# Ordination with covariates

Bert van der Veen

Department of Mathematical Sciences, NTNU

# Outline

- Background
- Constrained ordination
- Concurrent ordination

# 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
  - One form of that **constrained** ordination
- Will also cover: **residual** ordination, and **concurrent** ordination

# Residual ordination: the model

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

Covariates with species-specific coefficients ("conditioning")

- No longer an unconstrained ordination: covariates are involved
- For binary data it is a JSDM (more later)
- We can also use it to adjust the ordination (take an effect out)
- We estimate species-specific effect $\boldsymbol{\beta}_j$ so need a good amount of data

# 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?

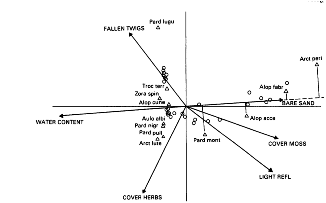


FIG. 1. The distribution of 12 species of hunting spiders caught in pitfall traps in a Dutch dune area.

Figure 1: ter Braak 1986

# Constrained ordination

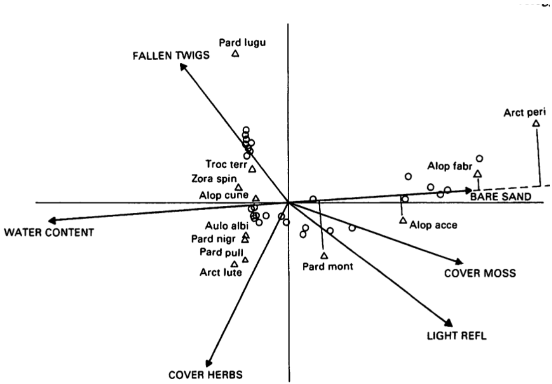



FIG. 1. The distribution of 12 species of hunting spiders caught in pitfall traps in a Dutch dune area.

Figure 2: ter Braak 1986

Now three quantities: so we call this a **triplot**. The arrows show

# Methods for constrained ordination

- Redundancy Analysis (Rao 1964)
- Canonical Correspondence Analysis (ter Braak 1986)
- RR-GLMs (Yee et al. 1996,2003,2010,2015)
- Row-column interaction models (Hawinkel et al. 2019)
- GLLVMs (van der Veen et al. 2023)

# 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)

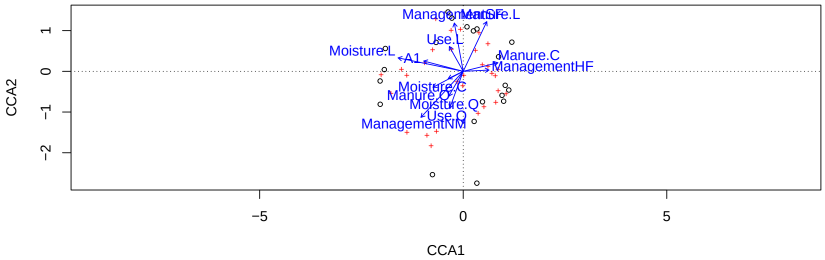
# Canonical Correspondence Analysis: arrows

The covariate coefficients $\mathbf{B}$ 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 $\mathbf{B}$ (more details later)

# Canonical Correspondence Analysis example

```
data(dune,dune.env, package = "vegan");Y=dune;X=dune.env
cca <- vegan::cca(Y, X)
vegan::ordiplot(cca)
```




Canonical coefficients can be retrieved `vegan::scores(cca, display = "reg")`

# Different scores

The CCA algorithm gives raise to two sets of site scores:

1) Weighted average (WA) scores
2) Linear combination (LC) scores

WA scores are usually recommended for plotting (Palmer, 1993)

# Constrained ordination

## In R e.g.

For constrained ordination:

- `vegan` - classical methods
- `VGAM` - cool algorithm, faster than `gllvm`, but not so easy to use (and no random effects)
- `gllvm` - easy to use

# Constrained ordination: the model

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

So far, we have assumed $\mathbf{z}_i = \boldsymbol{\epsilon}_i$
Constrained ordination instead assumes $\mathbf{z}_i = \mathbf{B}^\top \mathbf{x}_i$

# Constrained ordination: the model

Plugging in $\mathbf{z}_i = \boldsymbol{\epsilon}_i$ we get:

