



NTNU | Norwegian University of  
Science and Technology

# FROM MODEL-BASED TO HIERARCHICAL ORDINATION

perspectives from community ecology

Bert van der Veen



# Outline

**Part I:** Background

**Part II:** “Model-based” ordination

**Part III:** Getting the methods to the people

**Part IV:** Big picture stuff

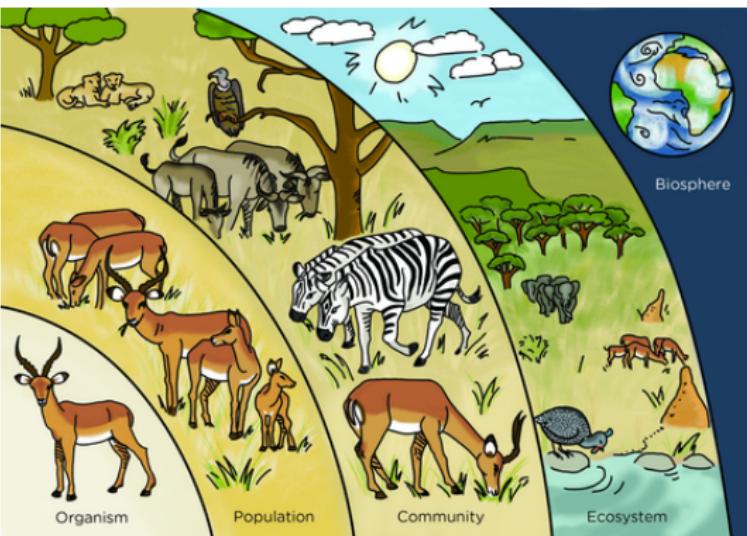
I will try to convince you that community ecology needs these methods.

Slides: <https://github.com/BertvanderVeen/VOC-talk>

Part I

# Background

# Community ecology



# Gathering data

Go out, register species at multiple sites



**Figure 1:** Geir-Harald Strand / NIBIO

# Gathering data

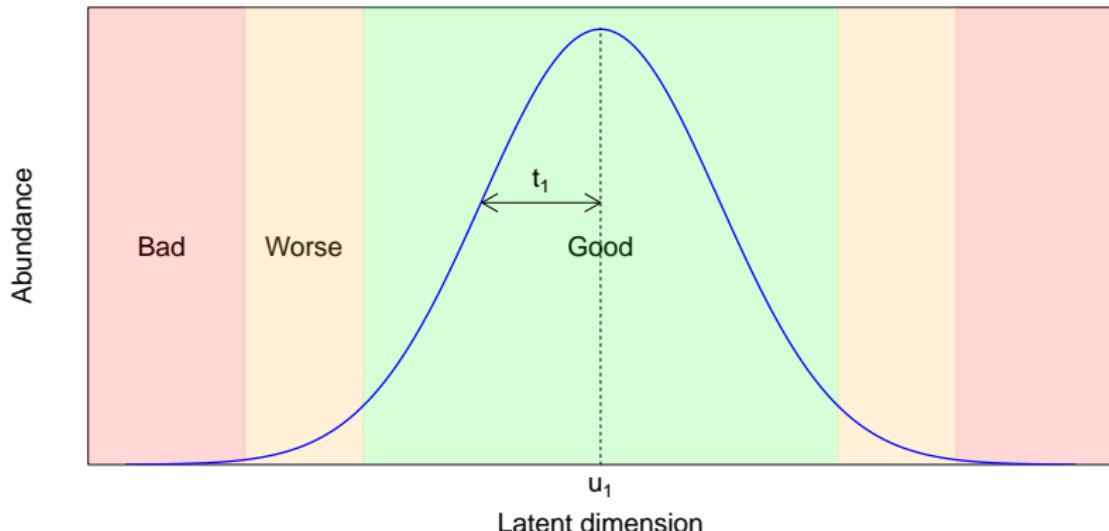
Go out, register species at multiple sites



**Figure 1:** Geir-Harald Strand / NIBIO

## Goal: analysis of sparse data

- ▶ Species occur at few places
- ▶ Shelford's law of tolerance (1931)
- ▶ Specialist and generalist species



# A brief history of ordination

## Milestones in Ordination: a timeline

- 1901 - Pearson develops PCA as a regression technique
- 1927 - Spearman applies factor analysis to psychology
- 1930 - Ramensky uses an informal ordination technique and the term "Ordnung" in ecology
- 1933 - Hotelling develops PCA for understanding the correlation matrix
- 1950 - Curtis and McIntosh employ the "continuum index" approach
- 1952 - Williams uses Correspondence Analysis
- 1954 - Goodall uses the term "ordination" for PCA
- 1957 - Bray-Curtis (Polar) ordination
- 1964 - Kruskal develops NMDS
- 1970's - Whittaker develops theoretical foundations of gradient analysis
- 1973 - Hill revises Correspondence Analysis
- 1976 - Canonical Correlation introduced in ecology
- 1977 - Fasham, Prentice use NMDS
- 1979 - DCA introduced by Hill and Gauch
- 1982 - Gauch's "Multivariate Analysis in Community Ecology"
- 1986 - CCA introduced by ter Braak
- 1986 - Fuzzy set ordination introduced by Roberts
- 1988 - ter Braak and Prentice's "Theory of Gradient Analysis"

This page was created and is maintained by [Michael Palmer](#).



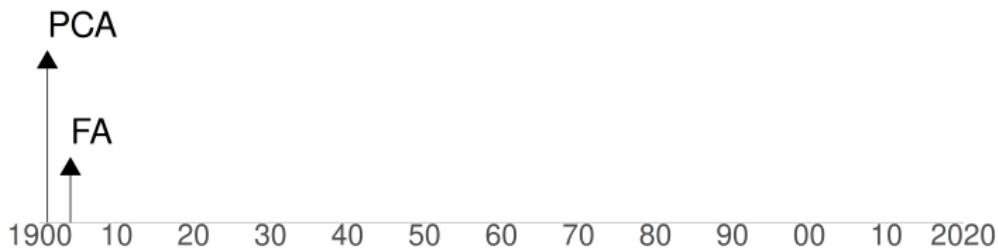
[In the ordination web page](#)

## Historical Perspective

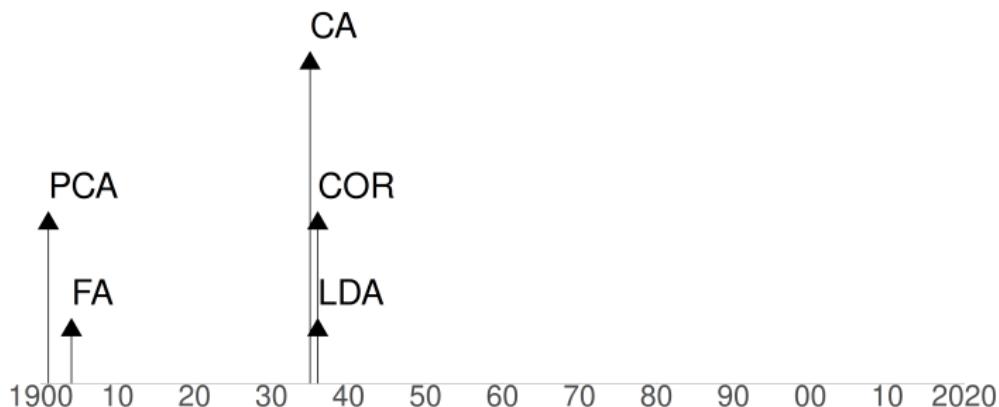
- 1901 - Pearson develops PCA as a regression technique.
- 1927 - Spearman applies factor analysis to psychology.
- 1930 - Ramensky uses an informal ordination technique and introduces the term 'ordnung' into ecology.
- 1954 - D.W. Goodall introduces PCA into ecology and proposes the term 'ordination'.
- 1970 - R.H. Whittaker develops theoretical foundations of gradient analysis, especially unimodal species responses and turnover along environmental gradients.
- 1971 - K.R. Gabriel develops biplot graphical display.
- 1973 - M.O. Hill re-invents correspondence analysis and introduces CA (as 'reciprocal averaging') into ecology.
- 1986 - Cajo ter Braak invents canonical correspondence analysis (CCA) and released CANOCO software.
- 1988 - Cajo ter Braak and Colin Prentice's "A theory of gradient analysis" (Advances in Ecological Research 18; 271-317) that unifies indirect and direct gradient analysis and highlights the importance of underlying species response models.
- 1998, 2002 - Cajo ter Braak and Petr Šmilauer CANOCO 4 & 4.5 software and manual.

- ▶ inspired by Michael Palmer's and John Birk's
- ▶ But those are outdated (like many resources)

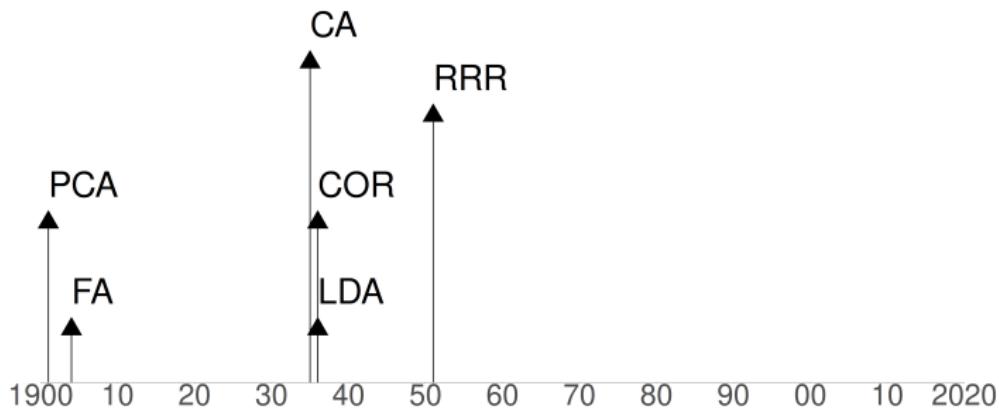
# In the beginning, there were PCA and FA



