The unimodal response model

Bert van der Veen

Department of Mathematical Sciences, NTNU

Questions so far?

Unimodal responses models and ordination

There has been a lot of research on the unimodal response model in multivariate analysis

- ▶ Whittaker (1956) introduced the idea
- ▶ MacArthur (1970) formalized it
- ▶ Gauch et al. (1974) proposed "Gaussian ordination" but it did not work
- ▶ ter Braak (1985) popularized the concepts for ordination in terms of CA
- ▶ Minchin (1987) and others opposed the idea of a symmetric model

Unimodal responses models and ordination

There has been a lot of research on the unimodal response model in multivariate analysis

- ▶ Whittaker (1956) introduced the idea
- ▶ MacArthur (1970) formalized it
- ▶ Gauch et al. (1974) proposed "Gaussian ordination" but it did not work
- ▶ ter Braak (1985) popularized the concepts for ordination in terms of CA
- ▶ Minchin (1987) and others opposed the idea of a symmetric model

There has been no succesful ordination method that explicitly incorporates a response model

The consensus

▶ PCA for short gradients (if you zoom in far enough, its linear)

▶ DCA for long gradients

▶ NMDS for robustness to deviations from unimodal curves

The consensus

▶ PCA for short gradients (if you zoom in far enough, its linear)
▶ DCA for long gradients
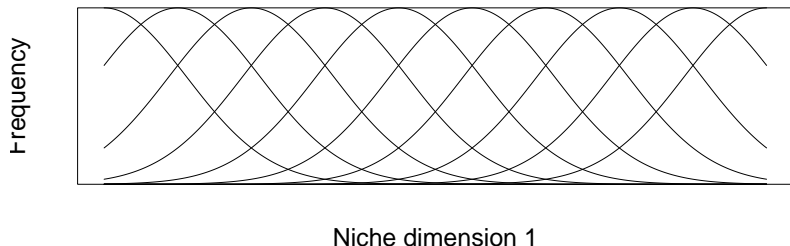▶ NMDS for robustness to deviations from unimodal curves

Based on how "good" an ordination looks: PCA gives horseshoe and CA gives arch.

Terminology

"Gaussian" = quadratic

Both are unimodal

Species packing


Niche dimension 1

- ▶ "Species packing" MacArthur (1969)
  - ▶ Competitive exclusion
  - ▶ Limiting similarity
  - ▶ Leads to CA assumptions (uniform optima, equal maxima, equal tolerances)

- Is the species packing assumption realistic?
- Optimal conditions are usually species-specific
- For example: temperature, space (environment or resources)
- But so is the tolerance
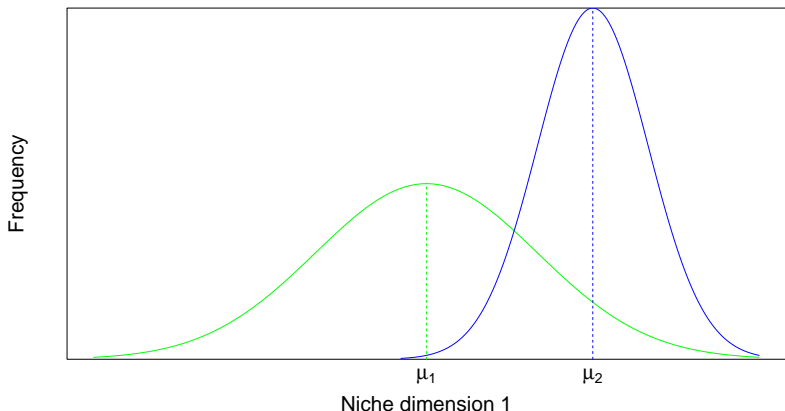
▶ Niche width
▶ Commonly assumed equal (not here)
▶ What is the probability of niche occurrence?

