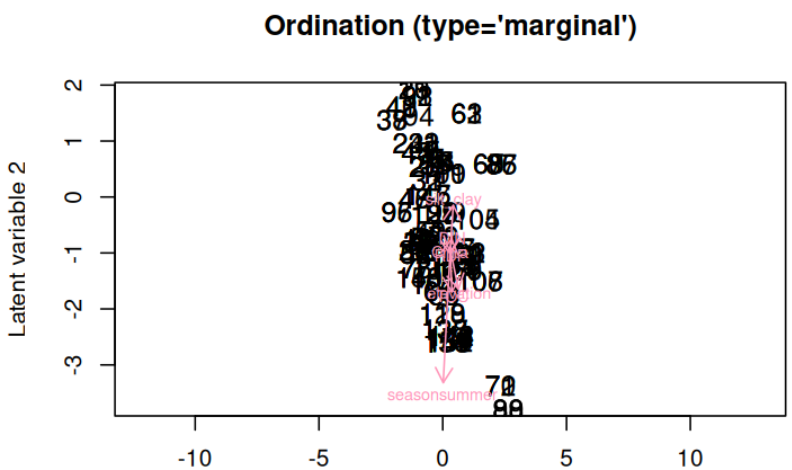
```
##                       CLV1        CLV2
## elevation      0.203985514 -0.44264857
## TOC            0.040909072 -0.47338687
## DIN            0.035673369  0.11267663
## RDP            0.013593778 -0.06123323
## Chl.a         -0.009550114 -0.02692006
## silt_clay      0.075088243  0.48653344
## seasonsummer -0.088791735 -0.74357339
```

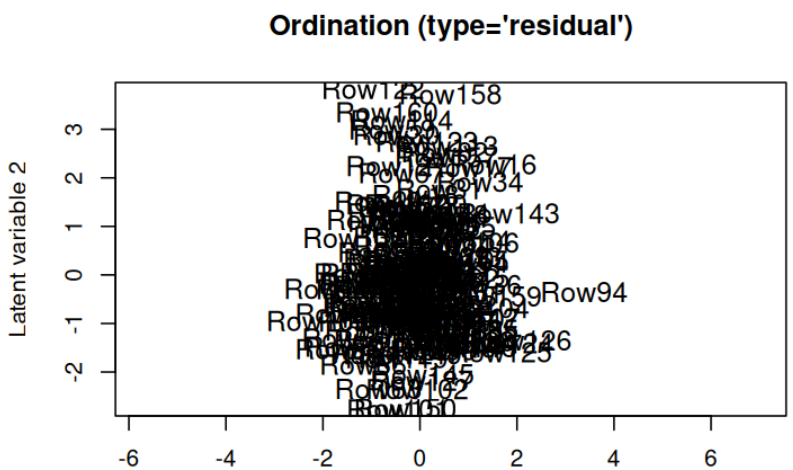
Example 2: Ordination plots of different scores

```
layout(matrix(c(1,1,2,3), 2, 2, byrow=TRUE))
gllvm::ordiplot(model3, type= "conditional", rotate = FALSE)
gllvm::ordiplot(model3, type = "marginal", rotate = FALSE)
gllvm::ordiplot(model3, type = "residual", rotate = FALSE)
```

Example 2: Ordination plots of different scores







The bouncing beta problem

When there is collinearity, and we resample the data, the covariate effect my change a lot (and thus the arrows in the ordination plot). Also if we change the model.

- This is due to estimation instability
- Due to identifiability problems (two covariate effects cannot be separated)
- When we resample, the collinearity structure may change considerably

It is not a problem for the model, but we might want to stabilize things. This is why classical ordination does not use the canonical coefficients.