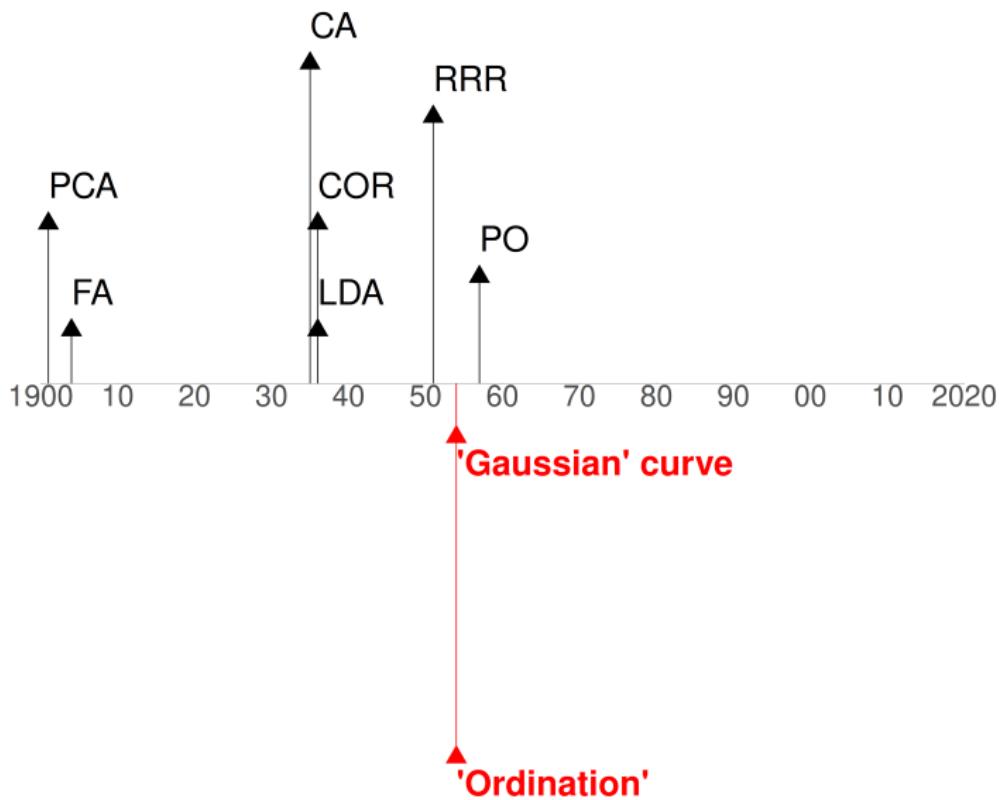
## Then, CA was developed



## Next, reduced rank regression



## Then, there was "Ordination"



# Whittaker 1956: Unimodal responses

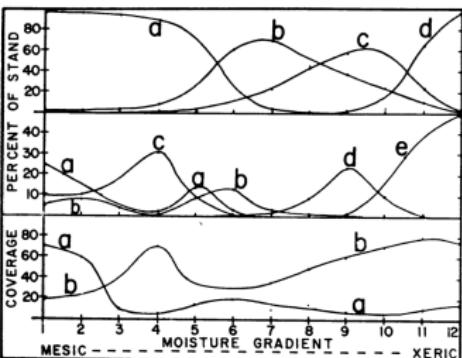
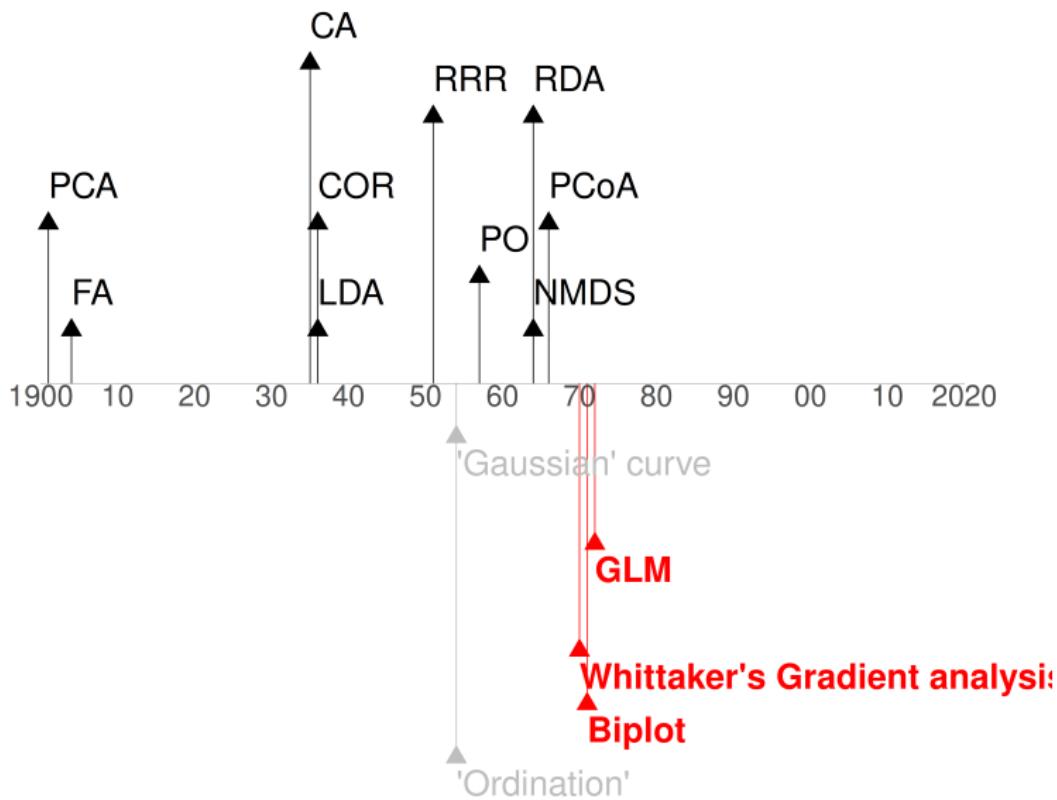
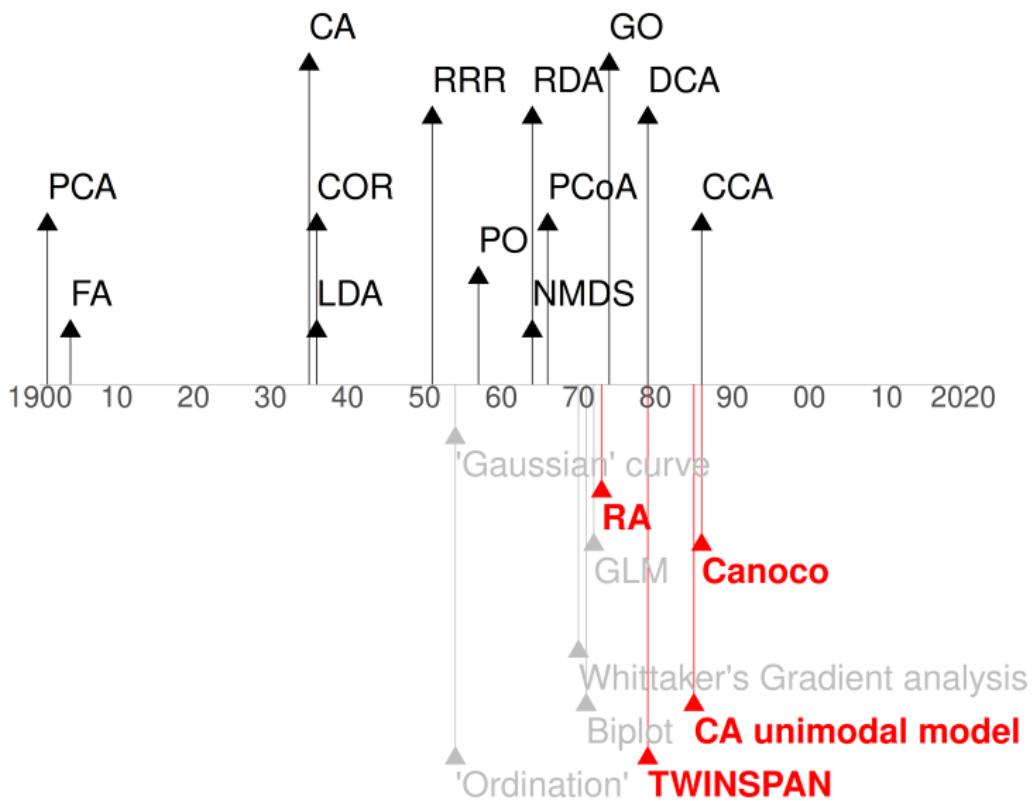


FIG. 4. Transect of the moisture gradient, 3500-4500 ft. Top—curves for tree classes; a, mesic; b, submesic; c, subxeric; d, xeric. Note expansion of mesic stands, compared with Figs. 2 and 3. Middle—curves for tree species: a, *Tilia heterophylla*; b, *Halesia monticola* (both the preceding are bimodal, with populations on each side of the mode of *Tsuga*); c, *Tsuga canadensis*; d, *Quercus alba*; e, *Pinus pungens*. Bottom—curves for undergrowth coverages: a, herbs; b, shrubs.

## Ordination became popular



## Rise of the unimodal model



# Unimodal model (1)

*Ecology* (1974) 55: pp. 1382–1390

## ORDINATION OF VEGETATION SAMPLES BY GAUSSIAN SPECIES DISTRIBUTIONS<sup>1</sup>

HUGH G. GAUCH, JR.

*Ecology and Systematics, Cornell University, Ithaca, New York 14850*

GENE B. CHASE<sup>2</sup>

*Education, Cornell University, Ithaca, New York 14850*

AND

ROBERT H. WHITTAKER

*Ecology and Systematics, Cornell University, Ithaca, New York 14850*

- ▶ Only for a single axis
- ▶ Not much use in practice
- ▶ Was “computationally intensive”

## Unimodal model (2)

BIOMETRICS 41, 859–873  
December 1985

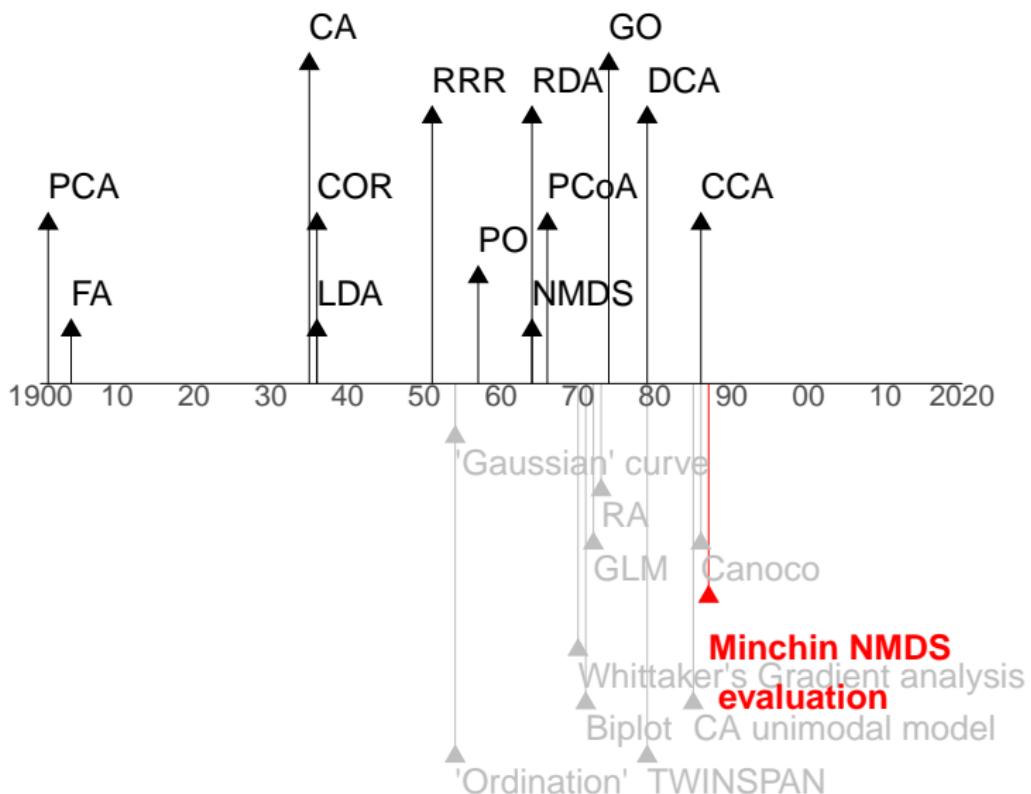
### Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model

Cajo J. F. ter Braak

TNO Institute of Mathematics, Information Processing and Statistics,  
P. O. Box 100, 6700 AC Wageningen, The Netherlands

- ▶ Part of the school of latent variable models
- ▶ Pitched CA approximately implementing a quadratic model

## NMDS as alternative to the unimodal model



# Are species response curves really symmetric?

Vegetatio 69: 89–107, 1987  
© Dr W. Junk Publishers, Dordrecht – Printed in the Netherlands

89

## An evaluation of the relative robustness of techniques for ecological ordination

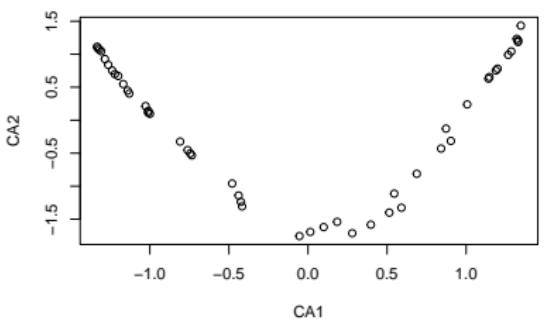
Peter R. Minchin\*

*CSIRO Division of Water and Land Resources, G.P.O. Box 1666, Canberra, 2601, Australia*