Niche dimension 1
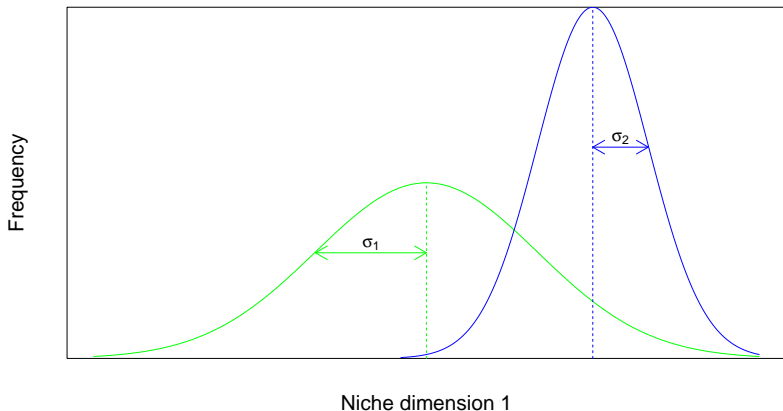
GLLVM with unimodal response models

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

Quadratic term (positive diagonal matrix)

This allows us to:

▶ Calculate species optima
▶ Calculate species tolerances
▶ Estimate gradient length
▶ Provide a more ecologically plausible ordination method

Niche width

- ▶ "Tolerance" is a measure of niche width
- ▶ Ecologically, we expect this to be different for species (some specialist, some generalist)
- ▶ CA requires the assumption that this is the same for all species
- ▶ We provide three options
    - ▶ the same ("equal tolerances", species packing)
    - ▶ the same per LV ("common tolerances")
    - ▶ not the same ("unequal tolerances")
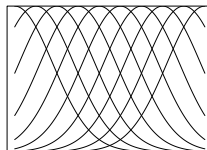- ▶ Species maxima are usually species-specific (but not in CA)

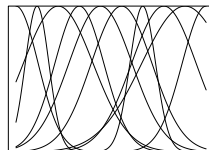## What does that mean?
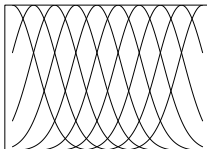

Equal tolerances   **Common tolerances**   **Unequal tolerances**
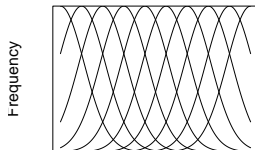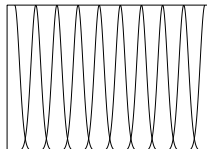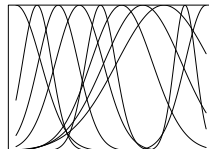
Rare species
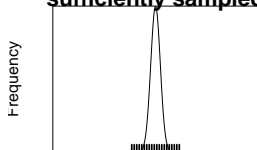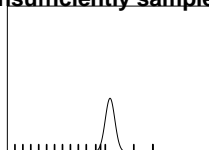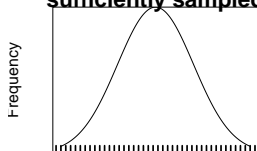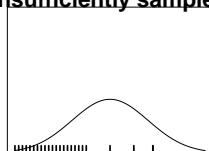
Consequences

▶ Unconstrained Quadratic Ordination (UQO) is a very complex
  method
▶ The most complex ecological ordinaiton method to date
  ▶ Still, it assumes a symmetric curve, which some deem an unrealistic
    ecological model
  ▶ But, no one has done better
▶ Unequal tolerances is not achievable for most datasets
▶ Equal or common tolerances is most suitable in practice
  ▶ An equal tolerances model can be fitted with an ordinary
    GLLVM

On non-linear distortions in GLLVMs

"The arch effect is simply a mathematical artifact, corresponding to no real structure in the data. It arises because the second axis (canonical variate) of RA is constrained to be uncorrelated with the first axis, but is in no way constrained to be independent of it."

(Hill and Gauch, 1980)

Recap: a-priori in GLLVMs the latent variables are normally distributed and uncorrelated, thus independent and orthogonal.

On non-linear distortions in GLLVMs

$$\mathrm{cov}(\mathbf{u}_i^\top \boldsymbol{\gamma}_j, \mathbf{u}_i^\top \mathbf{D}_j \mathbf{u}_i) = 0 \qquad (2)$$

The linear and quadratic terms are uncorrelated: it is an orthogonal polynomial (useful for interpretation and convergence).

This also means that the linear and quadratic terms can be separately visualized in an ordination plot, and lead to distortions. So we cannot use a biplot.

