

# Introduction

Bert van der Veen

Department of Mathematical Sciences, NTNU

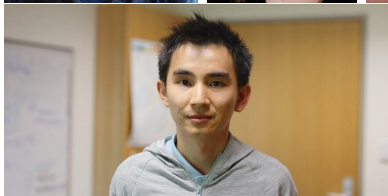
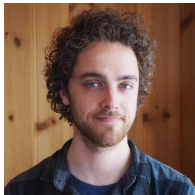
# Welcome!

---



# Who we are

---



# Logistics

---

- ▶ All happy with the rooms?
- ▶ Meals are covered by the summer school
- ▶ Lunch at 12:00, dinner at 18:00

Please keep an eye on slack.

## Credits and attendance

---

- ▶ ECTS require applying and an exam
- ▶ Can also opt in for “just” a letter of attendance

I will post a message on slack.

# Introduction

---

See [github](#) for all material

Sessions from 9:00 to 17:00. Sessions will consist of a mix of lectures, in-class discussion, and practical exercises / case studies over Slack and Zoom. And after dinner (?)

- ▶ Monday: Properties of community data, VGLMs, Model checking, intro to glvm
- ▶ Tuesday: VGLMMs, JSDM, hierarchical environmental responses
- ▶ Wednesday: Model-based ordination, Ordination with covariates, Conditioning, Unimodal responses
- ▶ Workflow A-Z, presentations, hike, article reanalysis
- ▶ Friday: Model-based clustering, Other packages, article reanalysis, analysis of own data

## How we will do it

---

Lectures of about 45 minutes

Practicals of about 60 minutes: datasets and R

- ▶ Practical “tasks”
- ▶ Potentially short, live, demonstrations

Questions: in person or on slack (“day channels”)

# Presentations

---

I will still make a schedule for Thursday.

- ▶ Low key
- ▶ Does not require slides (but nice if you have them)
- ▶ To tell everyone a little about your data
- ▶ And possibly think about what models to fit



## What I hope you take away

---

1. The gllvm R-package is great!
2. Performing multivariate analysis well is hard work
3. Model-based ordination methods do many things better (data properties, diagnostics)
4. One Method To Rule Them All

# Packages

---

Please make sure you've downloaded data and updated R/packages

## R-packages

---

For Friday:

- ▶ gllvm
- ▶ glmmTMB
- ▶ mvabund
- ▶ HMSC
- ▶ sjSDM
- ▶ ecopCopula
- ▶ borai

and some additional packages:

- ▶ DHARMA
- ▶ vegan
- ▶ labdsv
- ▶ ggplot2
- ▶ rnaturalearth
- ▶ rnaturalearthdata
- ▶ sf

## Some resources on classical ordination

---

- ▶ David Zeneley's website
- ▶ Michael Palmer's website
- ▶ Numerical ecology
- ▶ Numerical ecology with R
- ▶ Data analysis in Community and Landscape ecology
- ▶ Analysis of ecological communities

## Resources on model-based ordination

---

- ▶ Some of my other workshop repositories
- ▶ gllvm vignette website
- ▶ Oxford libraries article
- ▶ Warton 2022
- ▶ Fahrmeir and Tutz 2001
- ▶ Ovaskainen and Abrego
- ▶ Bartholomew et al. 2011
- ▶ Skrandal and Rabe-Hesketh 2004
- ▶ Zuur and Ieno (2025)

## Some recommended reading

---

- ▶ Halvorsen (2012)
- ▶ Wang et al. (2012)
- ▶ Warton et al. (2012)
- ▶ Clark et al. (2014)
- ▶ Warton et al. (2015)
- ▶ Warton et al. (2015)
- ▶ Hui et al. (2015)
- ▶ Pollock et al. (2015)
- ▶ ter Braak and Smilauer (2015)
- ▶ Hui et al. (2017)
- ▶ Niku et al. (2017)
- ▶ Ovaskainen et al. (2017)
- ▶ Roberts (2017)
- ▶ Warton et al. (2017)
- ▶ Niku et al. (2019)
- ▶ Niku et al. (2019)
- ▶ Roberts (2019)
- ▶ Paul (2020)
- ▶ Zurell et al. (2020)
- ▶ van der Veen et al. (2021)
- ▶ Blanchet et al. (2022)
- ▶ van der Veen (2022)
- ▶ van der Veen et al. (2023)
- ▶ Korhonen et al. (2024)
- ▶ van der Veen et al. (2024)
- ▶ Tang et al. (2025)