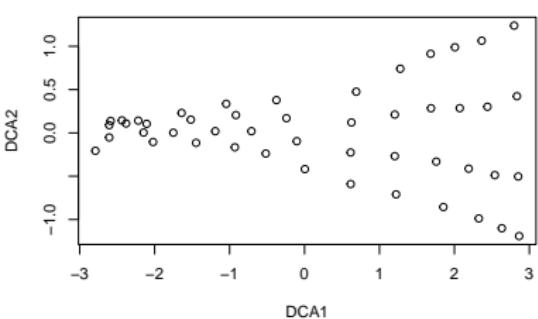
- ▶ A skewed response model might be more realistic
- ▶ NMDS is “robust”
- ▶ Moving away from process-based thinking

# Results Minchin dataset

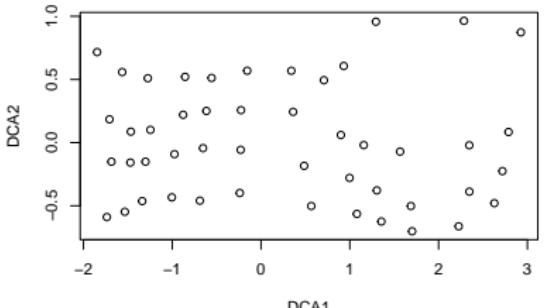
CA



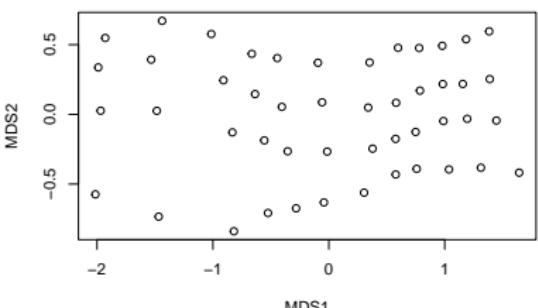
DCA



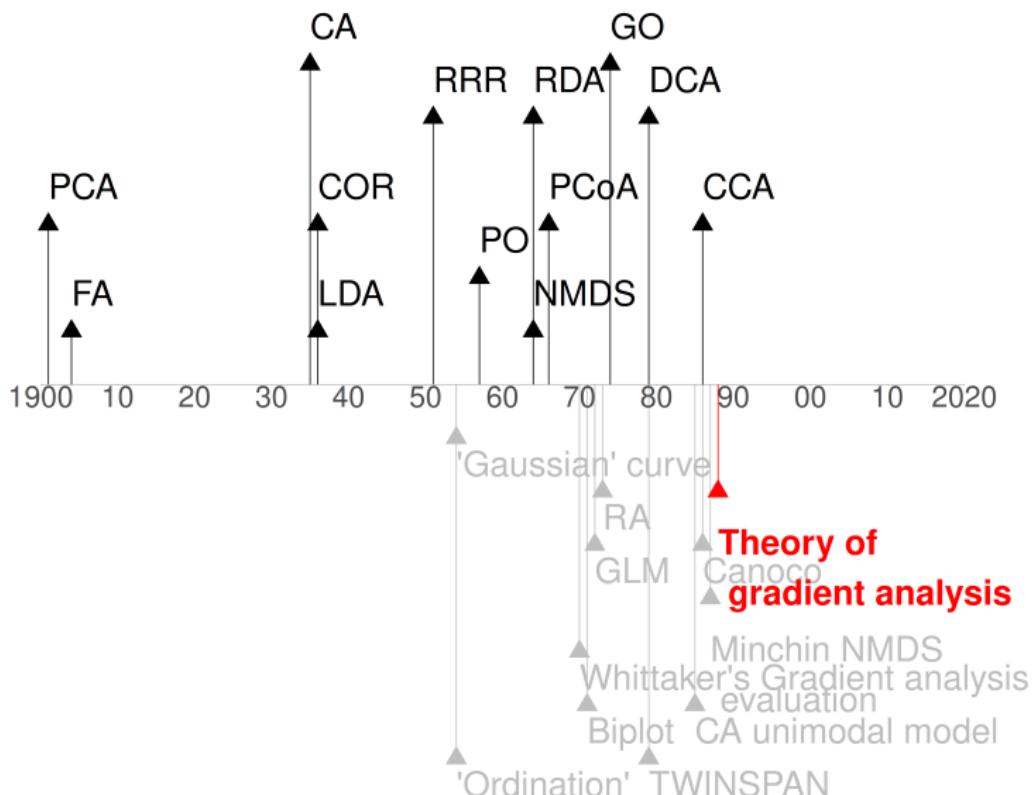
log-DCA c.q. ter Braak and Smilauer (2015)

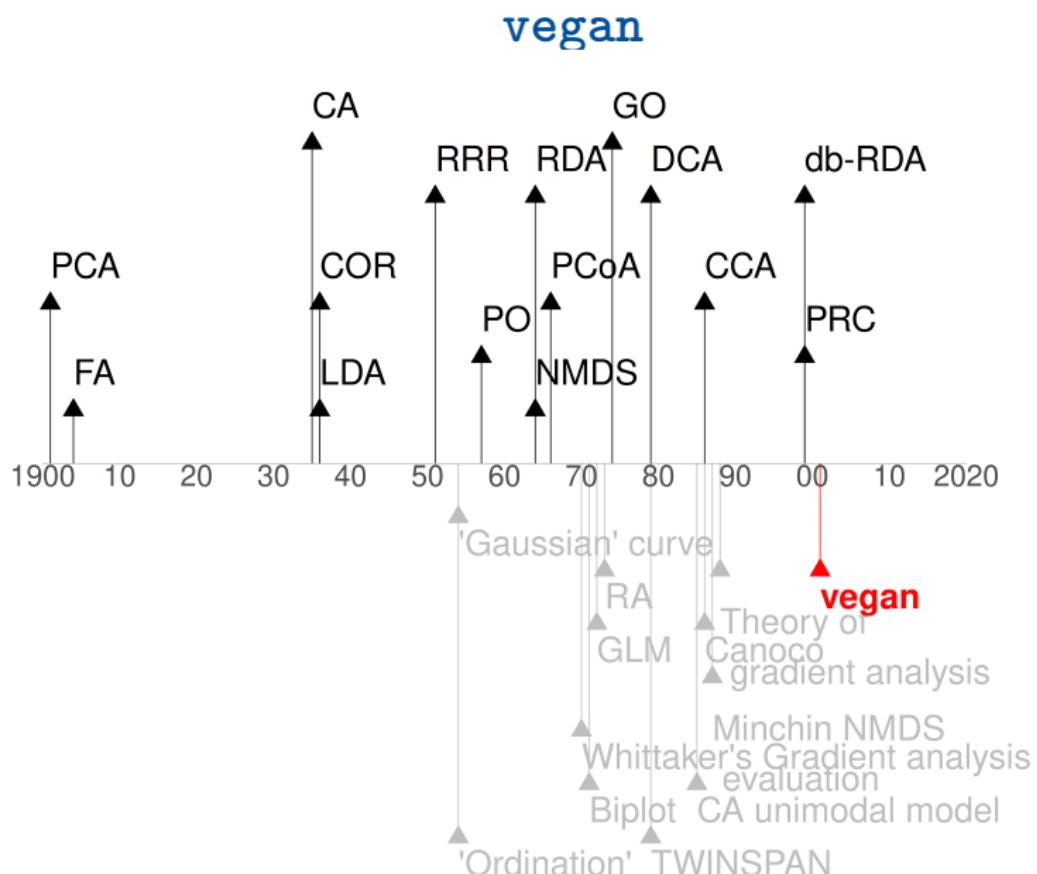


NMDS



## A theory for gradient analysis





## Ordination methods are well-established

- ▶ One of the successes currently: so many resources!
  - ▶ vegan, anadat, labdsv
  - ▶ many books, websites (Michael Palmer, David Zeleney)
  - ▶ run a search and you will find the answer
- ▶ This is where most courses stop
  - ▶ Contemporary ordination methods are not taught
  - ▶ E.g., in machine learning (t-SNE, UMAP and such)
  - ▶ But also model-based ordination methods are not taught

## Ordinations methods in status quo

There is now little discussion in ecology about ordination

## Ordinations methods in status quo

**There is now little discussion in ecology about ordination**

1. Application follows a standard narrative ter Braak and Prentice (1988)

## Ordinations methods in status quo

**There is now little discussion in ecology about ordination**

1. Application follows a standard narrative ter Braak and Prentice (1988)
2. NMDS when that does not work

## Ordinations methods in status quo

**There is now little discussion in ecology about ordination**

1. Application follows a standard narrative ter Braak and Prentice (1988)
2. NMDS when that does not work
3. Contemporary developments are limitedly adopted

## Ordinations methods in status quo

**There is now little discussion in ecology about ordination**

1. Application follows a standard narrative ter Braak and Prentice (1988)
2. NMDS when that does not work
3. Contemporary developments are limitedly adopted
4. Little discussion about performance or preference of methods

## Ordinations methods in status quo

There is now little discussion in ecology about ordination

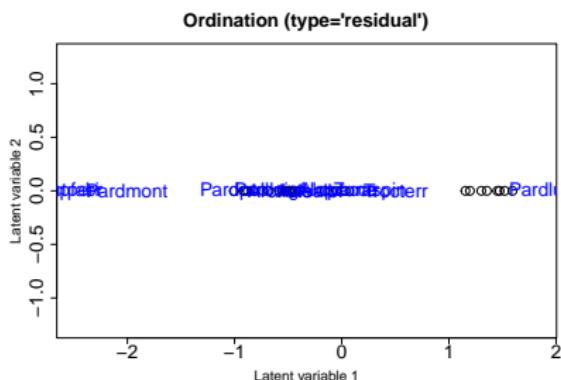
1. Application follows a standard narrative ter Braak and Prentice (1988)
2. NMDS when that does not work
3. Contemporary developments are limitedly adopted
4. Little discussion about performance or preference of methods
5. Ordination methods are now *descriptive*
  - ▶ One upon a time, ordination **was** the push for more quantitative methods

## The issue? Attitude.

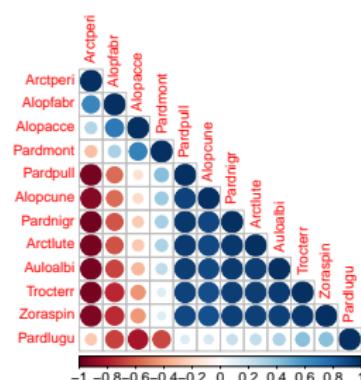
1. Ordination methods are descriptive and uncool
2. Complex methods are thought not *robust*
3. The methods are not understandable  
(because *linear regression* is hard to understand)

## Low-dimensional analysis

The narrative has changed: the right visualization is “cooler” than the left (despite the same underlying model)



**Figure 2:** ordination of spider data



**Figure 3:** correlations of spider data

We are working on telling people these are the same..