Random canonical coefficients

We can treat the canonical coefficients as random effects with `randomB` (with constrained and with concurrent ordination)

- This is usually faster
- Treats the "bouncing beta" problem
- Models associations of species due to environment (extracted with `getEnvironCov/r`)

1. LV: coefficients on the same axis have the same variance
   - Shrinkage over LVs
2. P: coefficient of the same covariate have the same variance
   - Shrinkage over LVs and covariates (e.g., for automated model selection)
3. single: all coefficients have the same variance

Random canonical coefficients

We can treat the canonical coefficients as random effects with `randomB` (with constrained and with concurrent ordination)

- This is usually faster
- Treats the "bouncing beta" problem
- Models associations of species due to environment (extracted with `getEnvironCov/r`)

1. LV: coefficients on the same axis have the same variance
   - Shrinkage over LVs
2. P: coefficient of the same covariate have the same variance
   - Shrinkage over LVs and covariates (e.g., for automated model selection)
3. single: all coefficients have the same variance

Random canonical coefficients

We can treat the canonical coefficients as random effects with `randomB` (with constrained and with concurrent ordination)

Consider this an RR-VGLMMs So, we can use the variance of the random effects for proportion of explained variation!

1. LV: coefficients on the same axis have the same variance
   - Shrinkage over LVs
2. P: coefficient of the same covariate have the same variance
   - Shrinkage over LVs and covariates (e.g., for automated model selection)
3. single: all coefficients have the same variance

Random canonical coefficients

We can also incorporate **correlation** of the effects. This is a model where we have:

- Associations due to environment `getEnvironCor`
- Correlation between environment effects (requires `lv.formula` `lme4`-style)
- Which means we can examine the reduced-rank approximated random effects `RandomCoefPlot`

Random canonical coefficients: mathy bits

Define $\mathbf{\Sigma}_{lv}$ as correlation matrix, then:

`randomB = "LV"`: $\mathbf{b}_q \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}\sigma_k^2)$
`randomB = "P"`:
$\mathbf{B} \sim \mathcal{MN}\{\mathbf{0}, \text{diag}(\sigma_1 \dots \sigma_k)\mathbf{\Sigma}\text{diag}(\sigma_1 \dots \sigma_k), \text{diag}(1, \sigma_2^2 \dots \sigma_d^2)\}$

but otherwise the same model. The estimate of $\mathbf{\Sigma}$ will get better with the number of LVs. This is the same matrix as in the VGLMM, just estimated in reduced rank. You can also combine this with full rank random effects in `formula`.

Example 3

```
model4 <- gllvm(y = Y, X = X, num.lv.c = 2, randomB = "P",
          family = "tweedie", Power =  NULL,
          lv.formula = ~(0+elevation+TOC+DIN+RDP+Chl.a+silt_clay+season|1),
          disp.formula = rep(1,ncol(Y)), n.init = 3, maxit = 10e5)
```

