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

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

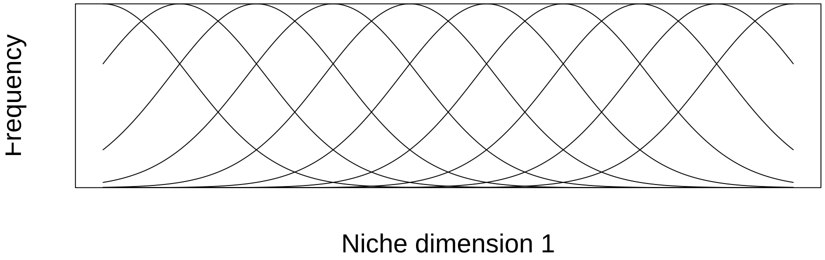
The consensus

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

Terminology

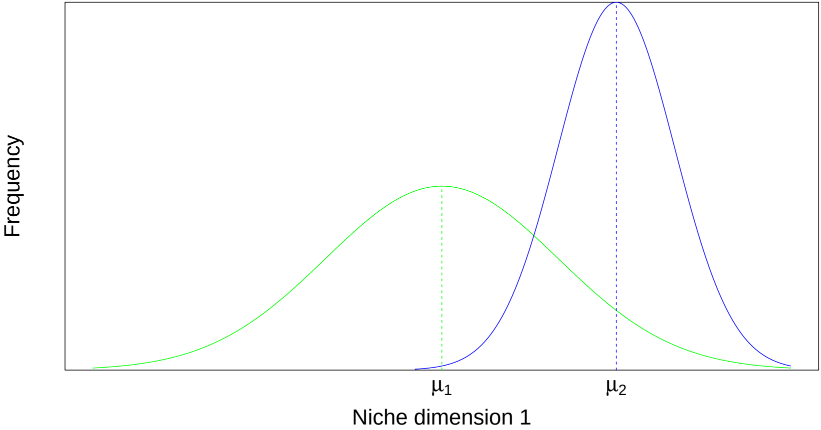
"Gaussian" quadratic unimodal

Species packing

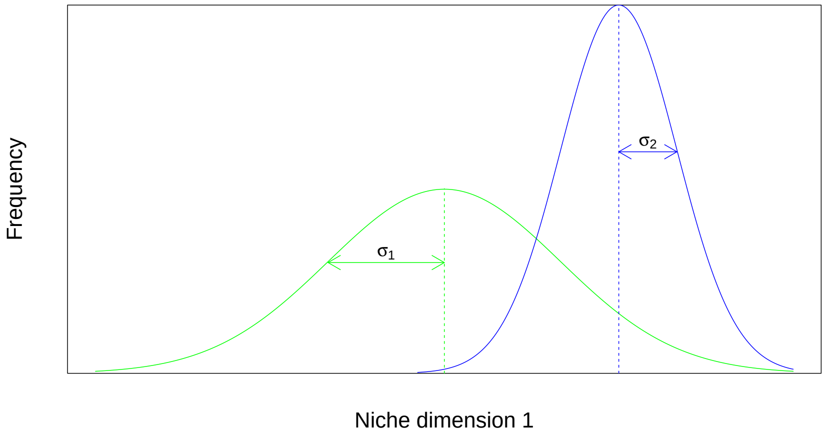


- "Species packing" MacArthur (1969)
  - Competitive exclusion
  - Limiting similarity
  - Leads to CA assumptions (uniform optima, equal maxima, equal tolerances)
- Quadratic in the environment/ niche

- 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?



GLLVM with unimodal response models

Received: 2 October 2020 | Accepted: 1 March 2021
DOI: 10.1111/2041-210X.13595

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Model-based ordination for species with unequal niche widths

Bert van der Veen[1,2,3] | Francis K. C. Hui[4] | Knut A. Hovstad[3,5] | Erik B. Solbu[1] | Robert B. O'Hara[2,3]

van der Veen et al. (2021) developed the first truly unimodal unconstrained ordination method