On plotting a quadratic response model

▶ One method is to plot the predicted curves per dimension
▶ For an ordination diagram we can plot the optima with site scores
▶ Tolerances can be used to draw suitability regions
▶ Very large optima (due to near linear responses) need to be drawn as arrows (`ordiplot(.)` does this by default)

On non-linear distortions in GLLVMs

▶ Horseshoe effect does not exist in GLLVM
▶ $\mathbb{E}(\mathbf{u}_i|\mathbf{y}_i)$ is more flexible than the (prior) normality assumption
▶ The model will always attempt to capture the data as well as possible
▶ Attitude: if we see something we do not like, we adjust the model
  ▶ I.e., if we see quadratic curvature, we apply a quadratic response model
  ▶ An ordination plot for quadratic response model cannot exhibit quadratic curvature

On non-linear distortions in GLLVMs

▶ A GLLVM without quadratic response model can exhibit non-linear patterns

▶ We can get into situations where linear terms approximate quadratic terms

▶ This might especially happen if the variance due to the quadratic term is larger than due to the linear term

The species associations matrix of a quadratic GLLVM is:

$$\mathbf{\Sigma} = \mathbf{\Gamma\Gamma}^\top + \sum_{q=1}^{d} 2\mathbf{\Delta\Delta}^\top \tag{3}$$

Quadratic model as row-effect

The quadratic response model with equal, or common, tolerance assumption is:

$$\eta_{ij} = \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j - \mathbf{u}_i^\top \mathbf{D} \mathbf{u}_i \qquad (4)$$

which we can write as a linear model with site-specific random effect:

$$\eta_{ij} = \alpha_i + \beta_{0j} + \mathbf{u}_i^\top \boldsymbol{\gamma}_j, \qquad \text{where } \alpha_i = \mathbf{u}_i^\top \mathbf{D} \mathbf{u}_i \qquad (5)$$

because the quadratic term does not depend on the species index. Under the common/equal tolerance assumption, the quadratic term only affects the total abundance at sites.

Quadratic model as row-effect

But the site random effect $\alpha_i \sim \mathcal{N}(\mathbf{0}, \sigma^2)$ while $\mathbf{u}_i^\top \mathbf{D} \mathbf{u}_i \sim d_q \chi_1^2$.

▶ The random row effect can **approximate** a quadratic term

▶ There is a distributional discrepancy; the quadratic form is generalized χ^2 distributed

  ▶ and χ^2 under equal tolerances

▶ A-posteriori (after the model has seen the data) this is not (very) relevant

▶ This is also why random site effects can accommodate non-linearity in the ordination

How to interpre the ordination plot

▶ On an ordination plot we can usually interpret species' locations as optima

▶ Instead of as the main direction of increase as in (for example) PCA

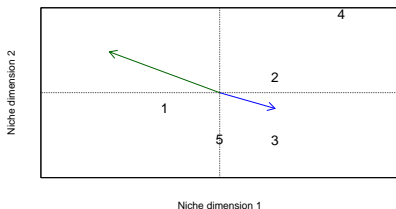▶ You need at least a random site effect in the model
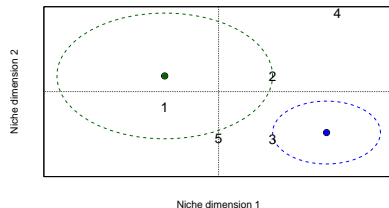

Figure 1: Species effects as increase direction

Figure 2: Species effects as centroids

Example: spider data

- ▶ Data by van der Aart and Smeek-Enserink (1975)
- ▶ 12 species of hunting spider at 100 sites (pitfall traps)
- ▶ 26 predictor variables at 28 of the sites (soil dry matter content, reflection, amount of fallen leaves and such)
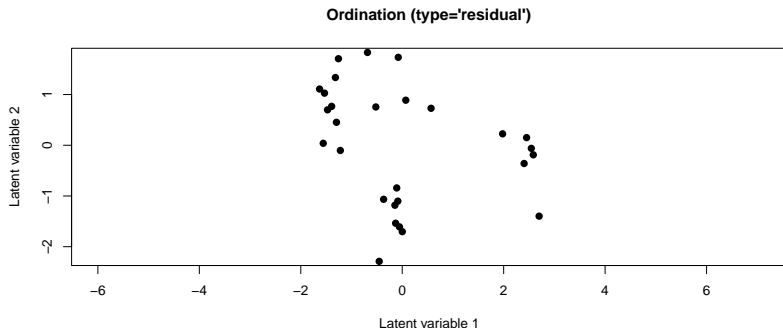- ▶ A classical dataset