## Resources that cover all kinds of ordination methods

---

(none)

# Motivation

---

## Methods in Ecology and Evolution



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

Figure 1: Warton et al. 2015



# Motivation

---

- ▶ We need formal, probabilistic, models for community ecology
- ▶ That incorporate data properties, rather than transforming our way out of them
- ▶ It makes for better/clearer inference
- ▶ For nicer for teaching
- ▶ Access to tools for testing assumptions
- ▶ Overall more flexibility

## Why GLLVMs?

---

GLLVMs are a formal statistical, fully probabilistic, tool for multivariate analysis.

- 1) To step up your multivariate analysis
- 2) Maybe you want to incorporate random effects
- 3) Negates the need for distances

Figure 3: JSMD of spider data

## Use of ordination

---

- ▶ One of the most used methods of multivariate analysis
  - ▶ but definitely not the only one
- ▶ Arranges species and samples in a low-dimensional space
  - ▶ or any column and row quantity really
- ▶ Summarizes data
- ▶ Find underlying structures (gradients/LVs)

# 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](#).



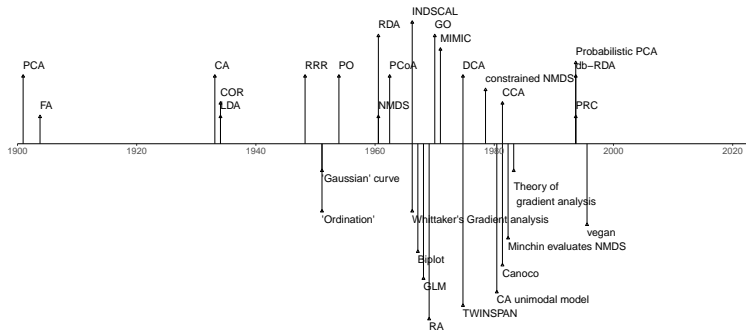
[To the ordination web page](#)

## Historical Perspective

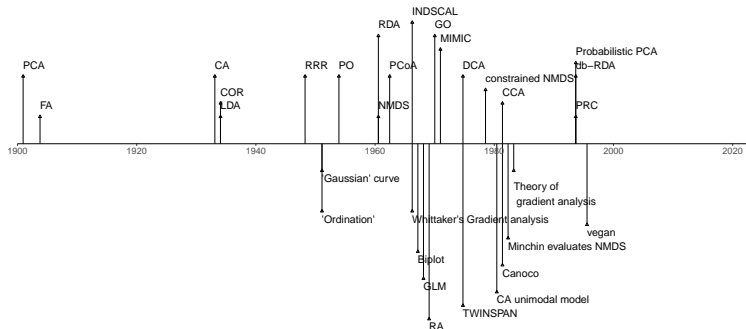
- 1901 - Pearson develops PCA as a regression technique.
- 1927 - Spearman applies factor analysis to psychology.
- 1930 - Ramensky uses an informal ordination technique and introduces the term 'ordnung' into ecology.
- 1954 - D.W. Goodall introduces PCA into ecology and proposes the term 'ordination'.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

# The rich history of multivariate analysis

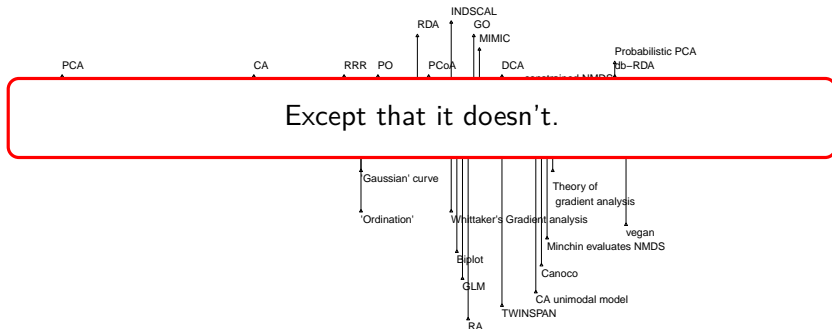


# The rich history of multivariate analysis



Seems like everything stops at the turn of the millennium?