**Use Generalised Linear Latent Variable Models  
to get the best of both community ecology  
and macroecology!**

R.B. O'Hara

A. Rugstad

B. van der Veen

## Correlation and interaction

# ECOLOGY LETTERS

Ideas and Perspectives

### Co-occurrence is not evidence of ecological interactions

F. Guillaume Blanchet ✉ Kevin Cazelles, Dominique Gravel

First published: 19 May 2020 | <https://doi.org/10.1111/ele.13525> | Citations: 429

## Part II

"Model-based" ordination

## Where it started

### Methods in Ecology and Evolution



| Free Access

Distance-based multivariate analyses confound location and dispersion effects

David I. Warton Stephen T. Wright, Yi Wang

First published: 06 June 2011 | <https://doi.org/10.1111/j.2041-210X.2011.00127.x> | Citations: 627

Correspondence site: <http://www.respond2articles.com/MEE/>

### Methods in Ecology and Evolution



Forum | Open Access |

The central role of mean-variance relationships in the analysis of multivariate abundance data: a response to Roberts (2017)

David I. Warton Francis K. C. Hui

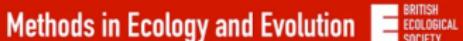
# Standing on the shoulder of giants



Article

Random-effects ordination: describing and predicting multivariate correlations and co-occurrences

Steven C. Walker Donald A. Jackson



Special Feature: New Opportunities at the Interface Between Ecology and Statistics | Free Access |

Model-based approaches to unconstrained ordination

Francis K.C. Hui Sara Taskinen, Shirley Pledger, Scott D. Foster, David I. Warton

First published: 23 July 2014 | <https://doi.org/10.1111/2041-210X.12236> | Citations: 57

## So Many Variables: Joint Modeling in Community Ecology

David I. Warton,<sup>1,\*</sup> F. Guillaume Blanchet,<sup>2</sup> Robert B. O'Hara,<sup>3</sup> Otso Ovaskainen,<sup>4,5</sup> Sara Taskinen,<sup>6</sup> Steven C. Walker,<sup>2</sup> and Francis K.C. Hui<sup>7</sup>

Ecological Monographs, 74(4), 2004, pp. 685–701  
© 2004 by the Ecological Society of America

A NEW TECHNIQUE FOR MAXIMUM-LIKELIHOOD CANONICAL GAUSSIAN ORDINATION

THOMAS W. YEE<sup>1</sup>

# What functionality was still needed to replace ordination

1. Unimodal responses
2. Constrained ordination (with random effects)
3. Additional random effects for nested designs
4. User friendliness (as much as possible)
5. Robustness c.q. Minchin (1980)

# Generalized Linear Latent Variable Model (GLLVM)

- ▶ A framework for model-based multivariate analysis
- ▶ That does dimension reduction (i.e., ordination)
- ▶ There is no distance measure to specify
- ▶ You do need to specify:
  1. A distribution
  2. A link function
  3. The model structure
- ▶ We (can) treat terms as random or fixed when appropriate

## Response distribution

$$y_{ij} \sim f \left\{ g^{-1} \left( \eta_{ij} \right), \phi_j \right\} \quad (1)$$

## Response distribution

$$y_{ij} \sim f \left\{ g^{-1} \left( \eta_{ij} \right), \phi_j \right\} \quad (1)$$

1. Community data



## Response distribution

$$y_{ij} \sim f \left\{ g^{-1} \left( \eta_{ij} \right), \phi_j \right\} \quad (1)$$

2. Response distribution



## Response distribution

$$y_{ij} \sim f \left\{ g^{-1} \left( \eta_{ij} \right), \phi_j \right\} \quad (1)$$

3. (inverse) Link function



## Response distribution

$$y_{ij} \sim f \left\{ g^{-1} \left( \eta_{ij} \right), \phi_j \right\} \quad (1)$$

4. Linear predictor ("the model") —

## Response distribution

$$y_{ij} \sim f\left\{ g^{-1}\left(\eta_{ij}\right), \phi_j \right\} \quad (1)$$

5. Dispersion parameter \_\_\_\_\_

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \beta_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} \implies \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

1. Linear predictor

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$



2. Species (column) intercept

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \beta_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

3. Sample (row) intercept \_\_\_\_\_

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

4. Fixed effects

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij}$$

(2)

5. Random effects

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

6. Residual error incorporating between-column correlation

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \beta_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

$$\mathcal{L}(\Theta) = \log \left\{ \int f\left(\mathbf{Y} | \epsilon, \mathbf{u}, \mathbf{r}_0\right) h\left(\epsilon, \mathbf{u}, \mathbf{r}_0; \Sigma_\epsilon, \Sigma_u, \Sigma_{r_0}\right) \right\} \partial(\epsilon, \mathbf{u}, \mathbf{r}_0) \quad (3)$$

## The reference model

$$g\{\mathbb{E}(y_{ij}|r_0, \epsilon, \mathbf{u}, \mathbf{r}_0)\} = \beta_0 + \beta_1 \epsilon_{ij} + \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

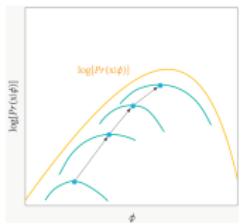
Requires fast methods for integration

$$\mathcal{L}(\Theta) = \log \left\{ \int f\left(\mathbf{Y} | \epsilon, \mathbf{u}, \mathbf{r}_0\right) h\left(\epsilon, \mathbf{u}, \mathbf{r}_0; \Sigma_\epsilon, \Sigma_u, \Sigma_{r_0}\right) \right\} \partial(\epsilon, \mathbf{u}, \mathbf{r}_0) \quad (3)$$

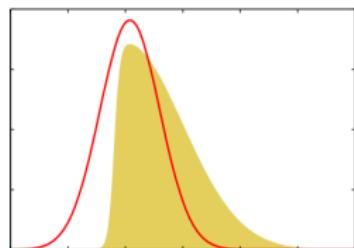
## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \beta_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

$$\mathcal{L}(\Theta) = \log \left\{ \int f\left(\mathbf{Y} | \epsilon, \mathbf{u}, \mathbf{r}_0\right) h\left(\epsilon, \mathbf{u}, \mathbf{r}_0; \Sigma_\epsilon, \Sigma_u, \Sigma_{r_0}\right) \right\} \partial(\epsilon, \mathbf{u}, \mathbf{r}_0) \quad (3)$$



**Figure 4:** Variational Approximations [borealisai.com](http://borealisai.com)

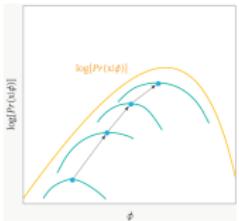


**Figure 5:** Laplace approximation [wiljohn.top](http://wiljohn.top)

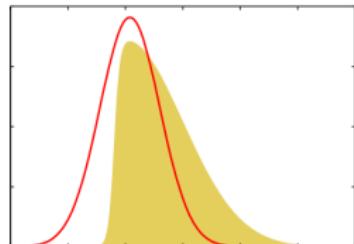
## The reference model

$$g\{\mathbb{E}(y_{ij}|r_{0i}, \mathbf{u}_j, \epsilon_{ij})\} = \eta_{ij} = \beta_{0j} + r_{0i} + \mathbf{x}_i^\top \beta_j + \mathbf{z}_i^\top \mathbf{u}_j + \epsilon_{ij} \quad (2)$$

$$\mathcal{L}(\Theta) = \log \left\{ \int f \left( \begin{array}{c} \text{Usually, we} \\ \text{cannot fit this} \end{array} \right) \epsilon, \Sigma_u, \Sigma_{r_0} \right\} \partial(\epsilon, \mathbf{u}, \mathbf{r}_0) \quad (3)$$



**Figure 4:** Variational Approximations [borealisai.com](http://borealisai.com)



**Figure 5:** Laplace approximation [wiljohn.top](http://wiljohn.top)

## Reduced-rank modeling

Instead of reducing the data,

**we reduce the model to a low-dimensional form**

So that we need to estimate fewer parameters, and can fit faster

# Generalized Linear (Mixed-effects) Models

$$\eta = \mathbf{1}\beta^\top + \dots + \mathbf{E}, \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (4)$$

↑  
This is a GLMM—Induces between-column correlation

# Generalized Linear (Mixed-effects) Models

$$\eta = \mathbf{1}\boldsymbol{\beta}^\top + \dots + \mathbf{E}, \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top) \quad (4)$$

This is an unconstrained ordination

# Generalized Linear (Mixed-effects) Models

$$\eta = \mathbf{1}\beta^\top + \dots + \mathbf{E}, \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \Gamma\Gamma^\top) \quad (4)$$

This is an unconstrained ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \mathbf{XB} \quad (5)$$

This is a GLM

# Generalized Linear (Mixed-effects) Models

$$\eta = \mathbf{1}\beta^\top + \dots + \mathbf{E}, \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \Gamma\Gamma^\top) \quad (4)$$

This is an unconstrained ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \mathbf{X}\mathbf{B}_{lv}\Gamma^\top \quad (5)$$

This is a constrained ordination

# Generalized Linear (Mixed-effects) Models

$$\eta = \mathbf{1}\beta^\top + \dots + \Lambda\Gamma^\top, \quad \lambda_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (4)$$

This is an unconstrained ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \Lambda\Gamma^\top \quad (5)$$

This is a constrained ordination

# Generalized Linear (Mixed-effects) Models

$$\eta = \mathbf{1}\beta^\top + \dots + \mathbf{\Lambda}\boldsymbol{\Gamma}^\top, \quad \lambda_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (4)$$

This is an unconstrained ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \mathbf{\Lambda}\boldsymbol{\Gamma}^\top \quad (5)$$

This is a constrained ordination

We could apply a decomposition *post-hoc*,  
but the idea is to do it *a-priori* and truncate it  
*Uniqueness* needs to be safeguarded with **constraints**

# A unifying framework for ordination

**Concurrent ordination: Simultaneous unconstrained and constrained latent variable modelling**

Bert van der Veen<sup>1,2,3</sup>  | Francis K. C. Hui<sup>4</sup>  | Knut A. Hovstad<sup>3,5</sup>  |  
Robert B. O'Hara<sup>2,3</sup> 

## Concurrent ordination

$$\eta = \underbrace{\mathbf{1}\beta^\top}_{\text{A vector}} + \dots + \underbrace{\Lambda \Gamma^\top}_{\text{A matrix}} \quad (6)$$

The model is flexible,  $\Lambda$  can be all kinds of things.

## Concurrent ordination

$$\eta = \underbrace{\mathbf{1}\beta^\top}_{\text{A vector}} + \dots + \underbrace{\Lambda \Gamma^\top}_{\text{A matrix}} \quad (6)$$

The model is flexible,  $\Lambda$  can be all kinds of things.

## Concurrent ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \Lambda \Gamma^\top \quad (6)$$

The model is flexible,  $\Lambda$  can be all kinds of things.

→ 1.  $\Lambda = E$ , unconstrained

## Concurrent ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \Lambda \Gamma^\top \quad (6)$$

The model is flexible,  $\Lambda$  can be all kinds of things.

→ 1.  $\Lambda = E$ , unconstrained

→ 2.  $\Lambda = XB$ , constrained

## Concurrent ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \Lambda \Gamma^\top \quad (6)$$

The model is flexible,  $\Lambda$  can be all kinds of things.

