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

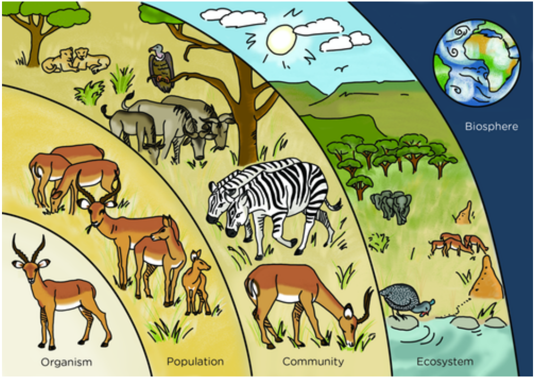
I will try to convince you that community ecology needs these methods.

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

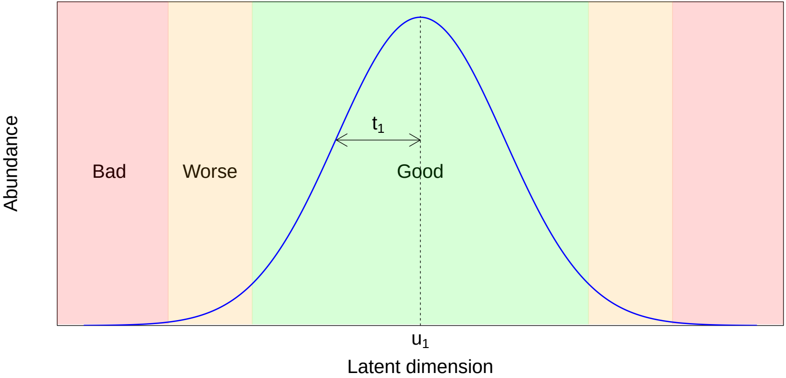
Gathering data

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

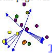


A brief history of ordination

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 revives Correspondence Analysis
- 1976 - Canonical Correlation introduced to 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.



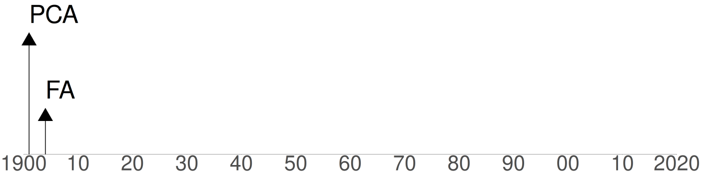
To the ordination web page

Historical Perspective

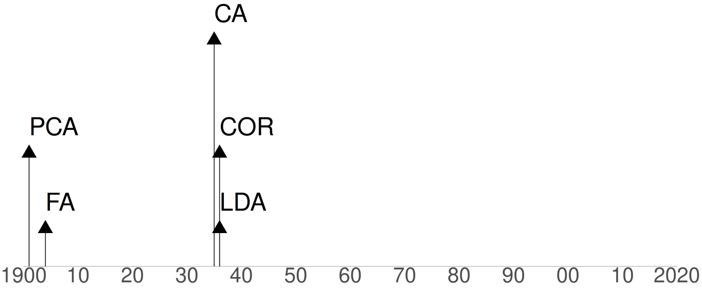
1901	-	Pearson develops PCA as a regression technique.
1927	-	Spearman applies factor analysis to psychology.
1930	-	Ramensky uses an informal ordination technique and introduces the term 'ordnung' into ecology.
1954	-	D.W. Goodall introduces PCA into ecology and proposes the term 'ordination'.
1970	-	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)

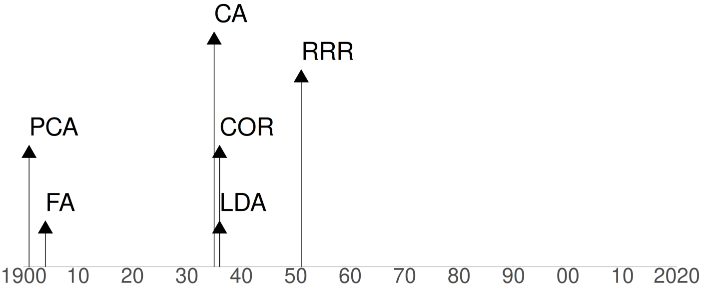
In the beginning, there were PCA and FA



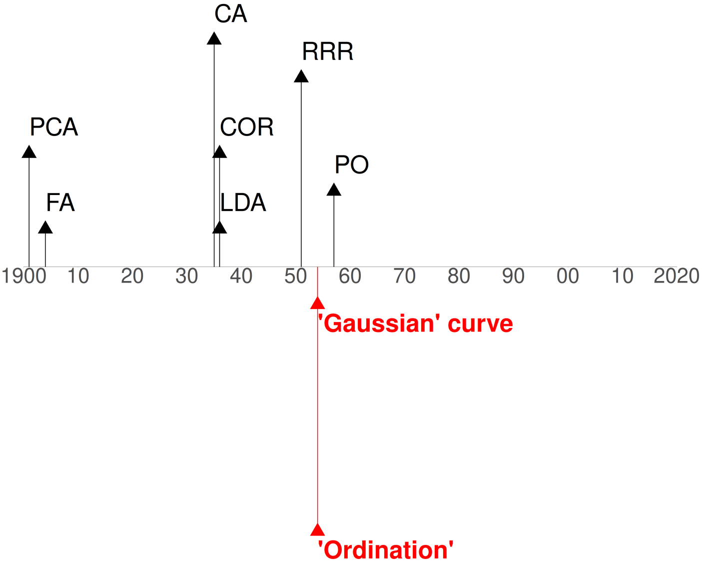
Then, CA was developed



Next, reduced rank regression



Then, there was "Ordination"