Example: non-linear distortion

```
model1 <- gllvm(Y, num.lv = 2, family = "poisson")
ordiplot(model1, symbols = TRUE, pch = 16)
plot(model1, which = 1)
```


Ordination (type='residual')

Example: non-linear distortion (2)

To deal with it: we adjust the model.

```
model2 <- gllvm(Y, num.lv = 2, family = "poisson", row.eff = "random")
ordiplot(model2, symbols = TRUE, pch = 16, main = "With row effect")
```
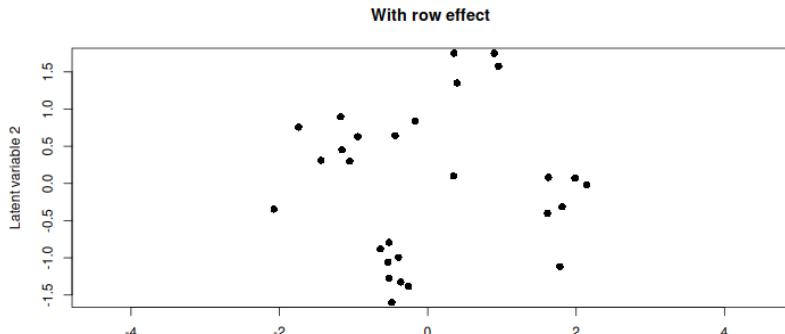
Example: non-linear distortion (2)

To deal with it: we adjust the model.
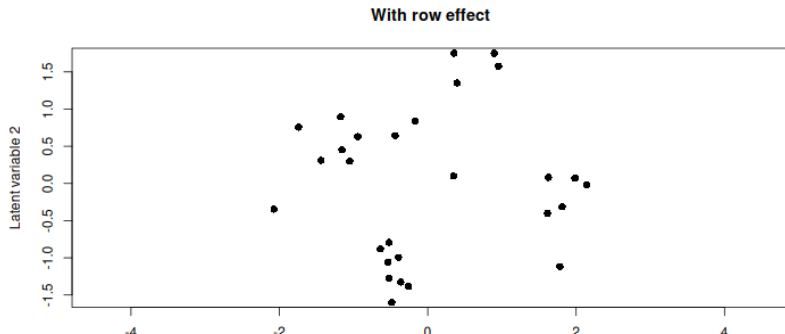
```
model2 <- gllvm(Y, num.lv = 2, family = "poisson", row.eff = "random")
ordiplot(model2, symbols = TRUE, pch = 16, main = "With row effect")
```


With row effect

## Example: information criteria confirms

```
AIC(model1, model2)
```

```
##        df      AIC
## model1 35 1761.655
## model2 36 1754.216
```

Gradient length

- ▶ One of the popular features of DCA: axes in terms of gradient length
- ▶ A unit tolerance curve falls and raises in about 4 units (Hill and Gauch 1980)
- ▶ To me, their method is a bit unclear
- ▶ But, we can calculate it in our own way (van der Veen et al. 2021)

Std. normal distribution
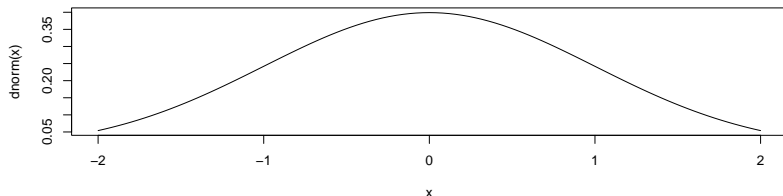
Gradient length

Gradient length is (the width of the standard normal for the LVs) is calculated as:

$$4\mathbf{G}^{\frac{1}{2}} \tag{6}$$

▶ For common tolerances, $\mathbf{G} = 2\mathbf{D}$
▶ Thus, gradient length is $\frac{4}{t}$
  ▶ t is the tolerance
  ▶ Note that van der Veen et al (2021) has an error
  ▶ So that curves are unit width
▶ For unequal tolerances we need to choose: mean or median?
▶ Ultimately, we can rescale in whatever way we want: its all relative

Turnover

We can calculate the (average) rate of turnover as (rate at which species appear-disappear):

$$2Q(\alpha; t) \qquad \text{(at least with a log-link function)} \tag{7}$$

▶ $Q(\cdot)$ is the **quantile** function of a normal distribution
  ▶ At least, that is (more or less) what it is for a log-link
▶ α is some level of error (since we need to cut off somewhere)
▶ t is the tolerance
▶ van der Veen et al. (2021) chose about $\alpha = .999$