GLLVM with unimodal response models

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

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

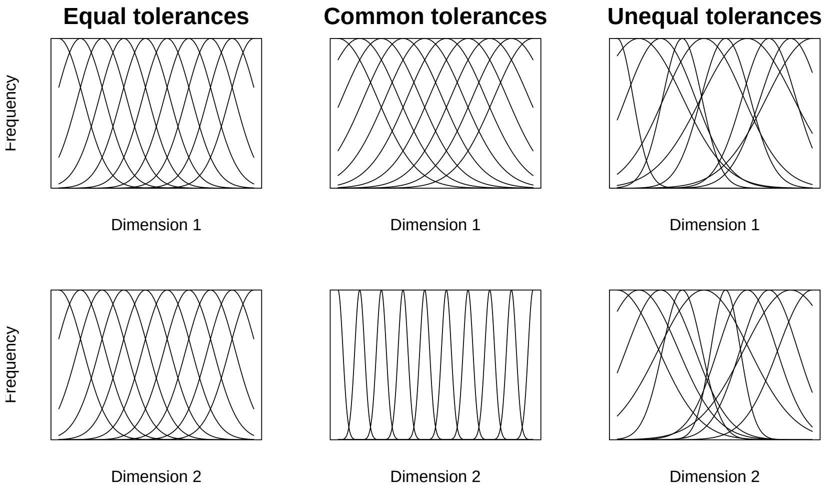
Distributions unimodal response model

Type of data	Distribution	Method	Link
Normal	Gaussian	VA/~~LA~~	identity
Counts	Poisson	VA/~~LA~~	log
	NB	VA/~~LA~~	log
	ZIP	VA/~~LA~~	log
	ZINB	VA/~~LA~~	log
	binomial	VA/~~LA~~	probit
	~~binomial~~	~~LA~~	~~logit~~
Binary	Bernoulli	~~EVA~~ VA/~~LA~~	probit ~~logit~~
Ordinal	Multinomial	VA	cumulative probit
~~Biomass~~	~~Tweedie~~	~~EVA/LA~~	~~log~~
Positive continuous	Gamma	VA/LA	log
	Exponential	VA/LA	log
~~Percent cover~~	~~beta~~	~~LA/EVA~~	~~probit/logit~~
~~with zeros or ones~~	~~ordered beta~~	~~EVA~~	~~probit~~

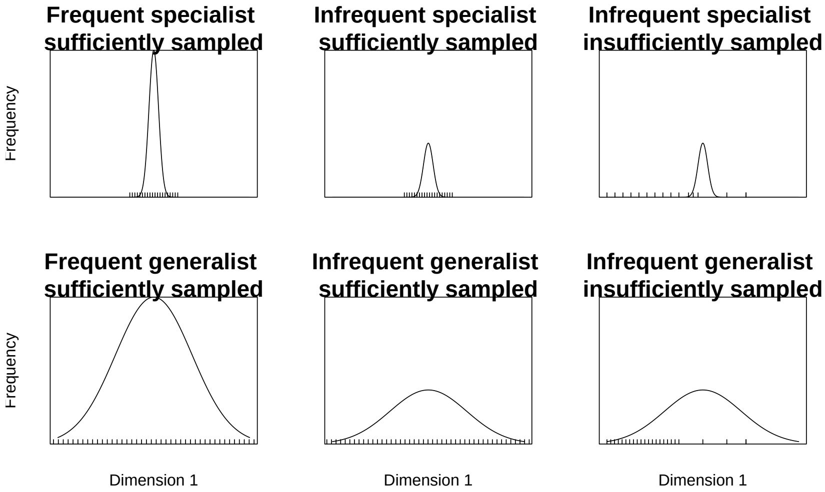
Niche width

- “Tolerance” is the same as niche width
- This is (obviously) 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?



Rare species



Consequences

- Unconstrained Quadratic Ordination (UQO) is a very complex method
- 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 uncorrelated and independent.

On non-linear distortions in GLLVMs

$$\text{cov}(\mathbf{z}_i^\top \boldsymbol{\gamma}_j, \mathbf{z}_i^\top \mathbf{D} \mathbf{z}_i) = 0 \tag{2}$$

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 plot the model as an ordinary ordination.

On plotting a quadratic response mode

- 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
- $\mathbb{E}(\mathbf{z}_i|\mathbf{y}_i)$ 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

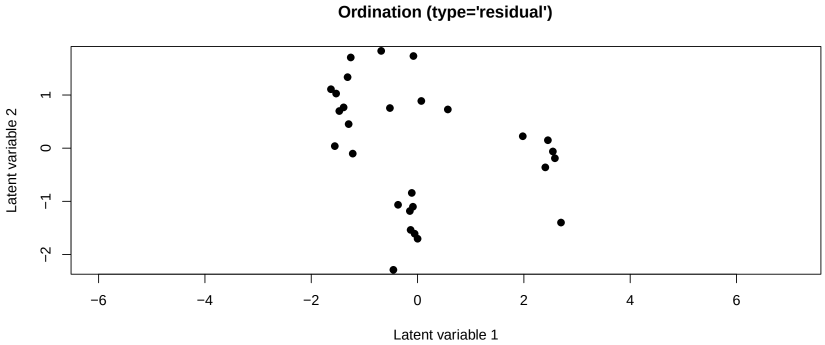
On non-linear distortions in GLLVMs

- 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

$$\Sigma_{j,l} = \gamma_j\gamma_l + \sum_{q=1}^{d} 2D_{jqq}D_{lqq} \tag{3}$$

Example non-linear distortion

```
library(gllvm)
data("spider", package = "mvabund");Y <- spider$abund;
model <- gllvm(Y, num.lv = 2, family = "poisson")
ordiplot(model, symbols = TRUE, pch = 16)
```



Equal tolerances

The quadratic response model with equal tolerances model is:

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

which is the same as:

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

Under the equal tolerances assumption, the quadratic term only affects the total abundance at sites.