Whittaker 1956: Unimodal responses

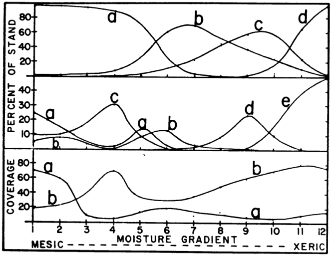
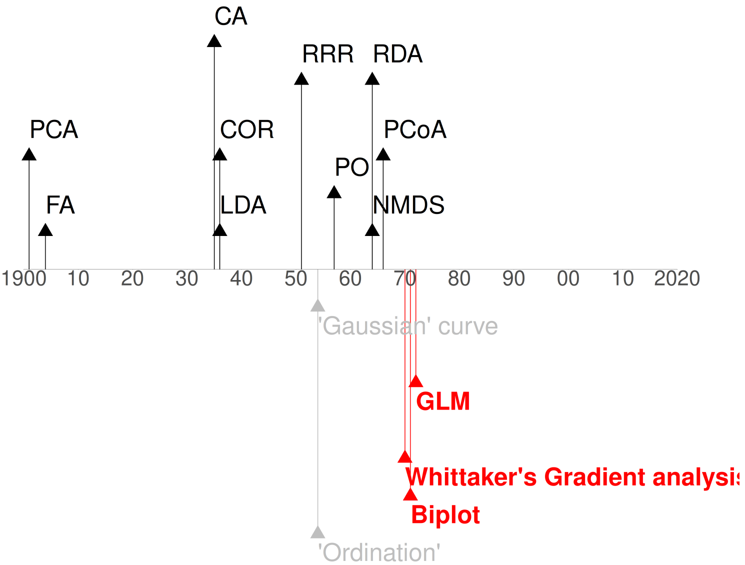
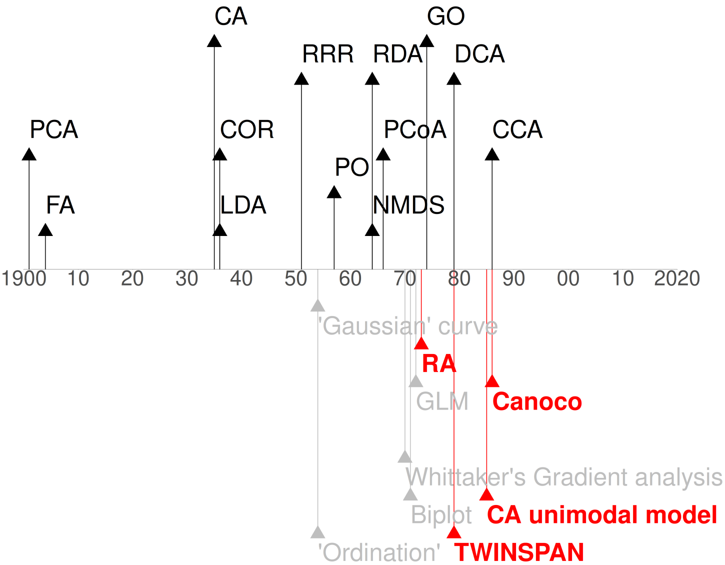


FIG. 4. Transect of the moisture gradient, 3500-4500 ft. Top—curves for tree classes; a, mesic; b, submesic; c, subxeric; d, xeric. Note expansion of mesic stands, compared with Figs. 2 and 3. Middle—curves for tree species: a, *Tilia heterophylla*; b, *Halesia monticola* (both the preceding are bimodal, 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[1]

HUGH G. GAUCH, JR.
Ecology and Systematics, Cornell University, Ithaca, New York 14850
GENE B. CHASE[2]
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)