Turnover

We can calculate the (average) rate of turnover as (rate at which species appear-disappear):

$$2Q(\alpha; t) \qquad \text{(at least with a log-link function)} \tag{7}$$

▶ $Q(\cdot)$ is the **quantile** function of a normal distribution
    ▶ At least, that is (more or less) what it is for a log-link
▶ α is some level of error (since we need to cut off somewhere)
▶ t is the tolerance
▶ van der Veen et al. (2021) chose about $\alpha = .999$

Note: we are working on a new framework that might incorporate a different definition.

When to go unimodal

▶ Quadratic can accommodate linear, not the other way around

▶ Linear = very wide unimodal curve

▶ Quadratic is more complex and involved to fit

▶ Simplify the model if possible

▶ Usually only when you have enough data/information can you fit the quadratic

　　▶ And then usually with common tolerances

When to go unimodal

▶ Quadratic can accommodate linear, not the other way around
▶ Linear = very wide unimodal curve
▶ Quadratic is more complex and involved to fit
▶ Simplify the model if possible
▶ Usually only when you have enough data/information can you fit the quadratic
  ▶ And then usually with common tolerances

If we misspecify the model as linear, we can miss the dominat gradient altogether.

Example 2: Swiss alpine plants

The Swiss alpine community, sampled on an elevation gradient.

```
Y <- read.csv("../data/alpineY.csv")[,-1]
X <- read.csv("../data/alpineX.csv")[,-c(1:3)]
X <- X[rowSums(Y)>0,]
Y <- Y[rowSums(Y)>0,]
```

Example 2: fit the models

```
model3 <- gllvm(Y, num.lv = 2, family = "binomial", quadratic = TRUE, sd.errors = FALSE, optim.method = "L-BFG
model4 <- gllvm(Y, num.lv = 2, family = "binomial", quadratic = "LV", sd.errors = FALSE, optim.method = "L-BFG
model5 <- gllvm(Y, num.lv = 2, family = "binomial", row.eff = "random", sd.errors = FALSE, optim.method = "L-B
model6 <- gllvm(Y, num.lv = 2, family = "binomial", sd.errors = FALSE, optim.method = "L-BFGS-B")
```

Example 2: the estimated latent variable

The ordination can differ significantly depending on the specified
model; it can miss the main gradient.

```
##                  model      DDEGO       SLOPE        MIND     SOLRAD         TPI
## quadratic unequal    -0.5648027 -0.18036226  0.49797218 -0.27420274  0.24038836
## quadratic unequal     0.7766327 -0.43089966 -0.77589746  0.23193303 -0.16812365
##  quadratic common    -0.2320379 -0.40938868  0.17643740 -0.16898682  0.09307503
##  quadratic common     0.8360569 -0.37122365 -0.83125090  0.27659502 -0.22457975
##         linear row    -0.7706684  0.06888324  0.67744559 -0.26946282  0.28891390
##         linear row     0.4393991 -0.47633001 -0.49331293  0.11878445 -0.10197456
##             linear    -0.8176695  0.25339964  0.77996367 -0.28037938  0.29646250
##             linear    -0.1495459  0.06442198  0.02183552  0.05711332  0.03319599
##         ELEV
##   0.55098202
##  -0.81011442
##   0.22031704
##  -0.87290047
##   0.74846137
##  -0.49446580
##   0.83835053
##   0.06243901
```