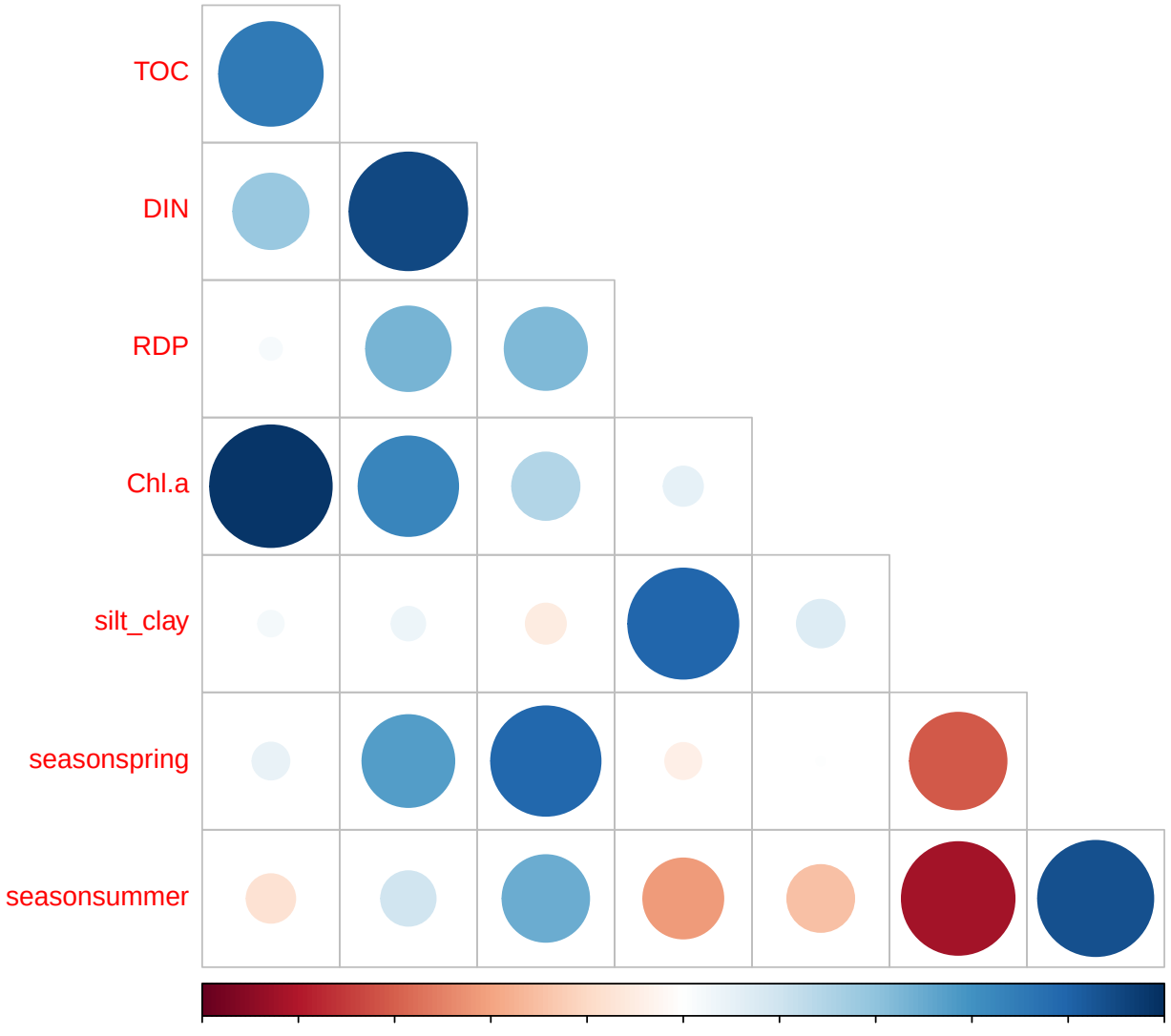
`lv.formula` follows standard random-effects style (correlation for effects in same brackets):

