Hierarchical Ordination

An efficient (if not necessarily fast) fitting with INLA

R.B. O'Hara & B. van der Veen

What is Model Based Ordination?

The Data

Communities: S species at N sites

Data y_{ij} $(i = 1 \ldots n,\ j = 1 \ldots p)$

$$\begin{pmatrix} 0 & 5 & 13 \\ 2 & 7 & 4 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{pmatrix}$$

The Model I: the GLM bit

Assume from the exponential family

▶ normal, Poisson, binomial etc.

$$g(E(y_{ij})) = \eta_{ij}$$

▶ $g()$ is a link function

$$\eta_{ij} = \alpha_i + \beta_j + o_{ij}$$

We can make $\alpha_i + \beta_j$ more complex, but we put the complexity on o_{ij}

The Model

$$g(E(Y_{is})) = \alpha_i + \beta_j + o_{ij}$$

$$o_{ij} = \sum_{l=1}^{L} z_{il} \Sigma_{ll} \gamma_{jl}$$

With $\text{Var}(z_{il}) = \text{Var}(\gamma_{jl}) = 1$

So $\text{Var}(o_{ij}) = \sum_{l=1}^{L} \Sigma_{ll}$

And $\text{Corr}(o_{is}, o_{it}) = \sum_{l=1}^{L} \gamma_{sl} \gamma_{tl}$

Concurrent Ordination

What if we model the site effects, z_{il}?

$$z_{il} = \sum_{k=1}^{K} X_{ik}\phi_{kl} + \varepsilon_{il}$$

With $\varepsilon_{il} = 0$ this is the same as a constrained 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]

Hierarchical Ordination

If we can model z_{il}, why not also model γ_{jl}?

$$\gamma_{jl} = \sum_{s=1}^{S} X_{js}\omega_{sl} + \epsilon_{jl}$$

Trait effects, phylogeny etc etc

Fitting a Hierarchical Ordination

Difficult!

▶ MCMC (incl. HMC) seems to do poorly.

So, let's try something different

Spot the GLM

If we condition on z_{jl}, this is a GLM:

$$g(E(Y_{is})) = \alpha_i + \beta_j + \sum_{s=1}^{S} X_{js}\omega_{sl}z_i + \epsilon_{jl}z_i$$

If we condition on γ_{il}, this is a GLM:

$$g(E(Y_{is})) = \alpha_i + \beta_j + \sum_{k=1}^{K} X_{ik}\gamma_j\phi_{kl} + \gamma_j\varepsilon_{il}$$

We use z_{jl} and γ_j as weights

In details

We can use this as a multivariate Gibbs sampler

Use INLA, so we have all the flexibility in there

We can marginalise over the posteriors, not just draw from them.

▶ improve the estimates of z and γ

In details

Estimate z_{jl}

▶ Fit $g(E(Y_{is})) = \alpha_i + \beta_j + \sum_{k=1}^{K} X_{ik}\gamma_j\phi_{kl} + \gamma_j\varepsilon_{il}$

▶ draw $z_{jl}^* = z_{jl}\sqrt{\Sigma_l} = \sum_{k=1}^{K} X_{ik}\phi_{kl} + \varepsilon_{il}$ from their posterior

▶ standardise: $z_{jl} = z_{jl}^*/sd(z_{\cdot l}^*)$
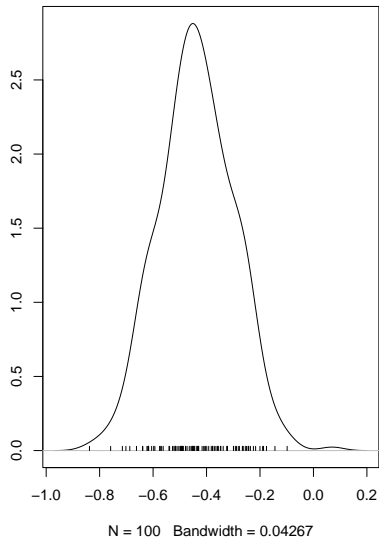
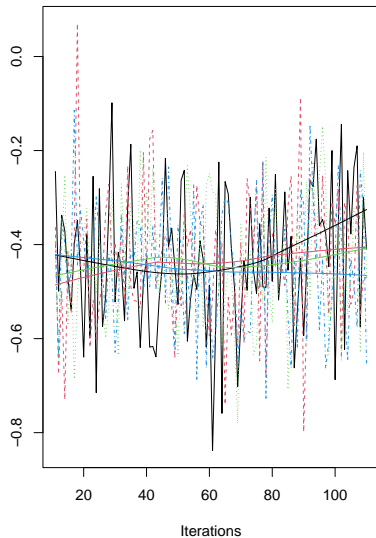Estimate γ_{il}

▶ as above

etc.