What does that mean?

- On an ordination plot we can usually interpret species' locations as optima
- Instead of as the main direction of increase
- You need a random site effect in the model

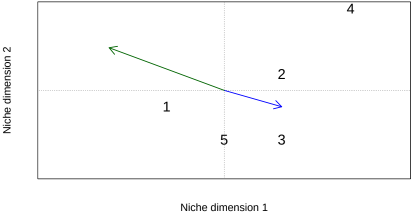


Figure 1: Species effects as increase direction

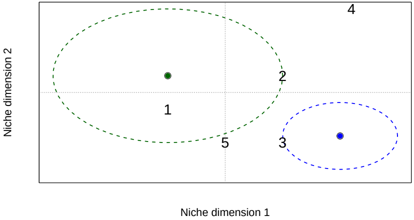
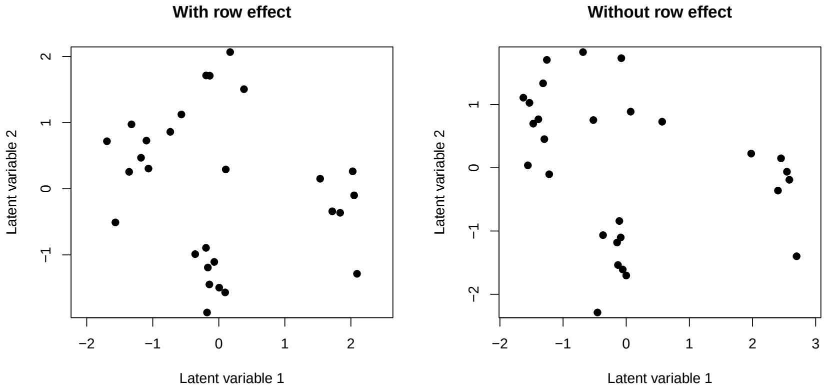


Figure 2: Species effects as centroids

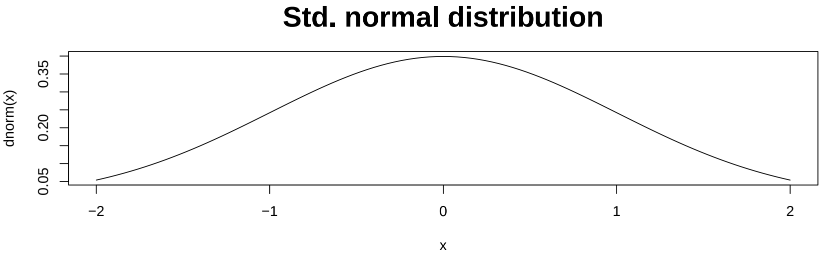
Example non-linear distortion (2)

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



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 our way (van der Veen et al. 2021)



Gradient length

Gradient length 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?

Turnover

We can calculate the rate of turnover as:

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

- $Q(\cdot)$ 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 .999

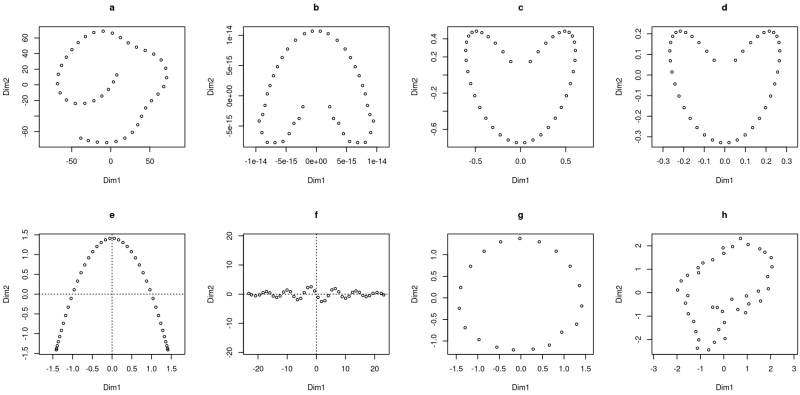
When to go unimodal

- 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

Testing

- Podani and Miklis (2002)
- Minchin (1987)