Example 3: summary

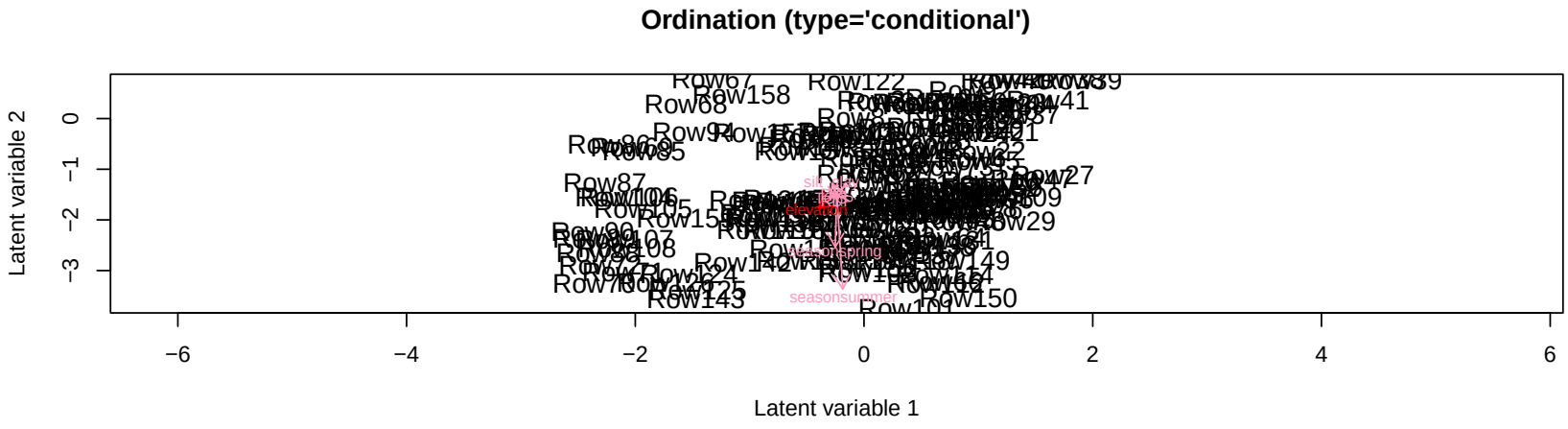
```
summary(model4)
```

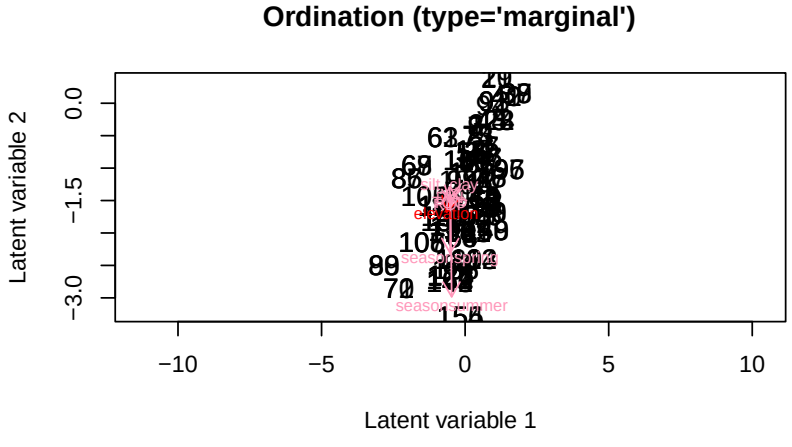
```
##  Name          Std.Dev.CLV1 Variance.CLV1 Std.Dev.CLV2 Variance.CLV2 Corr
##  elevation     0.2480       0.0615        0.6692       0.4478
##  TOC           0.0930       0.0086        0.2509       0.0629          0.7044
##  DIN           0.0621       0.0039        0.1675       0.0281          0.3735
##  RDP           0.0249       0.0006        0.0671       0.0045          0.0323
##  Chl.a         0.0721       0.0052        0.1945       0.0378          0.9700
##  silt_clay     0.1062       0.0113        0.2865       0.0821          0.0436
##  seasonspring 0.0147       0.0002        0.0397       0.0016          0.0901
##  seasonsummer 0.0951       0.0090        0.2566       0.0658         -0.1556
##
##
##
##    0.9087
##    0.4698  0.4447
##    0.6502  0.2985  0.1019
##    0.0750 -0.1061  0.7954  0.1499
##    0.5541  0.7855 -0.0862  0.0051 -0.6128
##    0.1958  0.4927 -0.4205 -0.2939 -0.8331  0.8703
```