Example: Ant traits

From the `gllvm` package

- ▶ 41 species of ant
- ▶ 30 sites
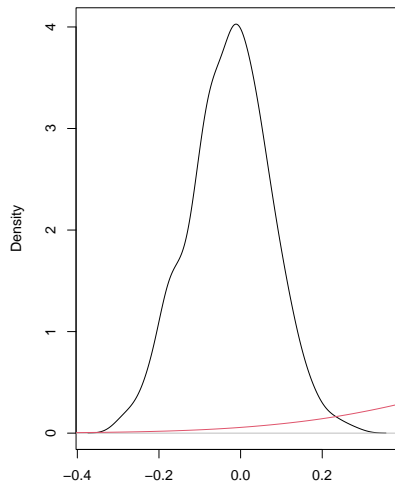- ▶ 5 environmental variables and
- ▶ 5 traits

Use 1 latent variable

Good Mixing

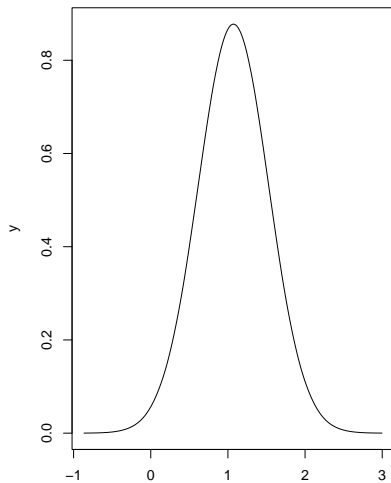
100 iterations, Camponotus.nigriceps
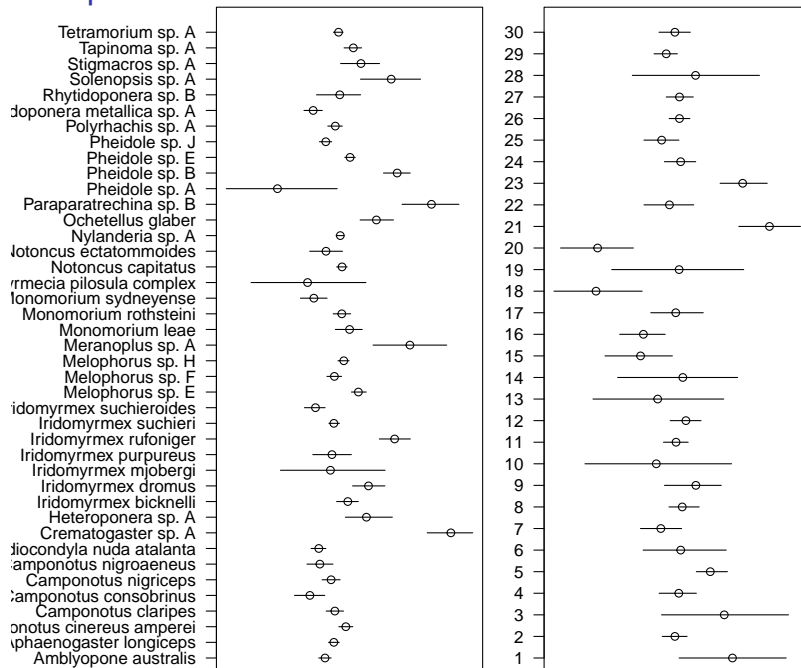
Good Density

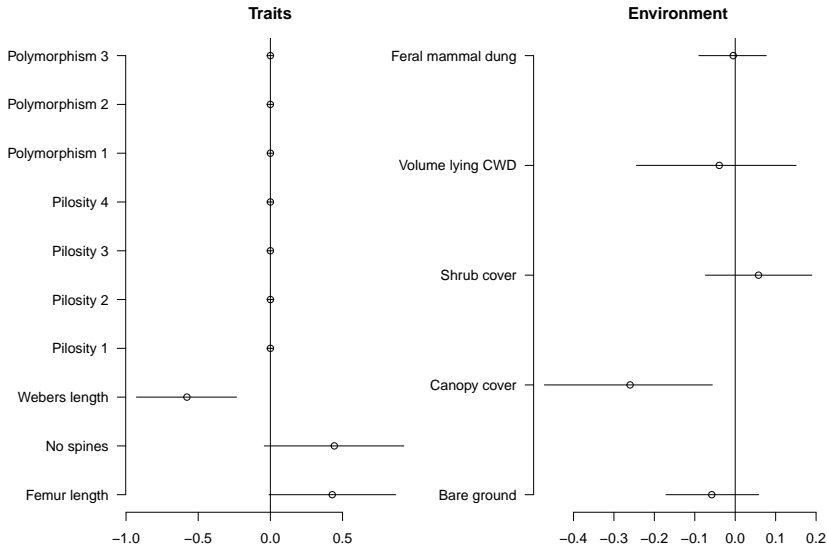**MCMC draws**   **INLA marginal**

N = 400   Bandwidth = 0.02617

Site & species scores

Covariate Effects

What is Weber's Length?

ML (Mesosomal or Weber's length). The diagonal length of the mesosoma in profile from the point at which the pronotum meets the cervical shield to the posterior basal angle of the metapleuron.

▶ Antwiki
(https://www.antwiki.org/wiki/Morphological_Measurements)

Abdomen length

The Current Status

How Well does it work?

I don't trust these results (yet)

Bugs need to be tracked and caught

>1 LV will need work to work out how best to rotate the LVs

Then play: Space, time, phylogeny

If you want to find bugs

(or laugh at my bad coding)

https://github.com/oharar/HierOrd