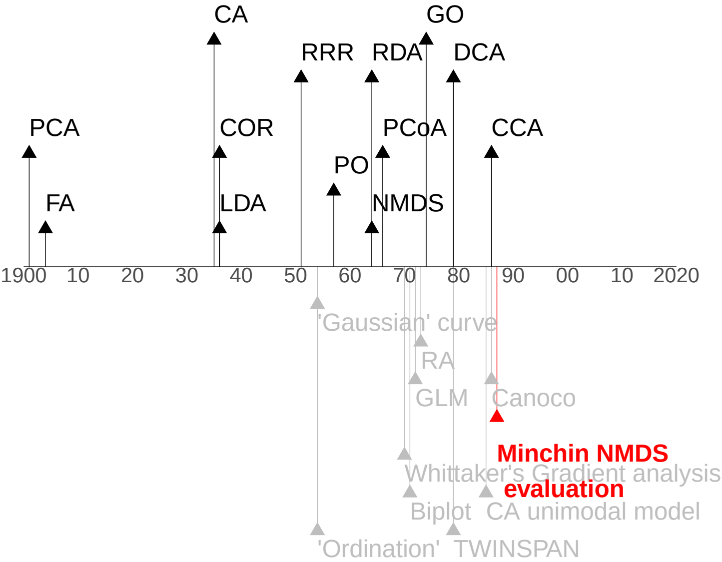
BIOMETRICS 41, 859–873
December 1985

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

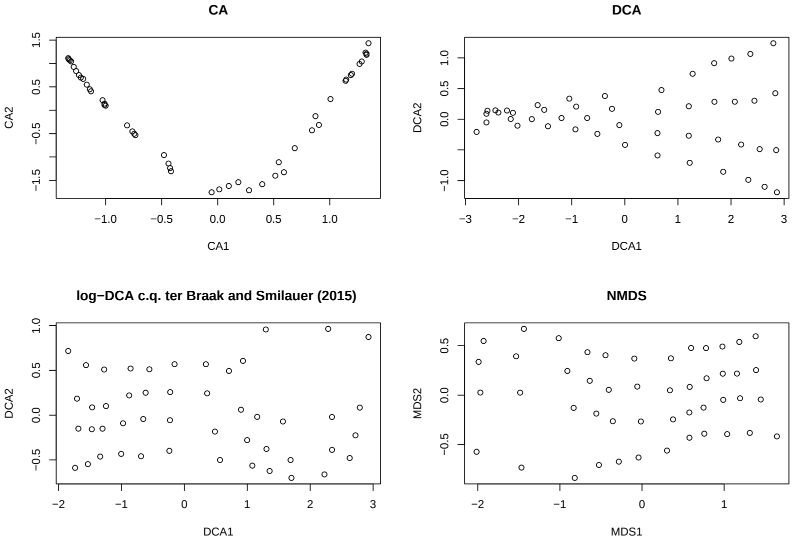
89

An evaluation of the relative robustness of techniques for ecological ordination

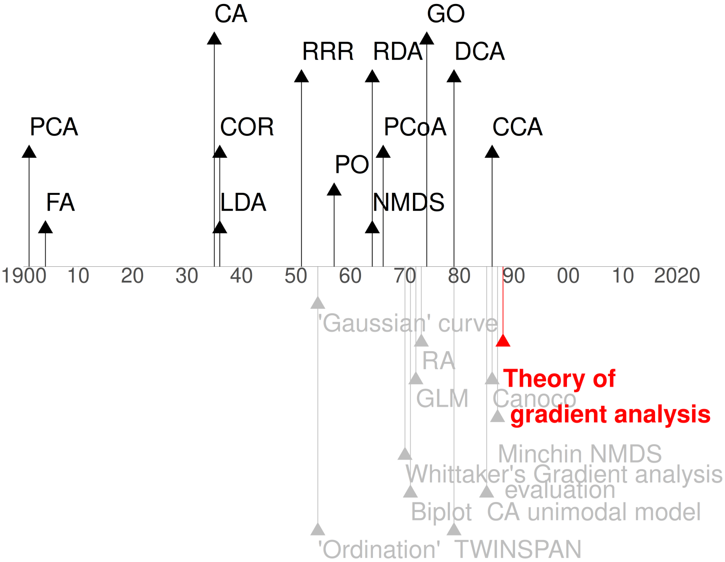
Peter R. Minchin*
CSIRO Division of Water and Land Resources, G.P.O. Box 1666, Canberra, 2601, Australia

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

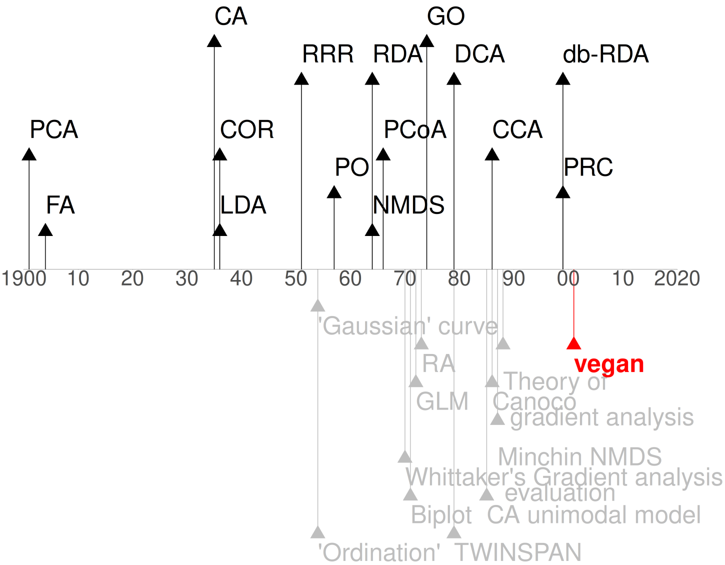
Results Minchin dataset



A theory for gradient analysis



vegan


CA
GO
RRR
RDA
DCA
db-RDA
PCA
COR
PCoA
CCA
PO
PRC
FA
LDA
NMDS
1900
10
20
30
40
50
60
70
80
90
00
10
2020
'Gaussian' curve
RA
vegan
Theory of
GLM
Canoco
gradient analysis
Minchin NMDS
Whittaker's Gradient analysis
evaluation
Biplot
CA unimodal model
'Ordination'
TWINSPAN

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 underlaying model)

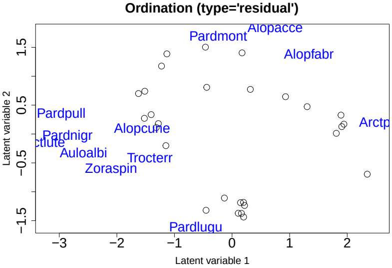


Figure 2: ordination of spider data

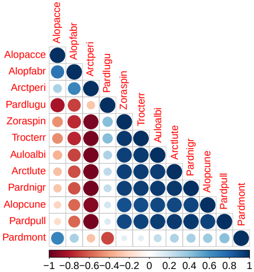


Figure 3: correlations of spider data

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 BRITISH ECOLOGICAL SOCIETY

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 BRITISH ECOLOGICAL SOCIETY

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

ECOLOGICAL MONOGRAPHS
ECOLOGICAL SOCIETY OF AMERICA

Article

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

Steven C. Walker, Donald A. Jackson

Methods in Ecology and Evolution — BRITISH ECOLOGICAL SOCIETY

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,[1,*] F. Guillaume Blanchet,[2] Robert B. O'Hara,[3] Otso Ovaskainen,[4,5] Sara Taskinen,[6] Steven C. Walker,[2] and Francis K.C. Hui[7]

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[1]

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

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

Response distribution

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

1. Community data

Response distribution

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

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

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

4. Linear predictor ("the model")

Response distribution

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

5. Dispersion parameter

The reference model

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

The reference model

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

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

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

3. Sample (row) intercept

The reference model

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

4. Fixed effects

The reference model

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

5. Random effects

The reference model

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

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

$$g\{\mathbb{E}(y_{ij}|r_{0} \quad \ldots \quad \mathbf{u}_j + \epsilon_{ij} \tag{2}$$