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

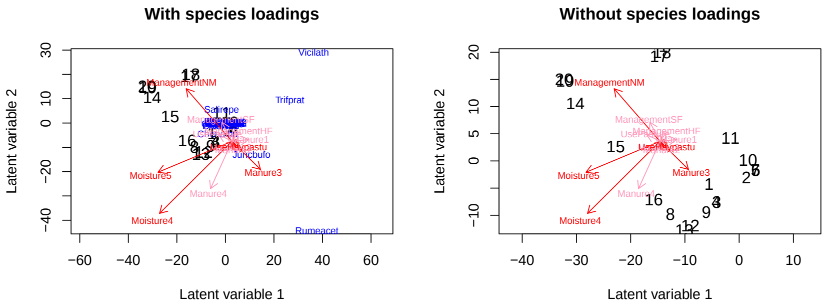
From this we see that $\boldsymbol{\beta}_j \overset{d}{\approx} \mathbf{B}\boldsymbol{\gamma}_j$

- 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

# Constrained ordination with `gllvm`

# Example with Dune data

```
X[,1] <- scale(X[,1]) # always center/scale your covariates for gllvm
X[,c(2:5)] <- data.frame(lapply(X[,c(2:5)],factor,ordered=FALSE)) # I do n
cord <- gllvm::gllvm(y = Y, X, num.RR = 2, family = "ordinal",
            starting.val="res",zeta.struc="common", seed = 2160)
```



# Example with Dune data

```r
summary(cord)
```

```
##
## Call:
## gllvm::gllvm(y = Y, X = X, family = "ordinal", num.RR = 2, zeta.struc = "common",
##     seed = 2160, starting.val = "res")
##
## Family:  ordinal
##
## AIC:  1121.67 AICc:  1182.297 BIC:  1649.302 LL:  -440.8 df:  120
##
## Informed LVs:  0
## Constrained LVs:  2
## Unconstrained LVs:  0
##
## Formula:  ~ 1
## LV formula:  ~A1 + Moisture + Management + Use + Manure
##
## Coefficients LV predictors:
##                    Estimate Std. Error z value Pr(>|z|)
## A1(CLV1)            0.27321    0.07951   3.436 0.000590 ***
## Moisture2(CLV1)    -0.03791    0.07702  -0.492 0.622632
## Moisture4(CLV1)    -1.88156    0.24130  -7.798 6.31e-15 ***
## Moisture5(CLV1)    -3.05399    0.21779 -14.023  < 2e-16 ***
## ManagementHF(CLV1)  0.22853    0.18213   1.255 0.209579
## ManagementNM(CLV1) -1.81691    0.02942 -61.761  < 2e-16 ***
```

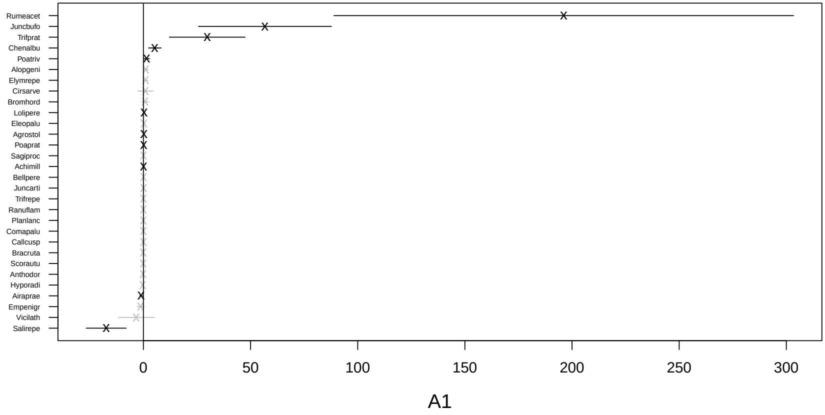
# Example with Dune data

```r
summary(cord, by = "terms")
```

```
##
## Call:
## gllvm::gllvm(y = Y, X = X, family = "ordinal", num.RR = 2, zeta.struc = "common",
##     seed = 2160, starting.val = "res")
##
## Family:  ordinal
##
## AIC:  1121.67 AICc:  1182.297 BIC:  1649.302 LL:  -440.8 df:  120
##
## Informed LVs:  0
## Constrained LVs:  2
## Unconstrained LVs:  0
##
## Formula:  ~ 1
## LV formula:  ~A1 + Moisture + Management + Use + Manure
##
## Coefficients LV predictors:
##                      Estimate Std. Error X2 value  Pr(>X2)
## A1(CLV1)              0.27321    0.07951    13.17 0.001382 **
## Moisture2(CLV1)      -0.03791    0.07702    14.47 0.000722 ***
## Moisture4(CLV1)      -1.88156    0.24130   135.00  < 2e-16 ***
## Moisture5(CLV1)      -3.05399    0.21779   587.32  < 2e-16 ***
## ManagementHF(CLV1)    0.22853    0.18213    13.80 0.001009 **
## ManagementNM(CLV1)   -1.81691    0.02942  8178.36  < 2e-16 ***
```