# The rich history of multivariate analysis



Seems like everything stops at the turn of the millennium?



## Contemporary multivariate analysis

---

- ▶ Vector GLM(M)s
- ▶ Joint Species Distribution Models
- ▶ Model-based ordination

All three of these can be fitted with the `gllvm` R-package.

## Vector Generalised Linear (Mixed effects) Models

---

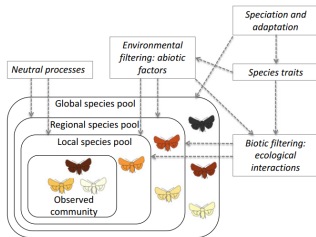
We can fit specialised mixed-effects models to community data:

$$g\{\mathbb{E}(y_{ij}|\boldsymbol{\lambda}_i)\} = \beta_{0j} + \mathbf{x}_i^\top \boldsymbol{\beta}_j + \mathbf{z}_i^\top \boldsymbol{\lambda}_j, \quad \boldsymbol{\lambda}_j \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (1)$$

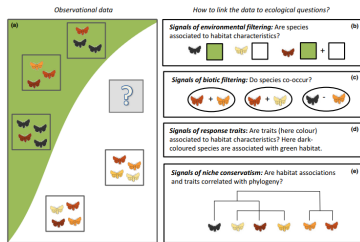
- ▶ The time that we were computationally limited is long past
- ▶ We can analyse multivariate data like any other
- ▶ Leverage information across all species in our data to learn about community processes

# Joint Species Distribution Modeling

- ▶ First suggested by Ovaskainen et al. 2010
- ▶ Named by Pollock et al. 2014



**Figure 1** A conceptual diagram of the assembly processes influencing ecological communities at different spatiotemporal scales. The composition and dynamics of local, regional and global communities are influenced by the combined effects of environmental filters, biotic interactions and neutral processes. The responses of the species to these factors depend on their traits, which are ultimately shaped by evolutionary history and therefore constrained by phylogenetic relationships.



**Figure 2** A conceptual illustration of some key questions in community ecology. The green and white colours represent differences in the environmental

Figure 4: Figures from Ovaskainen et al. 2017

## Joint Species Distribution Modeling (2)

- ▶ A niche process generates data
- ▶ Distributions are driven by 1) environment and 2) interactions
- ▶ Statistically, interaction = correlation of species
- ▶ So we model that:

$$g\{\mathbb{E}(y_{ij}|\epsilon_i)\} = \beta_{0j} + \mathbf{x}_i^\top \beta_j + \epsilon_{ij}, \quad \epsilon_i \sim \mathcal{N}(0, \Sigma) \quad (2)$$

Can be fit using standard mixed-effects modeling software:

```
glmer(abundance~species+x:species+(0+species|sites))
```

**Problem: Number of parameters grows quadratically**

## Model-based ordination

- ▶ Generalised Linear Latent Variable Model
- ▶ Adds “factor analytic” structure to  $\Sigma$
- ▶  $\epsilon_{ij} = \mathbf{u}_i^\top \boldsymbol{\gamma}_j$
- ▶ i.e.  $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \mathbf{\Gamma} \mathbf{\Gamma}^\top)$
- ▶ Faster and fewer parameters:
- ▶ Number of parameter doesn't grow so fast

$$\Sigma = \begin{bmatrix} \gamma_{11} & 0 & 0 \\ \gamma_{12} & \gamma_{22} & 0 \\ \vdots & \ddots & \vdots \\ \gamma_{1j} & \cdots & \gamma_{dj} \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1j} \\ 0 & \gamma_{22} & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_{dj} \end{bmatrix} \quad (3)$$

## Why model-based

---

- ▶ Flexibility
- ▶ Accommodate properties of data and sampling processes
- ▶ Now straightforward available and fast
- ▶ More advanced ecological inference and prediction

We can do all kinds of fun things

- ▶ Residual diagnostics
- ▶ Species-specific effects from constrained ordination
- ▶ Random-effects
- ▶ Flexible models
- ▶ Etc.

## Outline today

---

- ▶ Introduction
- ▶ Community data
- ▶ (Vector) GLMs
- ▶ Introduction to gllvm