Example 2: the elevation gradient
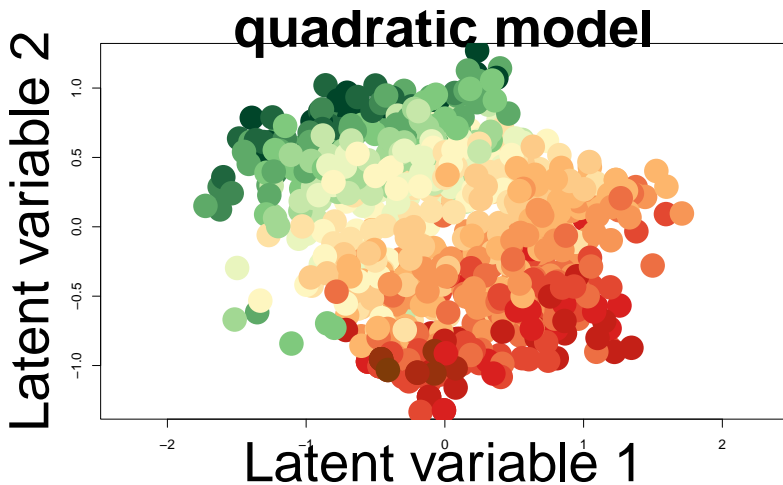
Example 2: curve plotting from common tolerances
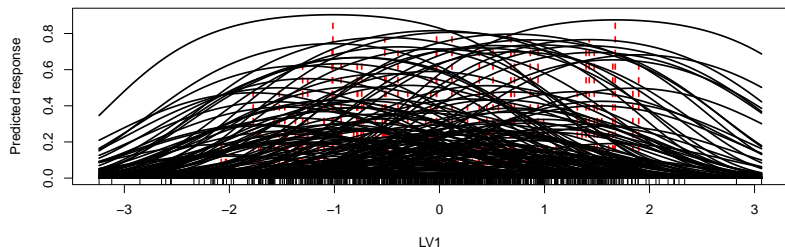
- Each species get a curve
- Red lines are the optima
- Some optima are unobserved
- Rate of turnover is: 7.4659171

Example 2: common tolerances (LV2)

- Each species get a curve
- Red lines are the optima
- Some optima are unobserved
- Rate of turnover is: 3.5387736

Comparison

There has been a lot of talk about the ideal ordination model.
Supposedly not skewed and such, and should avoid artifacts

▶ Podani and Miklos (2002)
▶ Minchin (1987)

Minchin we already know: simulated from skewed responses and
NMDS did best.
Podani and Miklos simulated data that creates artifacts, despite
following a linear model.
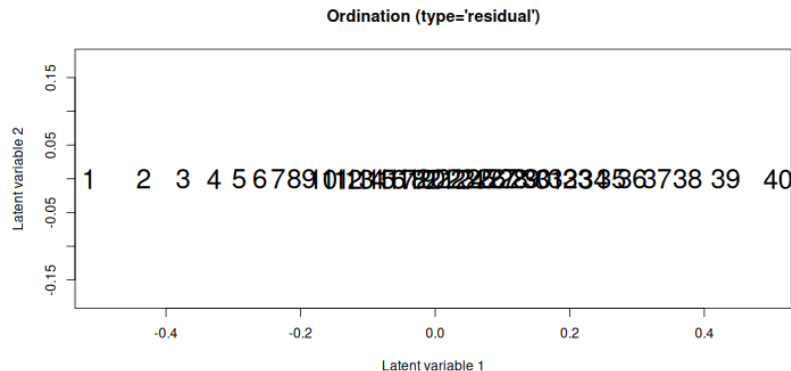
PM1: one-dimensional gradient
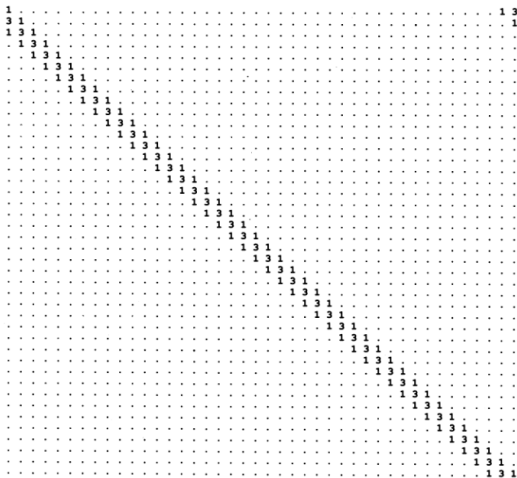
Figure 3: Podani and Miklos (2002) Fig. 1)
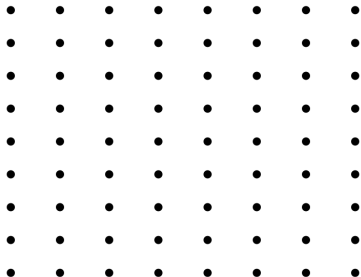
PM1: traditional methods
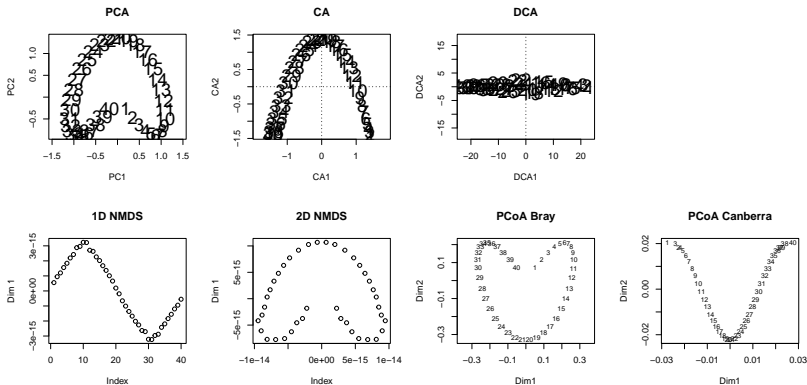
Safe to say, they all do not do great.