PM1: overview

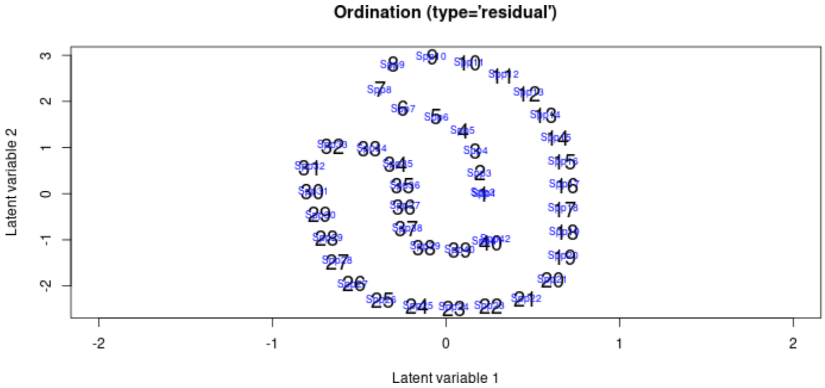


PM1: common tolerances

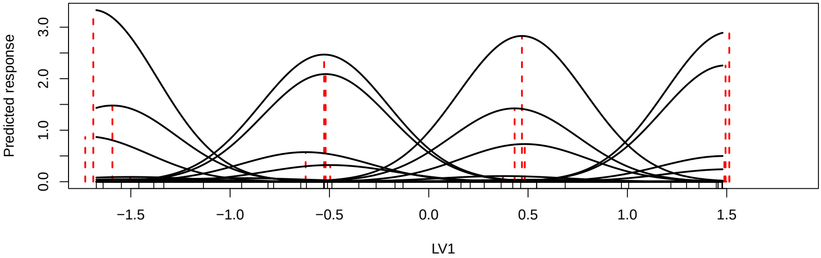
```
model <- gllvm(PM1, num.lv = 2, quadratic = "LV",
               family = "poisson", seed = 74)
ordiplot(model, biplot=TRUE, rotate = FALSE,
         spp.arrows = FALSE, xlim = c(-2,2))
```

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

PM1: common tolerances

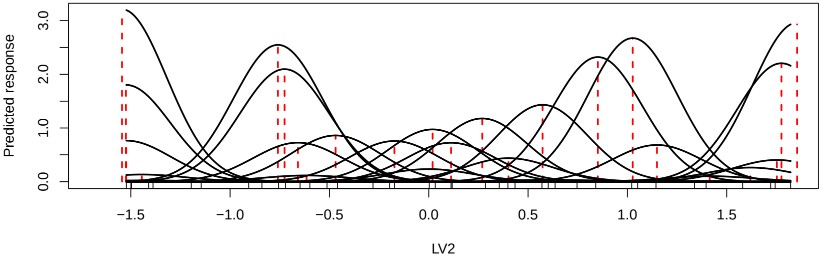


PM1: common tolerances (LV1)



- Each species get a curve
- Red lines are the optima
- Some optima are unobserved
- Rate of turnover is: 1.981247

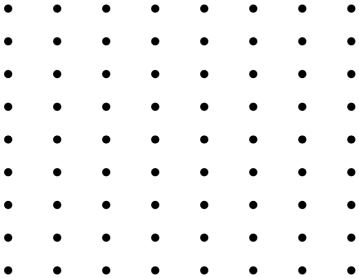
PM1: common tolerances (LV2)



- Each species get a curve
- Red lines are the optima
- Some optima are unobserved
- Rate of turnover is: 1.3878305

Minchin

```r
MC <- read.csv("https://raw.githubusercontent.com/BertvanderV
MC[is.na(MC)] <- 0
```

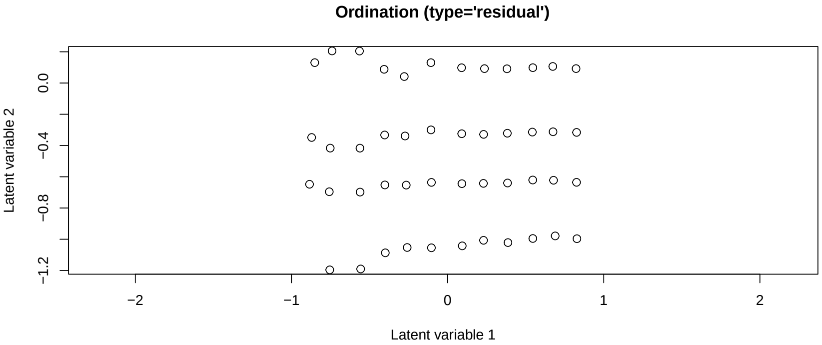


Minchin

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

Minchin

```
ordiplot(model, rotate = FALSE, symbols = TRUE)
```



Well, that looks like a lattice to me!

Summary

- For unimodal models we use `quadratic = TRUE` or `quadratic = "LV"`
- Alternatively, a simpler model is with site-specific random effects (equal 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