Requires fast methods for integration

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

The reference model

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

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

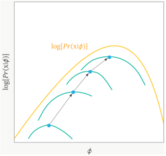


Figure 4: Variational Approximations borealisai.com

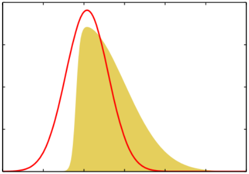


Figure 5: Laplace approximation wiljohn.top

The reference model

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

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

Usually, we cannot fit this

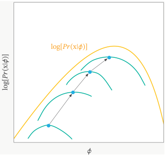


Figure 4: Variational Approximations borealisai.com

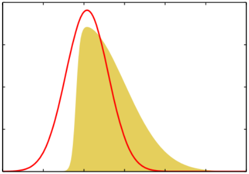


Figure 5: Laplace approximation wiljohn.top

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

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \mathbf{E}, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{4}$$

This is a GLMM Induces between-column correlation

Generalized Linear (Mixed-effects) Models

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + ... + \mathbf{E}, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top \tag{4}$$

This is an unconstrained ordination

Generalized Linear (Mixed-effects) Models

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \dots + \mathbf{E}, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boxed{\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top} \tag{4}$$

This is an unconstrained ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \dots + \boxed{\mathbf{X}\mathbf{B}} \tag{5}$$

This is a GLM

Generalized Linear (Mixed-effects) Models

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \mathbf{E}, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top \tag{4}$$

This is an unconstrained ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \mathbf{X}\mathbf{B}_{lv}\boldsymbol{\Gamma}^\top \tag{5}$$

This is a constrained ordination

Generalized Linear (Mixed-effects) Models

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + ... + \boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top, \qquad \boldsymbol{\lambda}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{4}$$

This is an unconstrained ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + ... + \boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top \tag{5}$$

This is a constrained ordination

Generalized Linear (Mixed-effects) Models

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top, \qquad \boldsymbol{\lambda}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{4}$$

This is an unconstrained ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top \tag{5}$$

This is a constrained ordination

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

A unifying framework for ordination

Concurrent ordination: Simultaneous unconstrained and constrained latent variable modelling

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

Concurrent ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \dots + \boldsymbol{\Lambda}\ \boldsymbol{\Gamma}^\top \tag{6}$$

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

Concurrent ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \boldsymbol{\Lambda}\ \boldsymbol{\Gamma}^\top \tag{6}$$

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

Concurrent ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \boldsymbol{\Lambda}\ \boldsymbol{\Gamma}^\top \tag{6}$$

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

1. **$\boldsymbol{\Lambda}$ = E, unconstrained**

Concurrent ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \boldsymbol{\Lambda}\ \boldsymbol{\Gamma}^\top \tag{6}$$

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

1. $\boldsymbol{\Lambda}$ $=$ **E**, **unconstrained**
2. $\boldsymbol{\Lambda}$ $=$ **XB**, **constrained**

Concurrent ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \boldsymbol{\Lambda}\ \boldsymbol{\Gamma}^\top \tag{6}$$

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

1. $\boldsymbol{\Lambda} = \mathbf{E}$, **unconstrained**
2. $\boldsymbol{\Lambda} = \mathbf{XB}$, **constrained**
3. $\boldsymbol{\Lambda} = \mathbf{XB} + \mathbf{E}$, **<u>concurrent</u>**

Concurrent ordination

$$\boldsymbol{\eta} = \mathbf{1}\boldsymbol{\beta}^\top + \ldots + \boldsymbol{\Lambda}\ \boldsymbol{\Gamma}^\top \tag{6}$$

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

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

1.
2. $\boldsymbol{\Lambda} = \mathbf{XB}$, **constrained**
3. $\boldsymbol{\Lambda} = \mathbf{XB} + \mathbf{E}$, <u>**concurrent**</u>

Unimodal responses

Vanilla GLLVM

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

Unimodal responses

Vanilla GLLVM

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

Quadratic extension

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

- 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 $\mathrm{D}_{jqq} = \mathrm{d}\ \forall \mathrm{q,j}$

Unimodal responses

Vanilla GLLVM

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

Quadratic extension

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

- 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 $\mathsf{D}_{jqq} = \mathsf{d} \; \forall \mathsf{q,j}$