- 1.  $\Lambda = E$ , **unconstrained**
- 2.  $\Lambda = XB$ , **constrained**
- 3.  $\Lambda = XB + E$ , **concurrent**

## Concurrent ordination

$$\eta = \mathbf{1}\beta^\top + \dots + \Lambda \Gamma^\top \quad (6)$$



The model is flexible.  $\Lambda$  can be all kinds of things.

i.e., we have a (hierarchical) Linear Mixed-effects model for the latent variables

- 1.  $\Lambda$  =  $\mathbf{X}\mathbf{B}$ , constrained
- 2.  $\Lambda$  =  $\mathbf{X}\mathbf{B} + \mathbf{E}$ , concurrent

# Unimodal responses

## Vanilla GLLVM

$$\eta_{ij} = \beta_{0j} + \dots + \lambda_i^\top \gamma_j \quad (7)$$

## Unimodal responses

**Vanilla GLLVM**

**Quadratic extension**

$$\eta_{ij} = \beta_{0j} + \dots + \boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_j \quad (7)$$

$$\eta_{ij} = \beta_{0j} + \dots + \boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_j - \boldsymbol{\lambda}_i^\top \mathbf{D}_j \boldsymbol{\lambda}_i \quad (8)$$

- ▶ For positive diagonal matrix of quadratic coefficients  $\mathbf{D}_j$
- ▶ I.e. non-linear (quadratic) in the LVs (van der Veen et al. 2021)
- ▶ Similar to e.g. Correspondence Analysis (ter Braak 1985) with  $D_{jqq} = d \ \forall q, j$

## Unimodal responses

**Vanilla GLLVM**

**Quadratic extension**

$$\eta_{ij} = \beta_{0j} + \dots + \boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_j \quad (7)$$

$$\eta_{ij} = \beta_{0j} + \dots + \boldsymbol{\lambda}_i^\top \boldsymbol{\gamma}_j - \boldsymbol{\lambda}_i^\top \mathbf{D}_j \boldsymbol{\lambda}_i \quad (8)$$

- ▶ For positive diagonal matrix of quadratic coefficients  $\mathbf{D}_j$
- ▶ I.e. non-linear (quadratic) in the LVs (van der Veen et al. 2021)
- ▶ Similar to e.g. Correspondence Analysis (ter Braak 1985) with  $D_{jqq} = d \forall q, j$

**Residual covariance**

$$\Sigma_{jk} = \sum_{q=1}^d \gamma_{jq} \gamma_{kq} + 2\mathbf{D}_{jqq} \mathbf{D}_{kqq} \quad (9)$$

$$= \sum_{q=1}^d (\mathbf{t}_{jq}^2 \mathbf{t}_{kq}^2)^{-1} + (0.5 + \mathbf{u}_{jq} \mathbf{u}_{kq}) \quad (10)$$

# Robustness

Vegetatio 69: 89–107, 1987

© Dr W. Junk Publishers, Dordrecht – Printed in the Netherlands

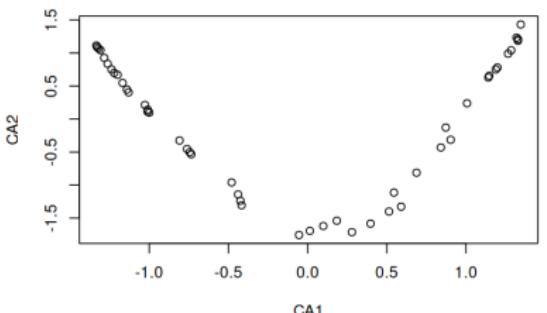
89

## An evaluation of the relative robustness of techniques for ecological ordination

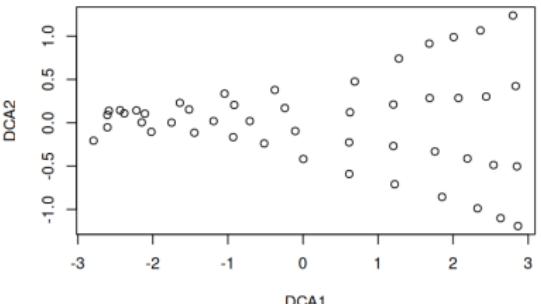
Peter R. Minchin\*

*CSIRO Division of Water and Land Resources, G.P.O. Box 1666, Canberra, 2601, Australia*

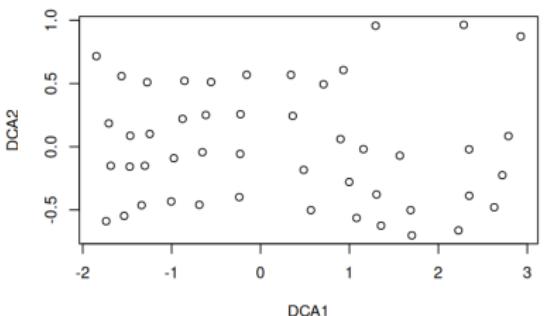
CA



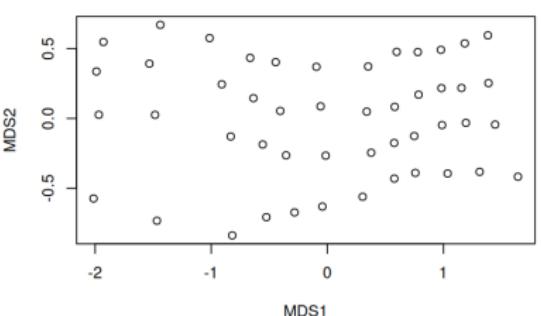
DCA

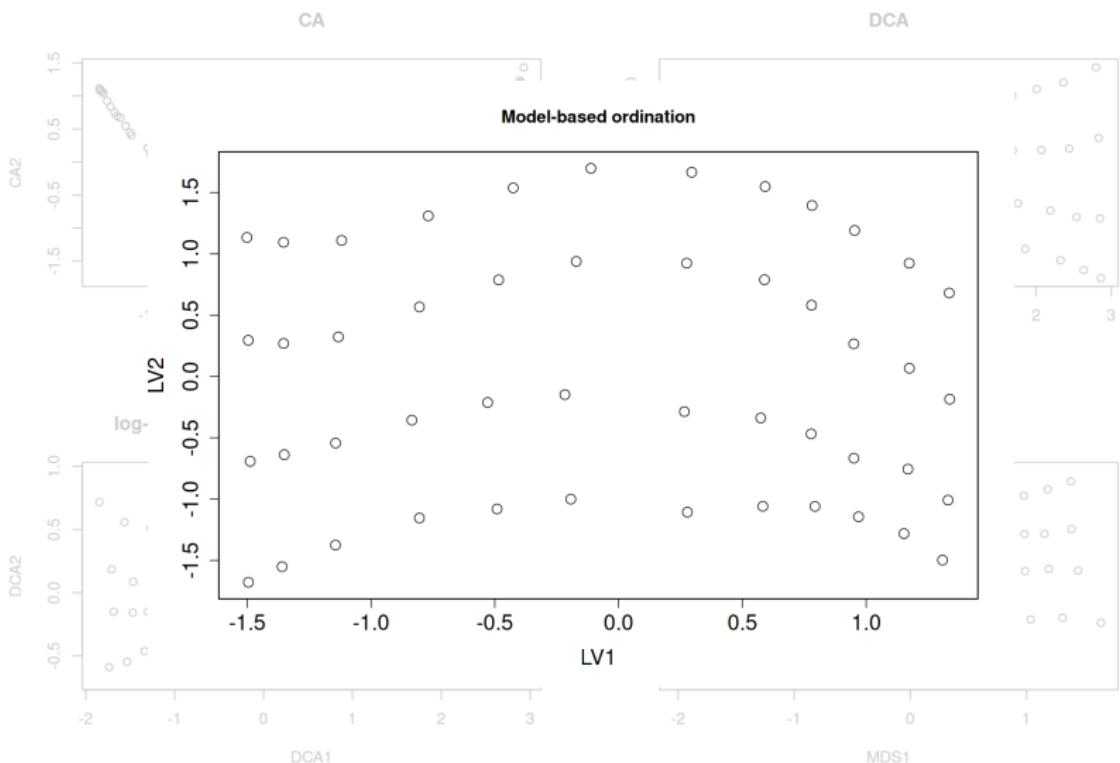


log-DCA c.q. ter Braak and Smilauer (2015)



NMDS

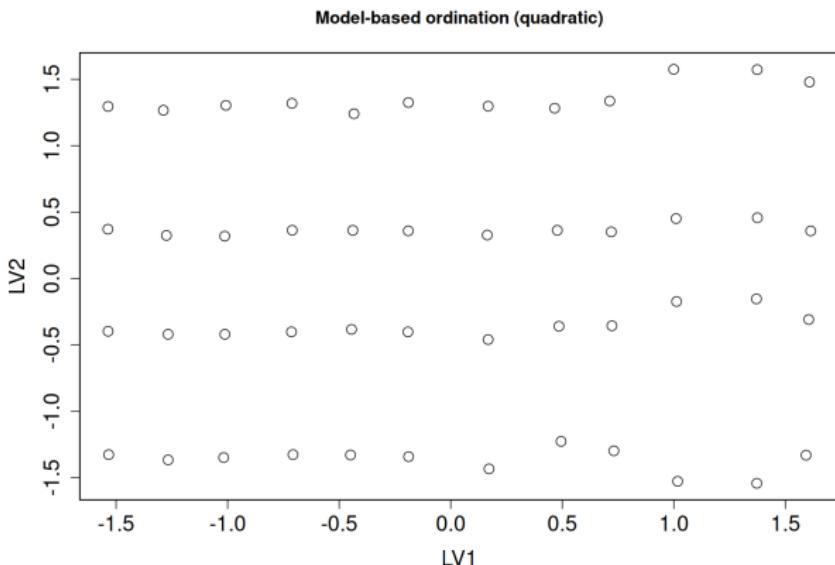




# Robustness

**Now with quadratic model (van der Veen et al. 2021)**

# Robustness



## Part III

# Getting methods to the people

# gllvm

Received: 7 May 2019 | Accepted: 5 September 2019

DOI: 10.1111/2041-210X.13303

## APPLICATION

Methods in Ecology and Evolution

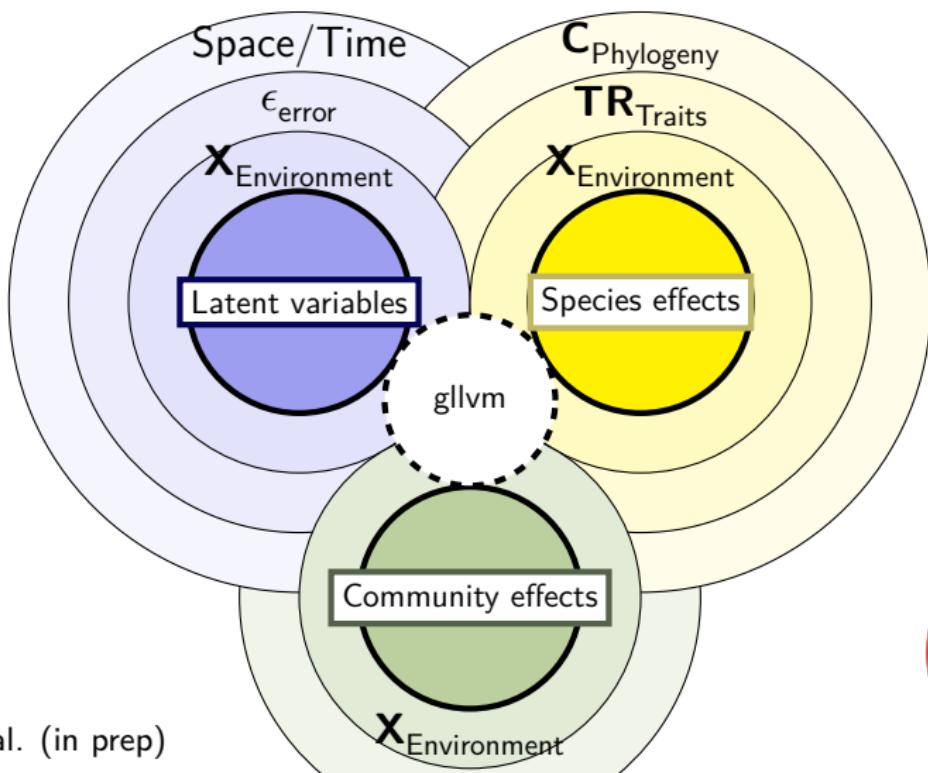


## gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in R

Jenni Niku<sup>1</sup> | Francis K. C. Hui<sup>2</sup> | Sara Taskinen<sup>1</sup> | David I. Warton<sup>3</sup>

- ▶ Originally published in 2019 by Niku et al. I "joined in" shortly after
- ▶ For model-based multivariate analysis of community ecological data
- ▶ Models are fitted in C++ (Kristensen et al. 2015)
- ▶ Estimated with general-purpose numerical optimisation

# gllvm 2.0



# Response distributions

Distributions for all common data types:

Continuous	Bounded	Discrete	Zero-inflated
▶ Gaussian	▶ exponential ▶ gamma ▶ Tweedie ▶ beta	▶ Poisson ▶ NB <sup>1</sup> ▶ binomial ▶ ordinal	▶ Poisson ▶ NB <sup>1</sup> ▶ hurdle beta ▶ ordered beta

---

<sup>1</sup>negative binomial

## Syntax headlines

- ▶ Main function: `gllvm(.)`
- ▶ Define the type of latent variable: `num.lv`, `num.RR`, `num.lv.c`
- ▶ Formula arguments to specify the main components:  
`formula`, `lv.formula`, `row.eff`
- ▶ `family` to specify the distribution