# Example with Dune data

```
gllvm::coefplot(cord, 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 unconstraied 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` 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 discared, 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. (accepted MEE)
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

$$\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j \tag{4}$$

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

# Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j \tag{4}$$

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

# Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j \tag{4}$$

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

1. $\mathbf{z}_i = \epsilon_i$, **unconstrained**

# Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j \tag{4}$$

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

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

# Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j \tag{4}$$

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

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

# Concurrent ordination: the model

$$\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j \tag{4}$$

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

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

Often unconstrained and concurrent ordinations are similar

# Concurrent ordination: site scores

$$\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j \tag{4}$$

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

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

Often unconstrained and concurrent ordinations are similar

# Concurrent ordination: the model

Essentially a linear mixed-effects model of $\mathbf{z}_i$

(pending an improved software implementation)

# Concurrent ordination with `gllvm`

# Example with Dune data

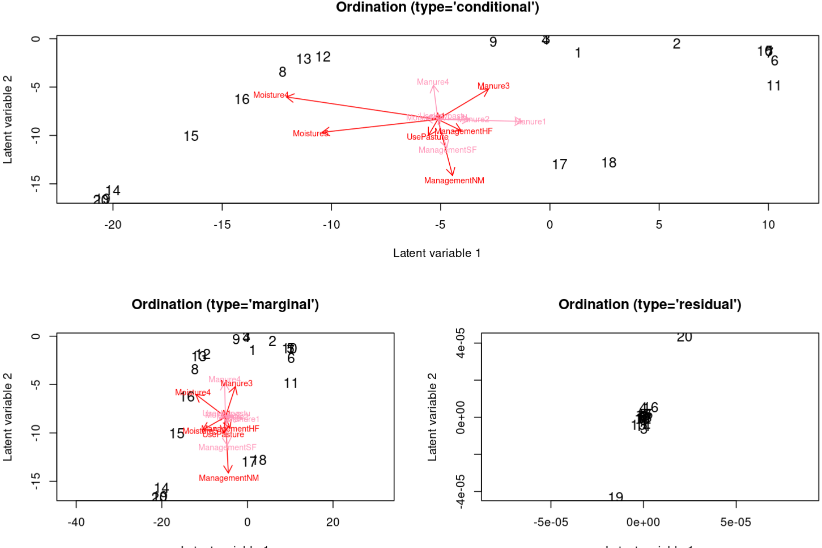
```
cnord <- gllvm::gllvm(y = Y, X, num.lv.c = 2, family = "ordinal",
            starting.val="res", seed = 5882)
coef(cnord, parm="Cancoef")
```

```
##                    CLV1         CLV2
## A1             0.11649617  0.437712324
## Moisture2     -0.28164717  0.071538765
## Moisture4     -2.59820715  0.868266220
## Moisture5     -3.15696632 -0.799328483
## ManagementHF   0.57700088 -0.628848908
## ManagementNM   0.38681278 -3.325594446
## ManagementSF   0.23562939 -1.681882745
## UseHaypastu    0.14173285  0.141490076
## UsePasture    -0.23444726 -0.905167360
## Manure1        1.72479564 -0.094193816
## Manure2        0.73338398  0.004163264
## Manure3        1.15889421  1.566272014
```

## Example with Dune data

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

# Example with Dune data



# Random canonical coefficients

We can treat the canonical coefficients as random with `randomB`

- This is usually faster
- Treats the "bouncing beta" problem
- Models correlation between species due to environment

1) LV: canonical coefficients of the same ordination axis come from the same distribution

- Shrinkage over LVs

2) P: canonical coefficients of the same covariate come from the same distribution

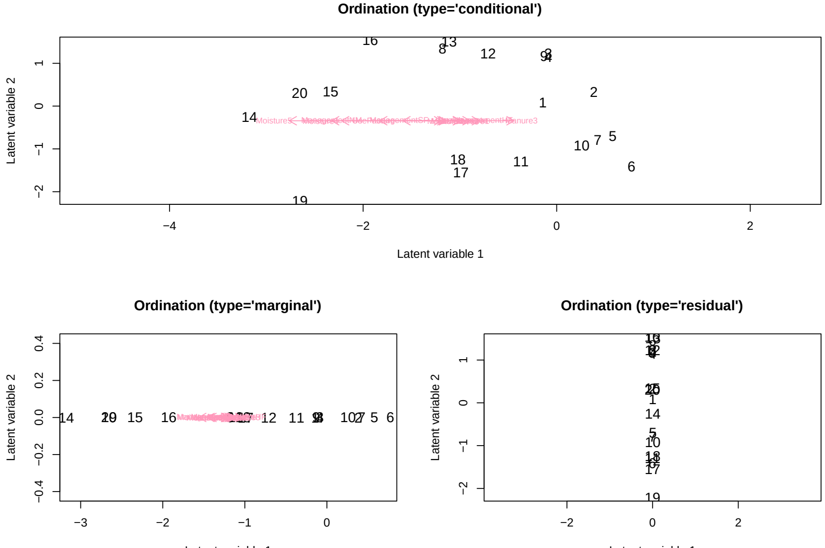
- Shrinkage over covariates

3) single: all come from the same distribution

# Example with Dune data