Residual covariance

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

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

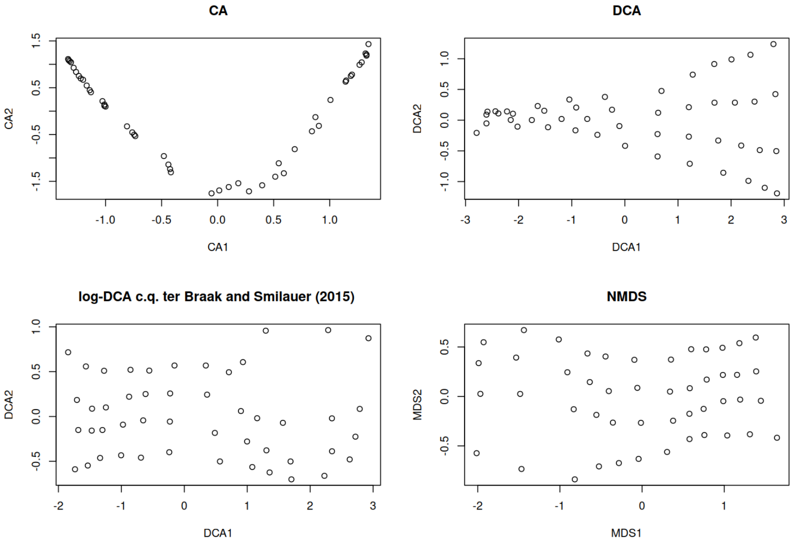
Robustness

Vegetatio 69: 89–107, 1987

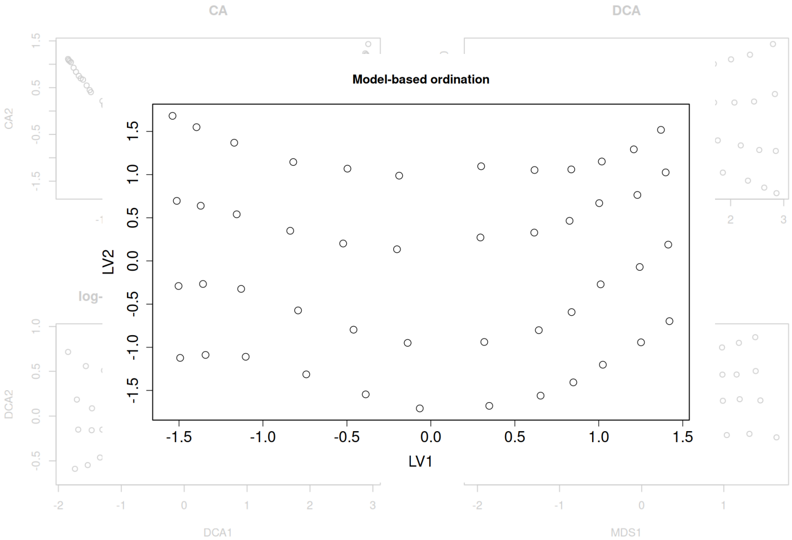
89

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



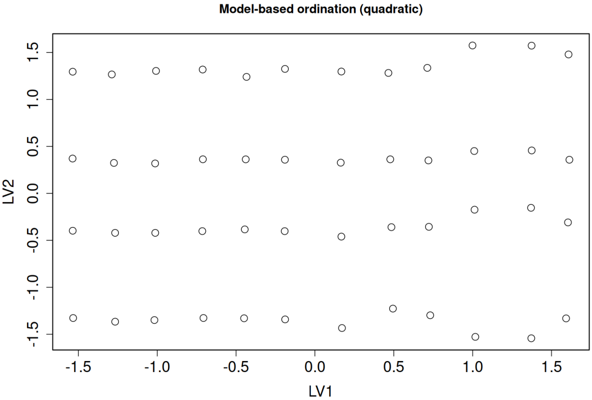
CA
CA1
CA2
DCA
DCA1
DCA2
log-DCA c.q. ter Braak and Smilauer (2015)
DCA1
DCA2
NMDS
MDS1
MDS2


Model-based ordination
LV1
LV2
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
CA
DCA
CA2
DCA1
DCA2
MDS1

Robustness

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

Robustness


Model-based ordination (quadratic)
LV2
LV1
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

Part III

Getting methods to the people

gllvm

Received: 7 May 2019 | Accepted: 5 September 2019
DOI: 10.1111/2041-210X.13303

APPLICATION

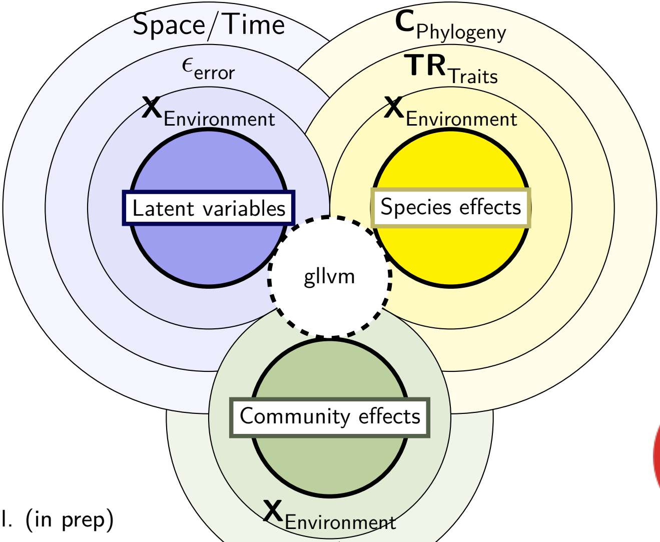
Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

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

Jenni Niku[1] | Francis K. C. Hui[2] | Sara Taskinen[1] | David I. Warton[3]

- 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



Korhonen et al. (in prep)

Response distributions

Distributions for all common data types:

Continuous

- Gaussian

Bounded

- exponential
- gamma
- Tweedie
- beta

Discrete

- Poisson
- NB[1]
- binomial
- ordinal

Zero-inflated

- Poisson
- NB[1]
- hurdle beta
- ordered beta

[1] 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)

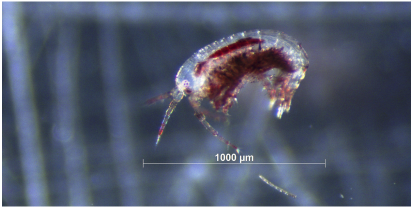
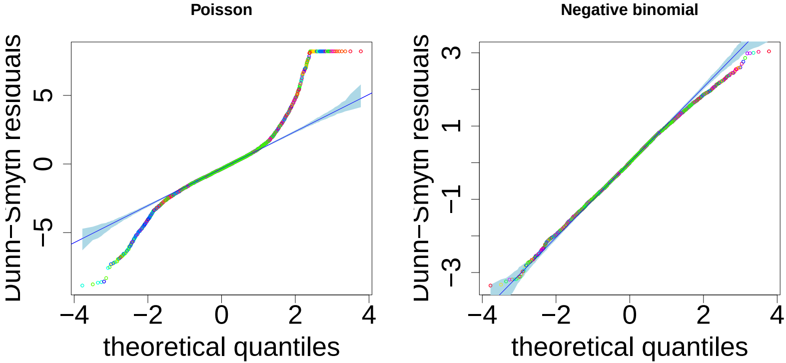


Figure 6: nioz.nl

```
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[,unlist(lapply(X,is.numeric))] <- scale(X[,unlist(lapply(X,i
```