# Demonstration

Wadden sea data [Dewenter et al. \(2023\)](#)

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

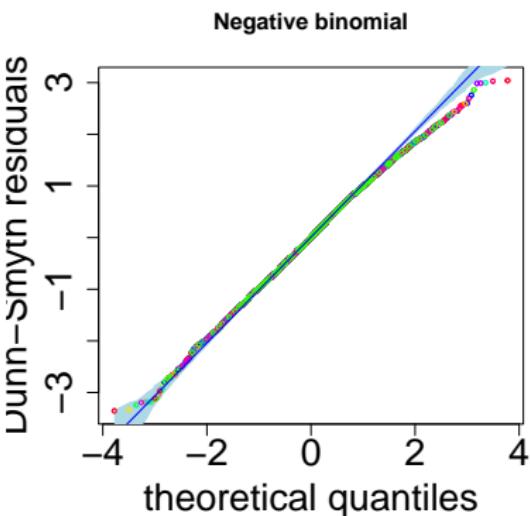
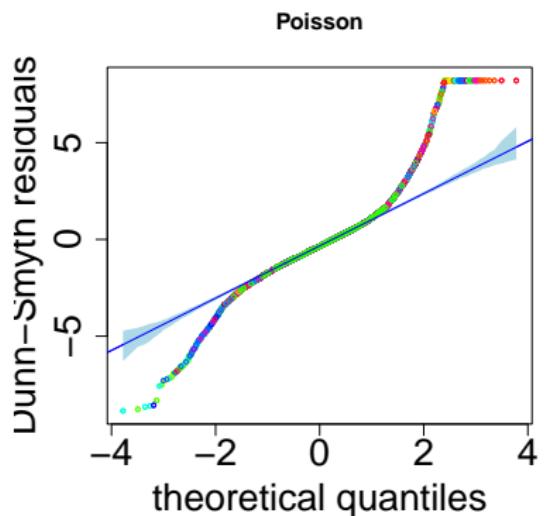


Figure 6: nioz.nl

```
Ya <- read.csv("../data/waddenY.csv") [,-c(1:2)];  
Ya <- Ya[, colSums(ifelse(Ya==0,0,1))>2]  
X <- read.csv("../data/waddenX.csv")  
X[,apply(X,2,is.numeric)] <- scale(X[,apply(X,2,is.numeric)])
```

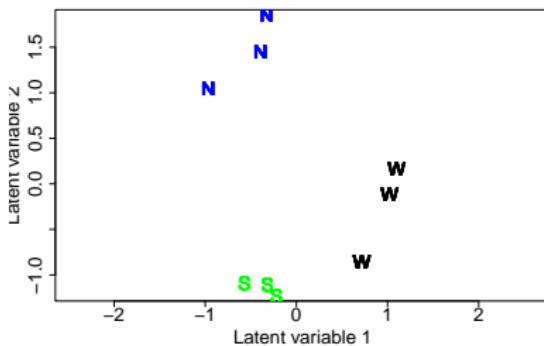
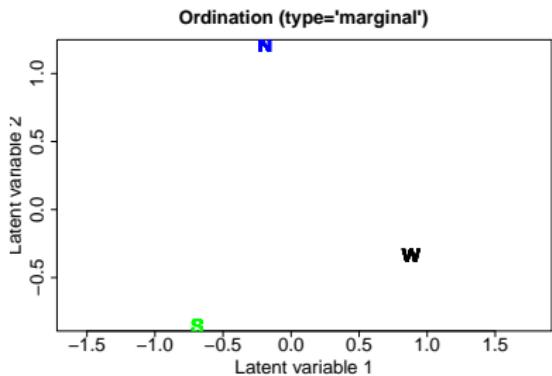
## Unconstrained ordination

```
uord1 <- gllvm(Ya, num.lv = 2, family = "poisson")
uord2 <- gllvm(Ya, num.lv = 2, family = "negative.binomial")
```



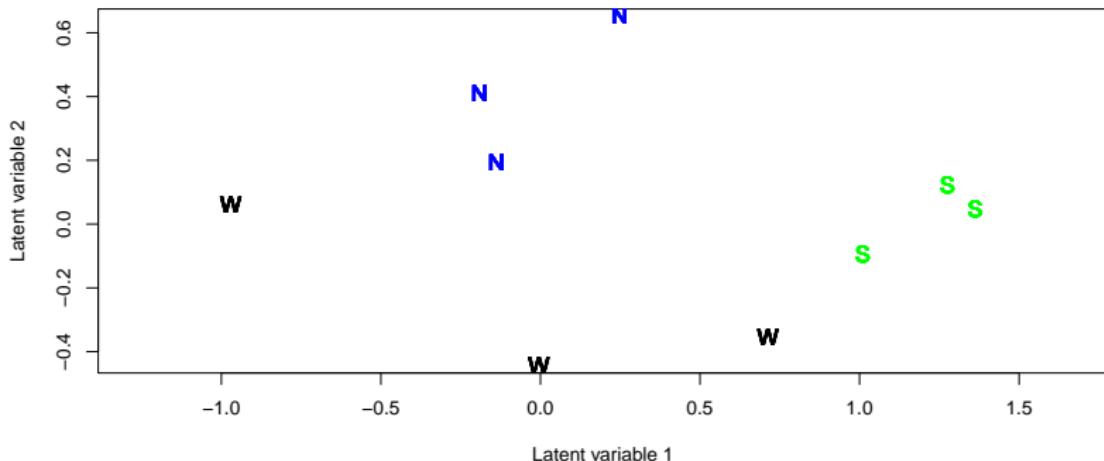
## Group-level unconstrained ordination

```
cord1 <- gllvm(Ya, X = X,
                  num.RR = 2, lv.formula = ~ (1|island), randomB = "LV",
                  family = "negative.binomial", disp.formula = rep(1,ncol(Ya)))
cord2 <- gllvm(Ya, X = X,
                  num.RR = 2, lv.formula = ~ (1|island/transect), randomB = "LV",
                  family = "negative.binomial", disp.formula = rep(1,ncol(Ya)))
```



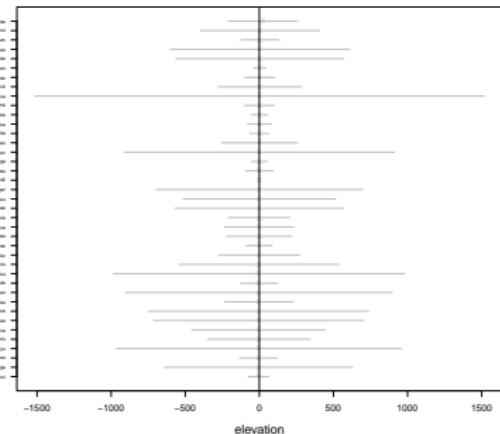
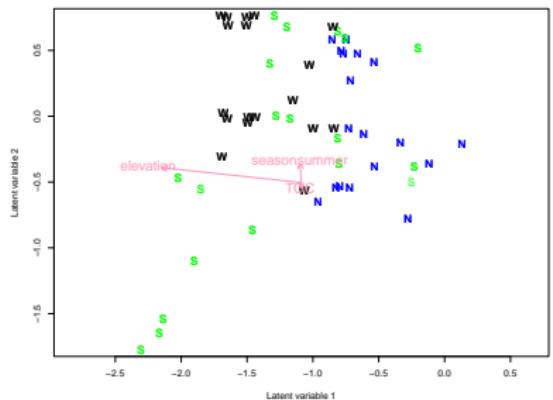
## Quadratic unconstrained ordination

```
cord3 <- gllvm(Ya, X = X, quadratic = TRUE,  
                 num.RR = 2, lv.formula = ~ (1|island/transect), randomB = "LV",  
                 family = "negative.binomial", disp.formula = rep(1, ncol(Ya)))
```



# Partial (fixed effects) constrained ordination

```
cord4 <- gllvm(Ya, X = X, studyDesign = X, num.RR = 2,
                 lv.formula = ~ elevation + silt_clay + season + TOC,
                 row.eff = ~ (1|island/transect),
                 family = "negative.binomial", disp.formula = rep(1,ncol(Ya)))
```



## Reduced-rank approximated effects

Since we have the relationship

$$\beta = \mathbf{B}\Gamma^\top \quad (11)$$

assuming normality and unbiasedness for both RHS quantities, we also have

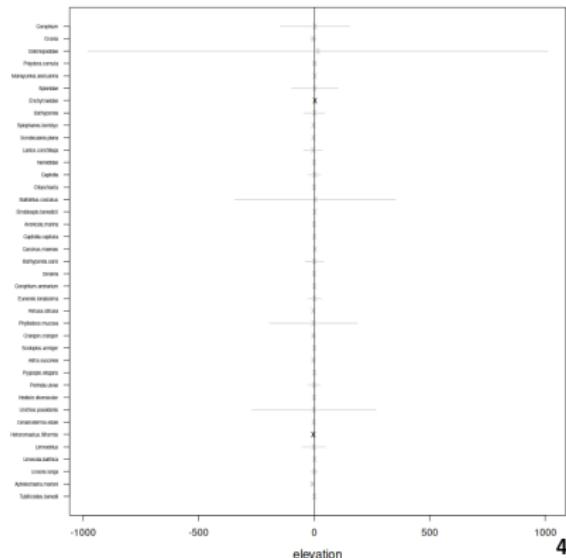
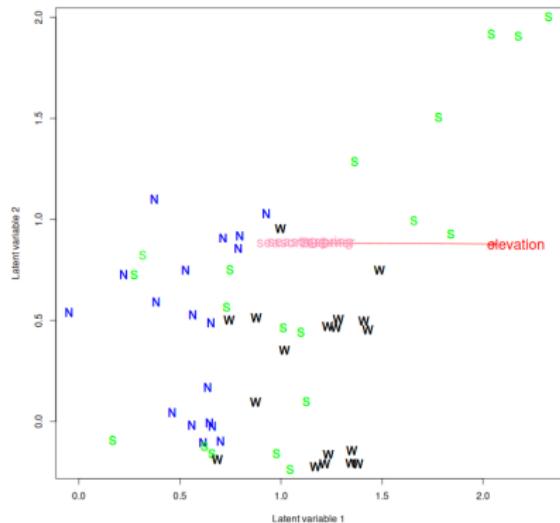
$$\begin{aligned} \text{var}(\hat{\mathbf{b}}_k \hat{\gamma}_j) &= \sum_{q=1}^d \sum_{r=1}^d \\ & b_{kj} b_{kr} \text{cov}(\hat{\gamma}_{qj}, \hat{\gamma}_{rj}) + \gamma_{rj} b_{kj} \text{cov}(\hat{\gamma}_{qj}, \hat{b}_{kr}) + \\ & \gamma_{qj} b_{kr} \text{cov}(\hat{\gamma}_{rj}, \hat{b}_{kj}) + \gamma_{qj} \gamma_{rj} \text{cov}(\hat{b}_{kj}, \hat{b}_{kr}) + \\ & \text{cov}(\hat{b}_{kj}, \hat{b}_{kr}) \text{cov}(\hat{\gamma}_{qj}, \hat{\gamma}_{rj}) + \text{cov}(\hat{\gamma}_{rj}, \hat{b}_{kj}) \text{cov}(\hat{\gamma}_{qj}, \hat{b}_{kr}) \end{aligned} \quad (12)$$

(for the quadratic model the solution is a bit more lengthy)

van der Veen et al. (2023)