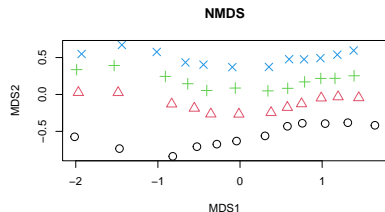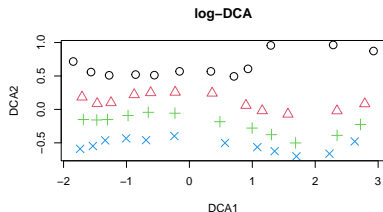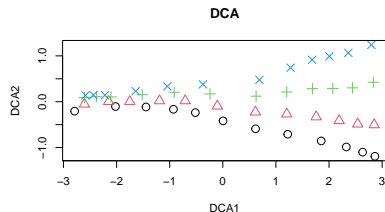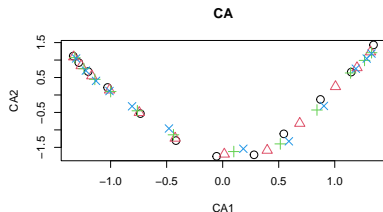# PM1: ordination plot from quadratic model

Note that successfully fitting a quadratic model usually takes a few runs, it tends to be numerically less stable than an ordinary GLLVM.
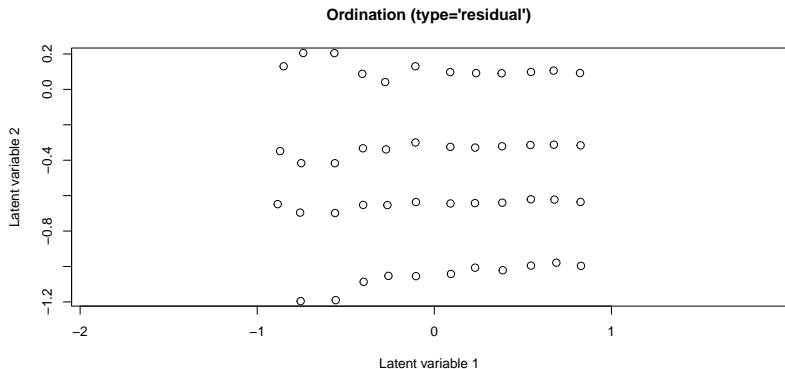
Minchin

Recall: Results Minchin dataset

Minchin

```
model <- gllvm(MC, num.lv = 2, family = "poisson",
          quadratic = TRUE, n.init = 100, n.init.max = 10)
```

Minchin

Ordination (type='residual')

Well, it really does not get better than that.

With covariates

The unimodal response model can be combined with all ordination methods;

▶ Unconstrained ordination
▶ Constrained ordination (random or fixed)
▶ Concurrent ordination (random or fixed canonical coefficients)

But, in the latter two the LVs are not orthogonal and convergence might be more challenging.

Summary

▶ To fit unimodal models we use `quadratic = TRUE` or `quadratic = "LV"`

▶ A simpler, sometimes more realistic, model is with site-specific random effects (equal/common tolerances)

▶ We can calculate gradient length, and rates of turnover, similar to DCA

▶ It is a complex model, and finding a good fit can be challenging
  ▶ But worth it!

▶ Nicely demonstrates the flexibility of GLLVMs for ecological purposes

▶ Handles almost any artifact splendidly