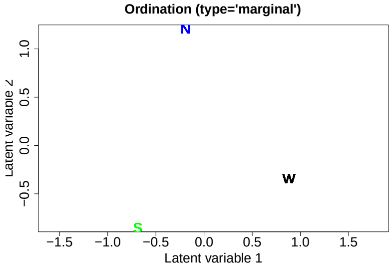
Unconstrained ordination

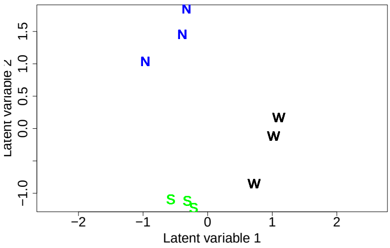
```r
uord1 <- gllvm(Ya, num.lv = 2, family = "poisson")
uord2 <- gllvm(Ya, num.lv = 2, family = "negative.binomial")
```



Group-level unconstrained ordination

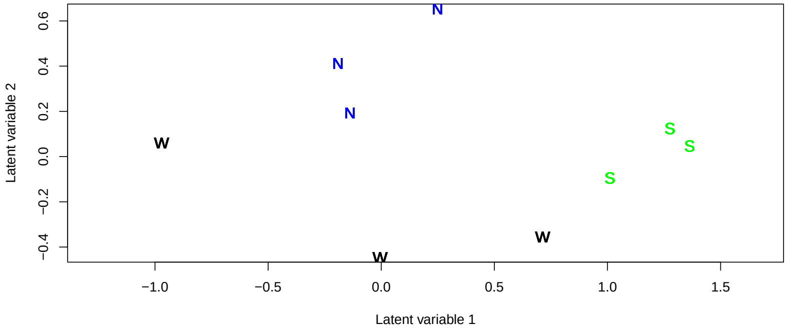
```
cord1 <- gllvm(Ya, X = X,
              num.RR = 2, lv.formula = ~(1|island), randomB = "LV",
              family = "negative.binomial", disp.formula = rep(1,ncol(Ya)))
cord2 <- gllvm(Ya, X = X,
              num.RR = 2, lv.formula = ~(1|island/transect), randomB = "LV",
              family = "negative.binomial",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





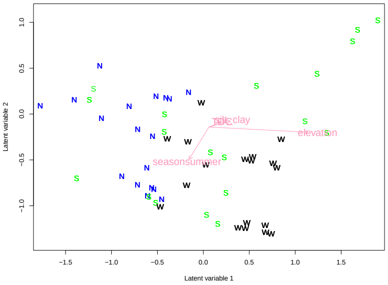
Quadratic unconstrained ordination

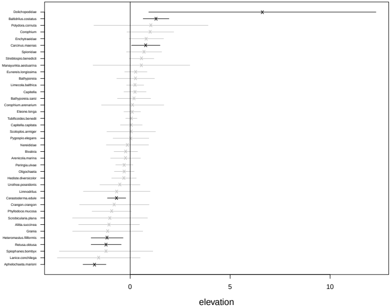
```
cord3 <- gllvm(Ya, X = X, quadratic = TRUE,
               num.RR = 2, lv.formula = ~(1|island/transect), randomB = "LV",
               family = "negative.binomial", disp.formula = rep(1,ncol(Ya)))
```



Partial (fixed effects) constrained ordination

```r
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

$$\boldsymbol{\beta} = \mathbf{B}\boldsymbol{\Gamma}^\top \tag{11}$$

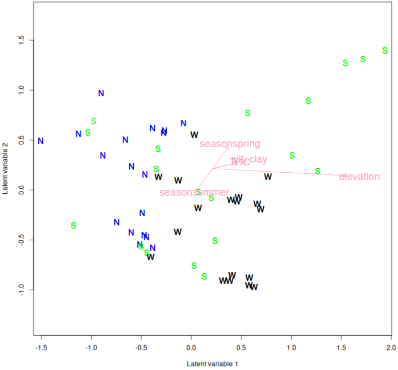
assuming normality and unbiasedness for both RHS quantities, we also have

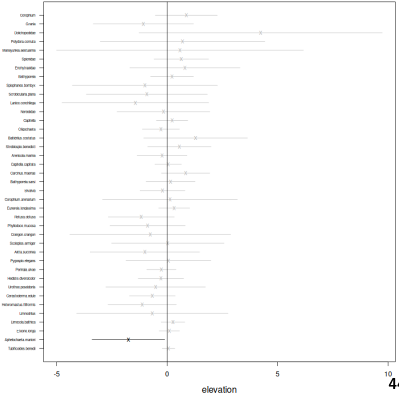
$$\begin{aligned}\text{var}(\widehat{\mathbf{b}_k \gamma_j}) = &\sum_{q=1}^{d}\sum_{r=1}^{d} \\ & b_{kq}b_{kr}\text{cov}(\widehat{\gamma_{qj}}, \widehat{\gamma_{rj}}) + \gamma_{rj}b_{kq}\text{cov}(\hat{\gamma}_{qj}, \hat{b}_{kr}) + \\ & \gamma_{qj}b_{kr}\text{cov}(\hat{\gamma}_{rj}, \hat{b}_{kq}) + \gamma_{qj}\gamma_{rj}\text{cov}(\hat{b}_{kq},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

(for the quadratic model the solution is a bit more lengthy) van der Veen et al. (2023)

Partial (random effects) constrained ordination

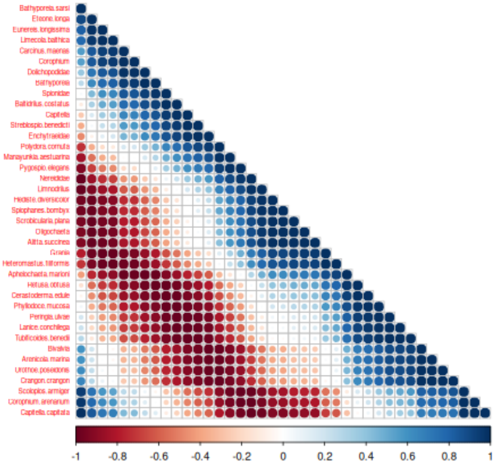
```
cord4b <- gllvm(Ya, X = X, studyDesign = X, num.RR = 2, randomB = "LV",
           lv.formula = ~ (0+elevation + silt_clay + season + TOC|1),
           row.eff = ~(1|island:transect), optimizer = "nlminb",
           family = "negative.binomial", disp.formula = rep(1,ncol(Ya)))
```