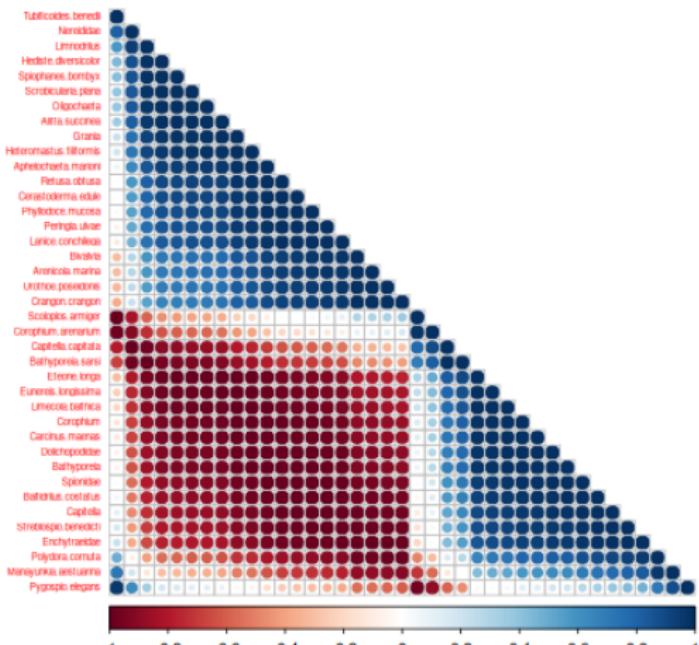
# Partial (random effects) constrained ordination

```
cord4b <- gllvm(Ya, X = X, studyDesign = X, num.RR = 2, randomB = "LV",
                  lv.formula = ~ (0 + elevation + silt_clay + season + TOC | 1),
                  row.eff = ~ (1 | island / transect),
                  family = "negative.binomial", disp.formula = rep(1, ncol(Ya)))
```



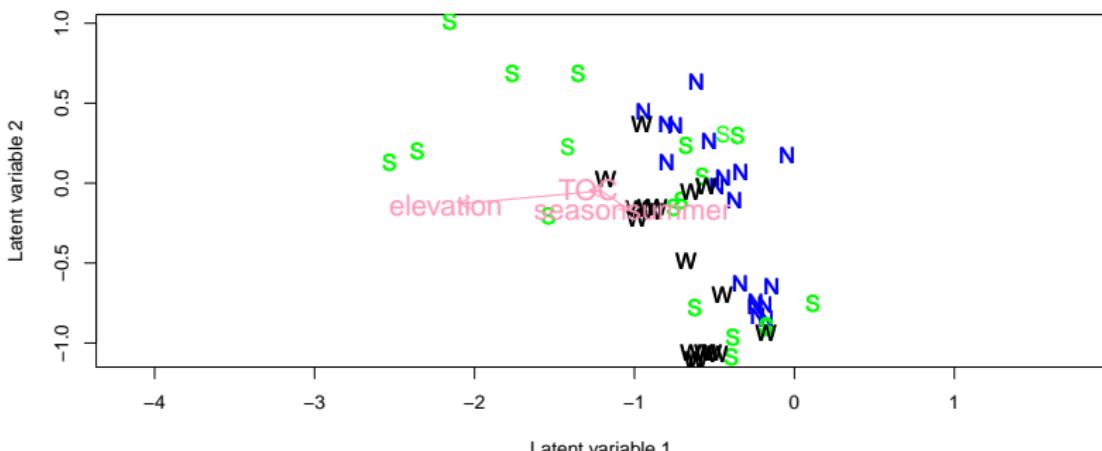
# Partial (random effects) constrained ordination

```
corrplot:::corrplot(getEnvironCor(cord4b), order = "AOE", type = "lower", tl.
```



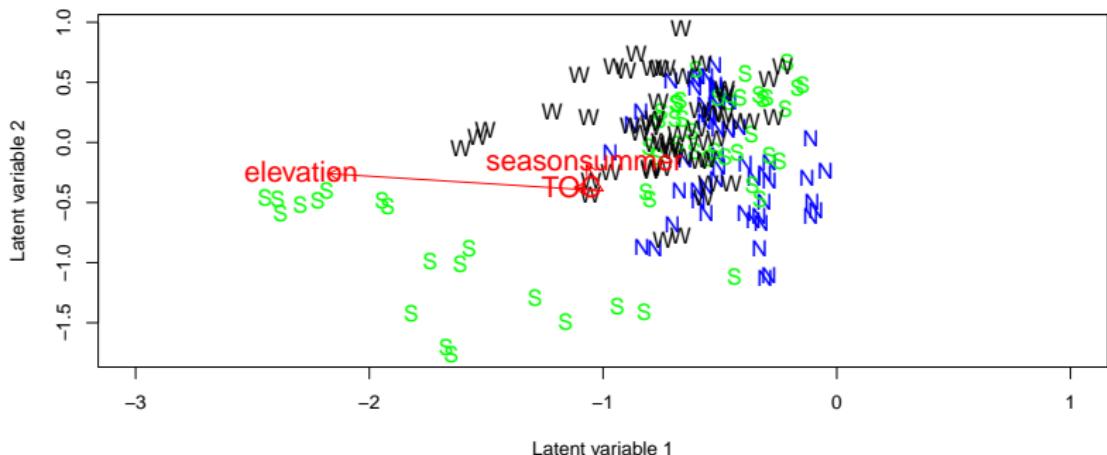
## Partial (fixed effects) constrained ordination

```
TMB::openmp(7, autopar=TRUE)
cord5 <- gllvm(Ya, X = X, studyDesign = X, num.RR = 2,
                lv.formula = ~ elevation + silt_clay + season + TOC,
                formula = ~diag(1|island/transect),
                family = "negative.binomial", disp.formula = rep(1, ncol(Ya)),
                Ab.struct = "diagonal", optim.method = "L-BFGS-B")
```



# Concurrent ordination

```
cnord1 <- gllvm(Ya, X = X, studyDesign = X, num.lv.c = 2,
                   lv.formula = ~ elevation + silt_clay + season + TOC,
                   formula = ~diag(1|island/transect),
                   family = "negative.binomial", disp.formula = rep(1,ncol(Ya)),
                   seed = 1, sd.errors = FALSE,
                   Ab.struct = "diagonal", optim.method = "L-BFGS-B")
```



Part IV

Big picture stuff

## Expanded reference model: fourth corner

$$\eta = r_0^\top \mathbf{1} + \mathbf{X}[\beta] + \boldsymbol{\Lambda} \boldsymbol{\Gamma}^\top \quad (13)$$

with hierarchical regressions for  $r_0$ ,  $\beta$ ,  $\boldsymbol{\Lambda}$  and  $\boldsymbol{\Gamma}$

## Expanded reference model: fourth corner

$$\eta = r_0^\top \mathbf{1} + \mathbf{X} \beta + \Lambda \Gamma^\top \quad (13)$$

with hierarchical regressions for  $r_0$ ,  $\beta$ ,  $\Lambda$  and  $\Gamma$

$$\beta = \mathbf{1}_m \beta_x^\top + \mathbf{T} \mathbf{R} \beta_{px} + \beta_\epsilon \quad (14)$$

## Expanded reference model: fourth corner

$$\eta = r_0^\top \mathbf{1} + \mathbf{X}[\beta] + \Lambda \Gamma^\top \quad (13)$$

with hierarchical regressions for  $r_0$ ,  $\beta$ ,  $\Lambda$  and  $\Gamma$

$$[\beta] = \mathbf{1}_m \beta_x^\top + \mathbf{T} \mathbf{R} \beta_{px} + \beta_\epsilon \quad (14)$$

$\Lambda$  and  $\Gamma$  are both as in concurrent ordination

# Preprint

We wrote a preprint, and fitted the model in NIMBLE (de Valpine et al. 2017)

Hierarchical Ordination, A unifying framework for drivers of  
community processes

R.B. O'Hara<sup>12</sup>

B. van der Veen<sup>123</sup>

# Overview

$$\eta = \mathbf{1}\beta^\top + \mathbf{r}_0\mathbf{1}^\top + \mathbf{Z}\Sigma\Gamma^\top \quad (15)$$

The diagram illustrates the components of the equation. The first term,  $\mathbf{1}\beta^\top$ , is highlighted in red and labeled 'Intercepts' with an arrow. The second term,  $\mathbf{r}_0\mathbf{1}^\top$ , is highlighted in blue and has no specific label. The third term,  $\mathbf{Z}\Sigma\Gamma^\top$ , is highlighted in yellow and labeled 'Ordination' with an arrow. Dashed arrows point from the labels to their respective terms.

Even more effects go into the ordination

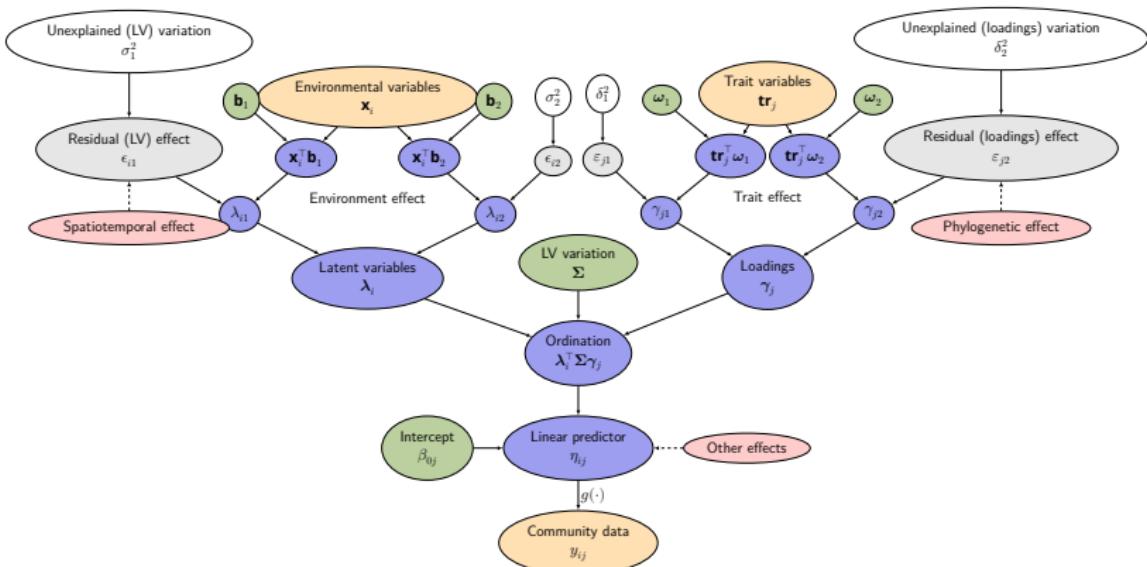
**Pros:** fewer parameters, easy visualization

**Cons:** generally difficult to fit/estimate

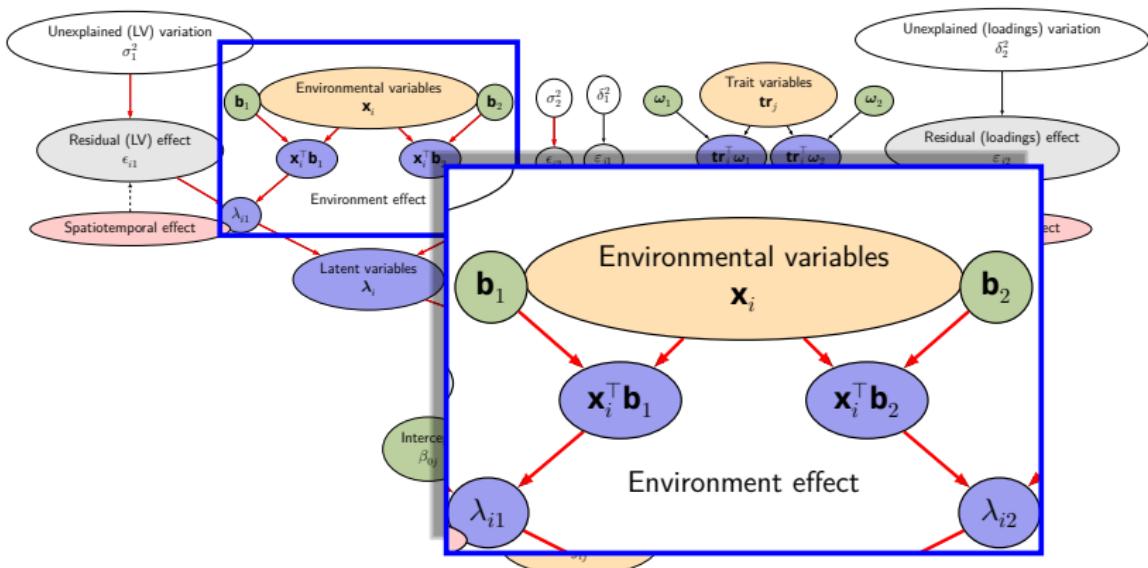
## Reduced-rank representation

$$\Lambda \Sigma \Gamma^\top = \underbrace{\mathbf{X} \mathbf{B} \Sigma \mathcal{E}^\top}_{\text{Main effects}} + \overbrace{\mathbf{X} \mathbf{B} \Sigma \Omega^\top \mathbf{T} \mathbf{R}^\top}^{\text{interaction terms}} + \underbrace{\mathbf{E} \Sigma (\mathbf{T} \mathbf{R} \Omega + \mathcal{E})^\top}_{\text{Other stuff}} \quad (16)$$