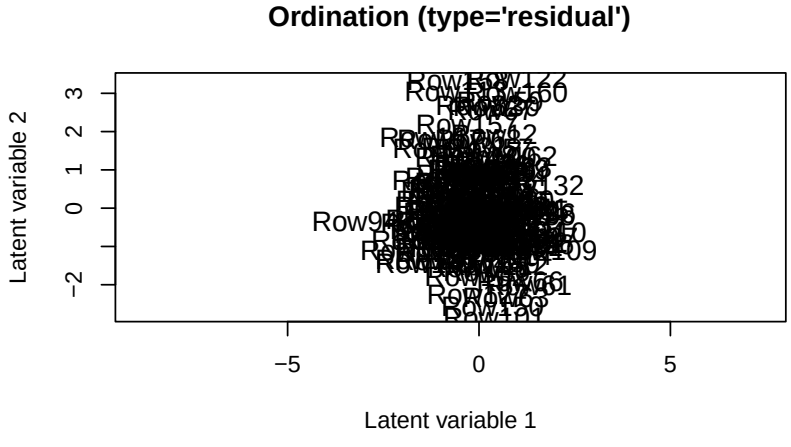
Example 3: correlation plot



Example 3: ordination plots

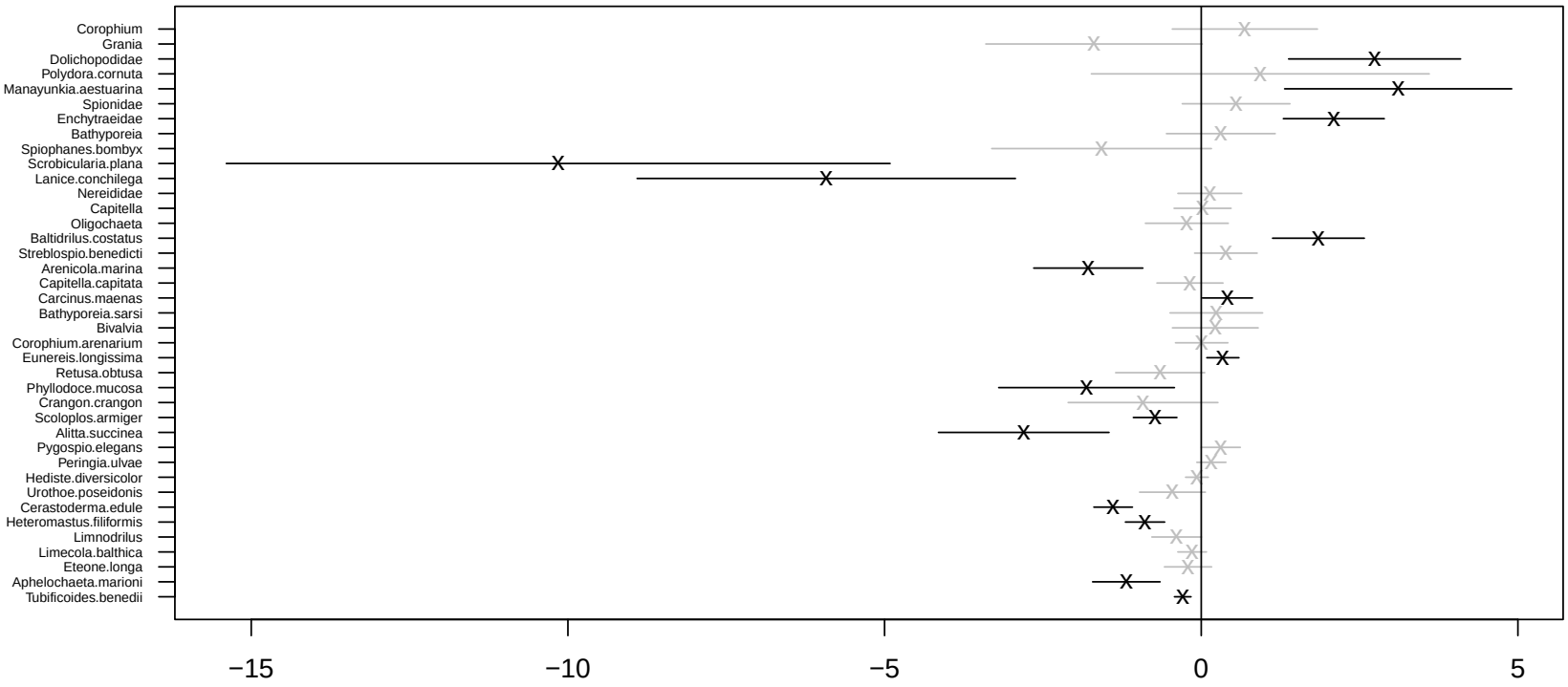






Example 3: Random coefficient plot of species

```
gllvm::randomCoefplot(model4, which.Xcoef="elevation")
```



Model-comparison note

The usual caveats apply:

- No parameters on the boundary
- Nestedness

Note:

- Random effect constrained ordination is nested in unconstrained ordination with `randomB = "LV"`
- Fixed effects constrained ordination is nested in fixed effects unconstrained ordination
- Constrained and unconstrained ordination are nested in concurrent ordination

`anova` and `AIC` can, in principle, be used.

summary

You can test effects over all LVs/covariates with `summary`:

- `summary(model, by = "all")`: test per coefficient (identify horizontal/vertical separation)
- `summary(model, by = "terms")`: per covariate across LVs ("does the ordination depend on **this** covariate")
- `summary(model, by = "LV)`: test per LV ("does LV depend on covariates at all?")

Wald statistics have low power (compared to LRT), and make strong (but usually) reasonable assumptions.

Summary

- Ordination with covariates has three flavours in GLLVM:
  - Residual ordination (covariate outside of the ordination)
  - Constrained ordination
  - Concurrent ordination (simultaneous unconstrained and constrained)
- In each of these we can apply conditioning for a partial ordination; with random effects or fixed effects
- Random canonical coefficients can be implemented via the `randomB` argument, for reduced rank effects
- Via `num.RR` you can also do fixed effects unconstrained ordination