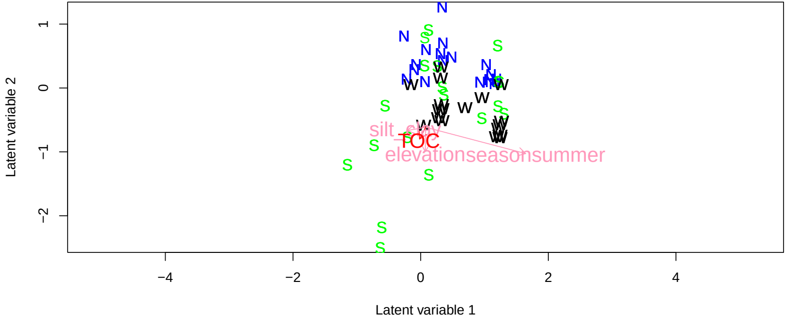
Partial (random effects) constrained ordination

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



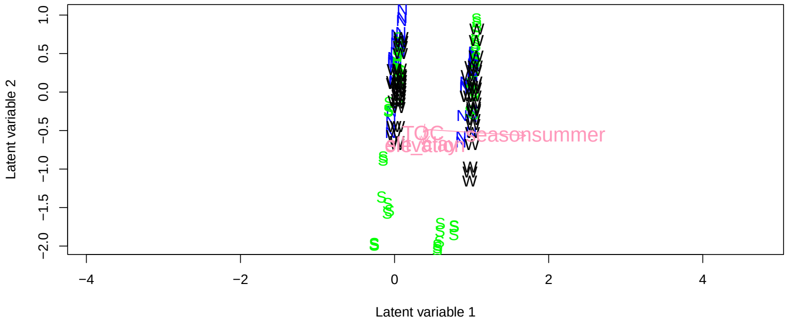
Partial (fixed effects) constrained ordination

```r
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), starting.val = "zero",
            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,
            Ab.struct = "diagonal", optim.method = "L-BFGS-B")
```



Part IV

Big picture stuff

Expanded reference model: fourth corner

$$\boldsymbol{\eta} = \mathbf{r}_0^\top \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top \tag{13}$$

with hierarchical regressions for $\mathbf{r}_0$, $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$

Expanded reference model: fourth corner

$$\boldsymbol{\eta} = \mathbf{r}_0^\top \mathbf{1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\Lambda\Gamma}^\top \tag{13}$$

with hierarchical regressions for $\mathbf{r}_0$, $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$

$$\boldsymbol{\beta} = \mathbf{1}_m \boldsymbol{\beta}_x^\top + \mathbf{TR}\boldsymbol{\beta}_{px} + \boldsymbol{\beta}_\epsilon \tag{14}$$

Expanded reference model: fourth corner

$$\boldsymbol{\eta} = \mathbf{r}_0^\top \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top \tag{13}$$

with hierarchical regressions for $\mathbf{r}_0$, $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$

$$\boldsymbol{\beta} = \mathbf{1}_m \boldsymbol{\beta}_x^\top + \mathbf{T}\mathbf{R}\boldsymbol{\beta}_{px} + \boldsymbol{\beta}_\epsilon \tag{14}$$

$\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ 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[1,2] B. van der Veen[1,2,3]

Overview

$$\boxed{\boldsymbol{\eta}} = \underbrace{\boxed{\mathbf{1}\boldsymbol{\beta}^\top} + \boxed{\mathbf{r}_0\mathbf{1}^\top}}_{\text{Intercepts}} + \underbrace{\boxed{\mathbf{Z}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top}}_{\text{Ordination}} \tag{15}$$

Even more effects go into the ordination

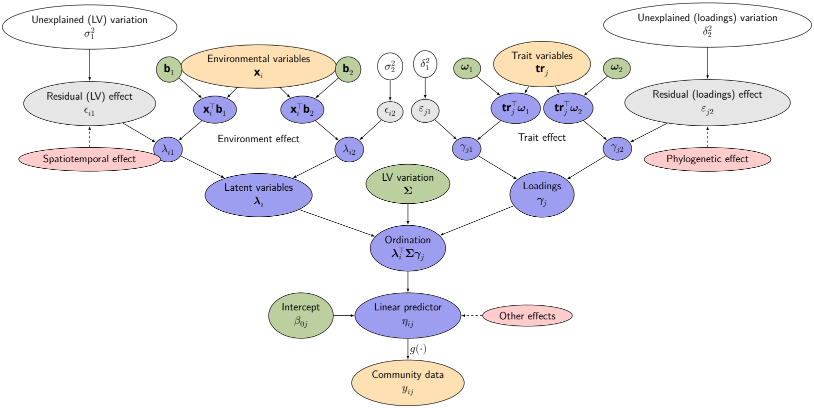
Pros: fewer parameters, easy visualization