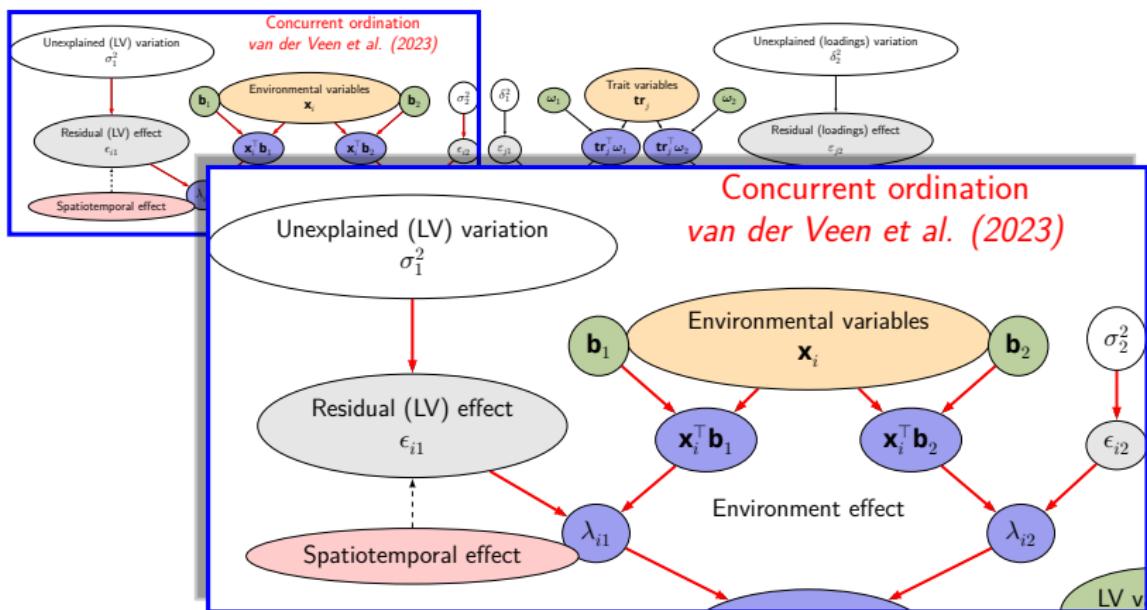
# Hierarchical ordination



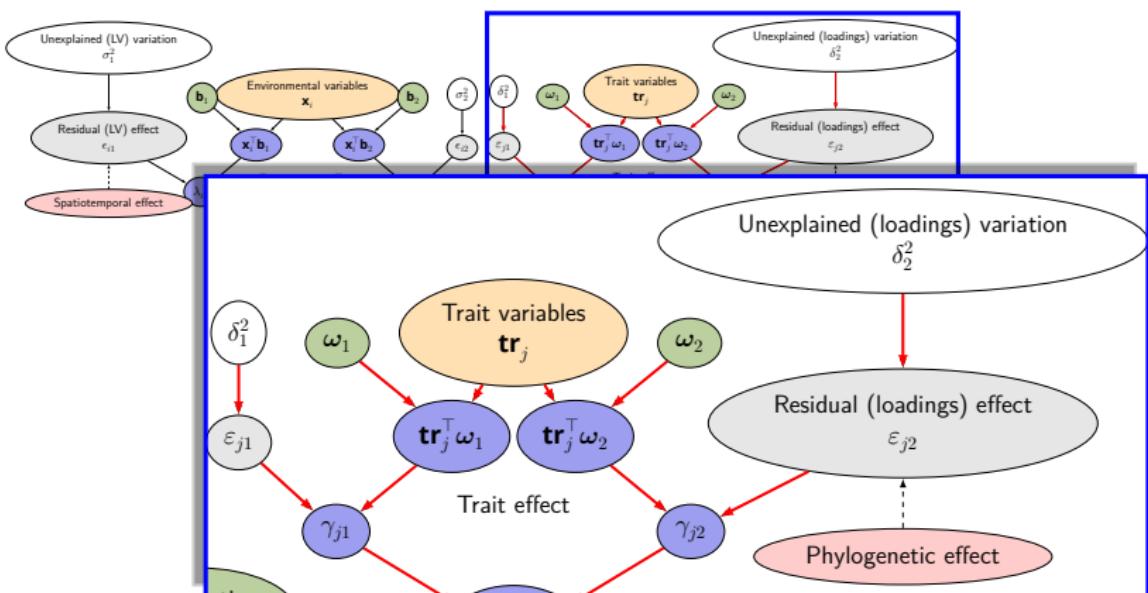
# Hierarchical ordination



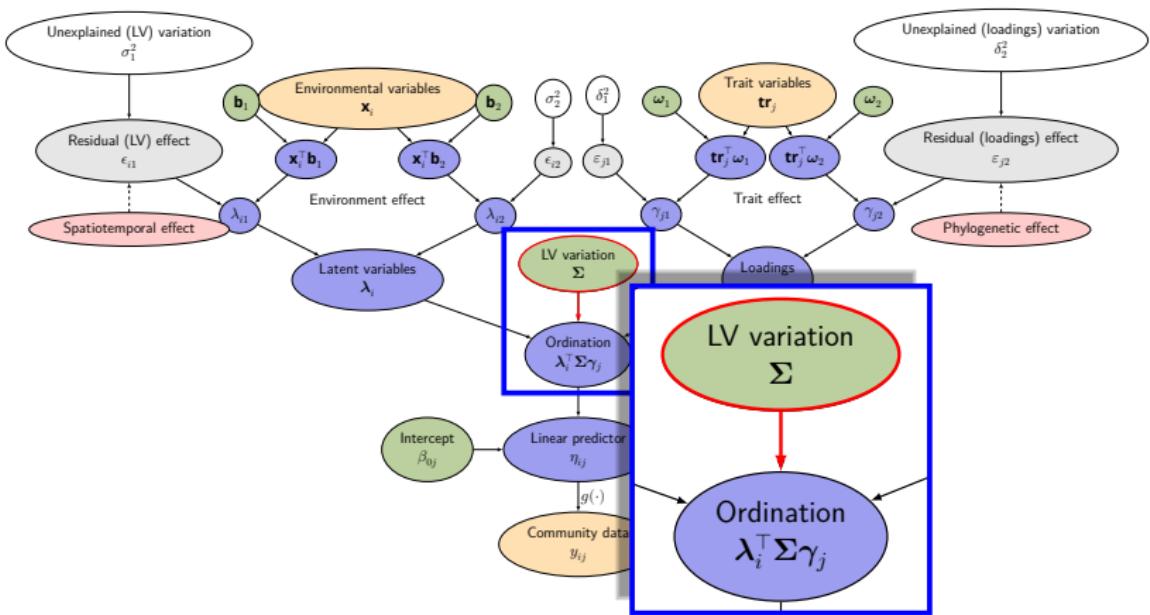
# Hierarchical ordination



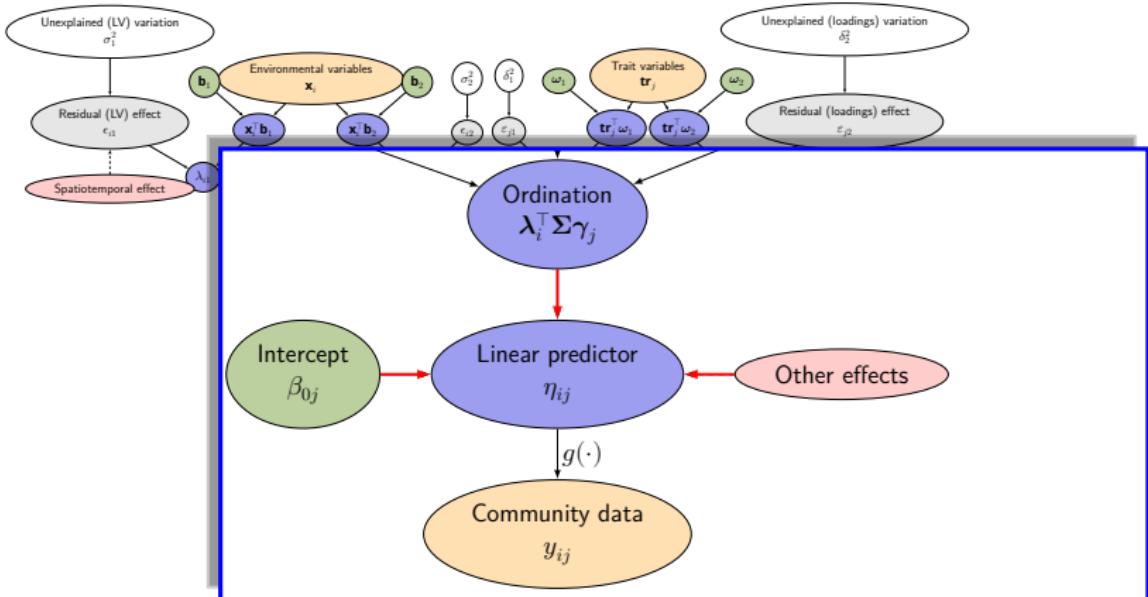
# Hierarchical ordination



# Hierarchical ordination



# Hierarchical ordination



# Challenges

**The more complex the model, the more exacerbated the issues**

Ordinary GLLVMs are hard to fit:

- ▶ Local convergence/multimodality of the likelihood
- ▶ Sensitivity to starting values
- ▶ Computationally intensive
- ▶ Complex methods with a steep learning curve

This is exacerbated when we add products of random effects  
(quadratic models and hierarchical ordination)

## Solutions

1. We already deal with local convergence issues, that's why we refit models multiple times
2. We already deal with the limitations of numerical integration, that's why we implement multiple methods
3. We are working to deal with usability and learning curve issues, that's why I give workshops

A lot of work remains to be done, on all three of these topics.  
But, we are limited by the state-of-the-art.

# Workshops

## GLLVMs: Advanced multivariate analysis of ecological communities in R

### Physalia GLLVM workshop

Bert van der Veen

This repository includes material for the Physalia workshop on Generalized linear Latent Variable Models, 10-13 June 2024. Feel free to share, alter, or re-use this material with appropriate referencing of this repository.

Workshop webpage: <https://www.physalia-courses.org/courses-workshops/gllvm/>

**Figure 7:** <https://github.com/BertvanderVeen/GLLVM-workshop>

**Analysing multivariate ecological data with Generalized Linear Latent Variable Models: 2-part workshop**

*Bert van der Veen, Norwegian Institute of Bioeconomy Research*

*Robert B. O'Hara, Norwegian University of Science and Technology*

*Sam Perrin, Norwegian University of Science and Technology*

*Jenni Niku, University of Jyväskylä*

## A broad class of ordination methods

- ▶ The framework that ecologists could only dream of for decades
- ▶ GLLVMs represent a broad class of ordination methods
- ▶ For a wide range of data types
- ▶ Everything classical methods can do, and more
- ▶ We can adjust to model to suit the data (generating process)
- ▶ There remain technical challenges
- ▶ Which is good, otherwise I would be out of a job