```r
cnord2 <- gllvm::gllvm(y = Y, X = X, num.lv.c = 2, family = "ordinal",
            starting.val="res", randomB="LV", seed = 318)
```

# Example with Dune data



# Example with Dune data

```r
summary(cnord2)
```

```
##
## Call:
## gllvm::gllvm(y = Y, X = X, family = "ordinal", num.lv.c = 2,
##     randomB = "LV", seed = 318, starting.val = "res")
##
## Family:  ordinal
##
## AIC:  1246.197 AICc:  1285.797 BIC:  1681.493 LL:  -524.1 df:  99
##
## Informed LVs:  2
## Constrained LVs:  0
## Unconstrained LVs:  0
## Residual standard deviation of LVs:  0.6789 1.7596
##
## Formula:  ~ 1
```

# Example with Dune data

```
coef(cnord2, "Cancoef")
```

```
##                    CLV1          CLV2
## A1            0.03690025  1.890662e-09
## Moisture2     0.05399321  6.163276e-10
## Moisture4    -1.00360899  1.632500e-09
## Moisture5    -2.21711704  1.600129e-09
## ManagementHF  0.51842381 -5.363600e-10
## ManagementNM -1.39261919 -1.756312e-09
## ManagementSF -0.52927193  2.625725e-09
## UseHaypastu   0.30077712  1.333141e-09
## UsePasture   -0.80856897 -2.477133e-11
## Manure1       0.32530176 -1.219668e-09
## Manure2       0.24723326 -9.307297e-10
## Manure3       0.90499761  9.808843e-10
## Manure4       0.06690356 -7.178333e-11
```

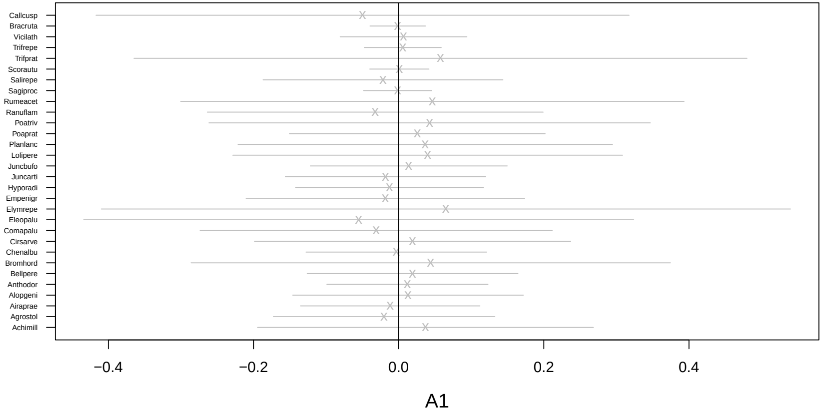
# Example with Dune data

```
gllvm::getPredictErr(cnord2)$b.lv
```

```
##             [,1]         [,2]
##  [1,] 0.1180142 2.720757e-05
##  [2,] 0.2544217 2.720757e-05
##  [3,] 0.6376393 2.720757e-05
##  [4,] 0.8998611 2.720757e-05
##  [5,] 0.3515302 2.720757e-05
##  [6,] 0.6433694 2.720757e-05
##  [7,] 0.4528023 2.720757e-05
##  [8,] 0.2597520 2.720757e-05
##  [9,] 0.3952956 2.720757e-05
## [10,] 0.4672414 2.720757e-05
## [11,] 0.4555743 2.720757e-05
## [12,] 0.5920449 2.720757e-05
## [13,] 0.5473763 2.720757e-05
```

# Example with Dune data

```
gllvm::randomCoefplot(cnord2, which.Xcoef="A1")
```



# Summary

- Ordination with covariates has three flavours in GLLVM:
  - Residual ordination (actually, not really an ordination with covariates)
  - Constrained ordination
  - Concurrent ordination (combining unconstrained and constrained)
- Random canonical coefficients via the `randomB` argument
- Many other useful functions for these models