Cons: generally difficult to fit/estimate

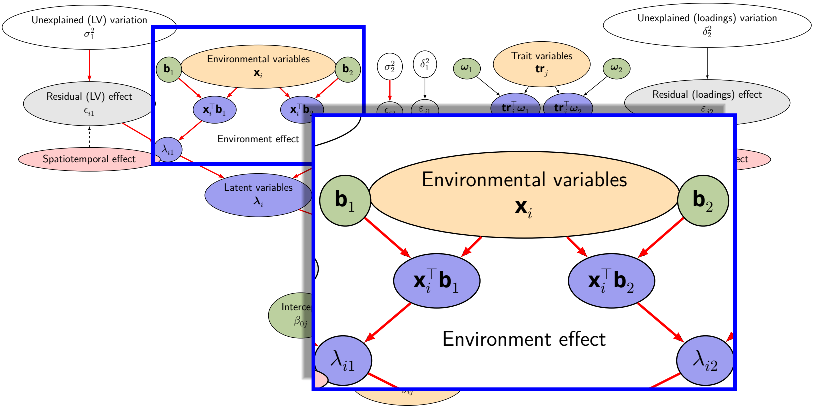
Reduced-rank representation

$$\mathbf{\Lambda}\,\mathbf{\Sigma}\,\mathbf{\Gamma}^\top = \underbrace{\mathbf{XB}\,\mathbf{\Sigma}\,\mathcal{E}^\top}_{\text{Main effects}} + \overbrace{\mathbf{XB}\,\mathbf{\Sigma}\,\mathbf{\Omega}^\top\mathbf{TR}^\top}^{\text{interaction terms}} + \underbrace{\mathbf{E}\,\mathbf{\Sigma}\,(\mathbf{TR\Omega}+\mathcal{E})^\top}_{\text{Other stuff}} \tag{16}$$

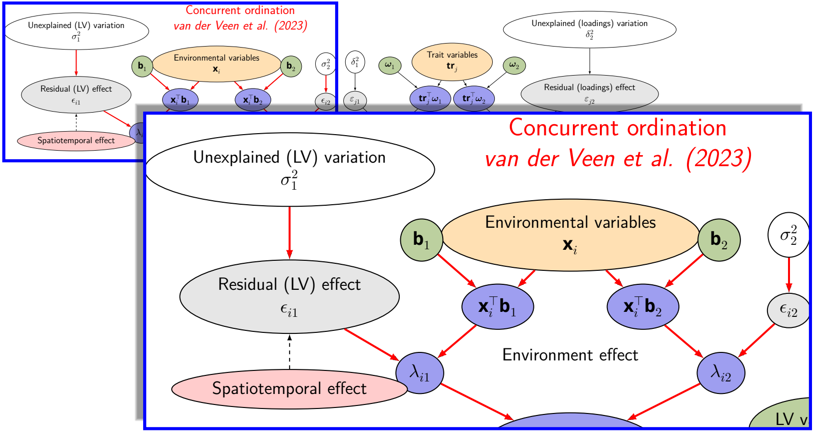
Hierarchical ordination



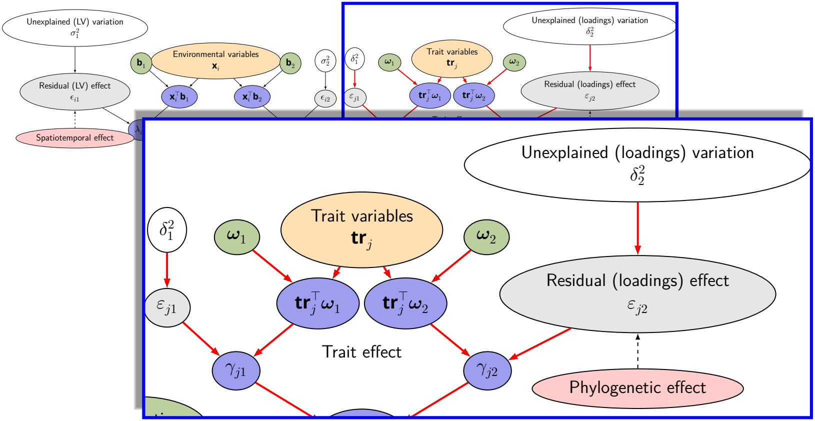
Hierarchical ordination



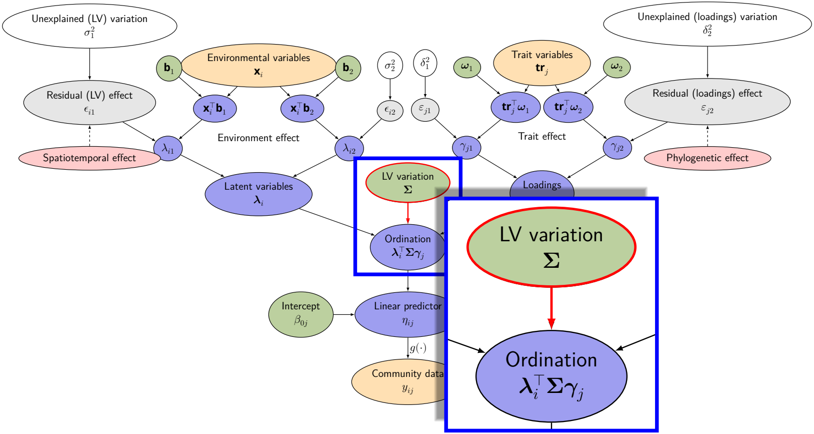
Hierarchical ordination



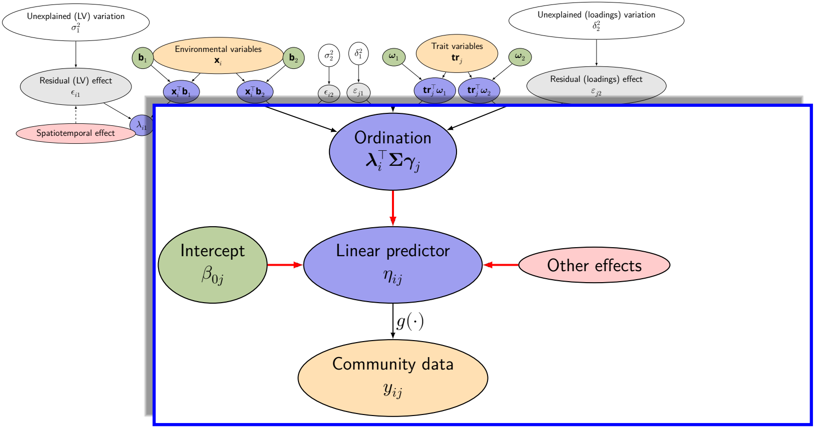
Hierarchical ordination



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


MAKE ORDINATION
GREAT AGAIN
imgflip.com