Hierarchical species response curves in package eHOF

Florian Jansen

November 30, 2013

Abstract

This is an example session to show how to use enhanced hierarchical logistic regression modeling in R with package *eHOF*. Only a few possibilities and applications can be covered. Use the usual help functions for further information or contact the author if in doubt. The package is still in it's early stage. After the installation of package eHOF you can access this PDF [1] with

```
vignette("eHOF")
```

1 Preparations

```
library(eHOF)

## Loading required package:  mgcv
## Loading required package:  nlme
## This is mgcv 1.7-27.  For overview type 'help("mgcv-package")'.
## Loading required package:  lattice
## This is eHOF 1.1
```

1.1 Load vegetation data

Hierarchical logistic regressions can be used in many fields. We will use vegetation data and we can use package `vegdata` [Jansen and Dengler, 2010] to load vegetation data from a Turboveg [Hennekens and Schaminée, 2001] database. Taxon names can be evaluated Jansen and Dengler [2010] if they are referenced with an appropriate taxonomic reference list [Jansen and Dengler, 2008]. Within this sample session we will use the build-in vegetation dataset of package `vegdata` from the floodplain of river Elbe [Leyer and Wesche, 2007] and a dataset delivered with `eHOF` from arable land of North-Eastern Germany with measured pH.

```
library(vegdata)

## Loading required package:  foreign
## This is vegdata 0.6.6

db <- "elbaue"
site <- tv.site(db)

##
## ########################################################
## Turboveg root directory is set to "/home/jansen/.wine/drive_c/Turbowin"
## If you want to change this use:  options(tv_home="<path_to_your_Turbowin_root>")
## ########################################################
```

---

[1] Id: eHOF.Rnw processed with eHOF 1.1 in R version 3.0.2 (2013-09-25) on November 30, 2013

```
veg <- tv.veg(db, tax = FALSE, spcnames = "numbers")
obs <- tv.obs(db)
# taxa <- tax(unique(obs$TaxonUsageID), verbose=TRUE) write.csv2(taxa,
# file='taxa.csv')
taxa <- read.delim("taxonnames.csv")
names(veg) <- sub(".0", "", names(veg), fixed = TRUE)
for (i in 1:ncol(veg)) names(veg)[i] <- as.character(taxa$LETTERCODE[match(as.numeric(names(veg)[i]),
    taxa$TaxonUsageID)])
```

Normally you will use the capabilities of package vegdata and the taxonomic reference list of your Turboveg installation to replace species numbers with Shortletters or real names. For the sake of CRAN policies we set tax=FALSE and replace taxnumbers manually.

1.2 Cover transformation

If you want to use cover values as performance, it might be better to use cover transformed abundance values instead of the original percentage cover.

```
veg.sqrt <- tv.veg(db, cover.transform = "sqrt", tax = FALSE, spcnames = "numbers")
names(veg.sqrt) <- sub(".0", "", names(veg.sqrt), fixed = TRUE)
names(veg.sqrt) <- taxa$LETTERCODE[match(names(veg.sqrt), taxa$TaxonUsageID)]
```

or even presence-absence information only.

```
veg.pa <- tv.veg(db, cover.transform = "pa", tax = FALSE, spcnames = "numbers")
names(veg.pa) <- sub(".0", "", names(veg.pa), fixed = TRUE)
names(veg.pa) <- taxa$LETTERCODE[match(names(veg.pa), taxa$TaxonUsageID)]
```

2 Modeling hierarchical logistic regressions

Introduced by Ramenskij in the early 20th century and named direct gradient analysis by Whittaker [1967] is modeling species responses along environmental gradients still a frequent task in vegetation ecology. Several algorithms are available to map the performance of a species along changing ecological conditions. Often it is difficult to decide which level of complexity is needed to get an adequate simplification of the data. In 1993 Huisman, Olff and Fresco introduced a set of hierarchical models to combine the wish for simple and easy to interpret response models with the need to catch different kinds of niche types and species response data [Huisman et al., 1993].

Species responses along measured environmental gradients are cross-sections of the species hyperniche. Even if we assume, that the physiological niche of a species should be simple (e.g. unimodal or with a specific threshold) we can not expect, that realized niches should be that simple too. Nevertheless, due to the generally high number of hidden gradients in field data and the resulting unbalanced datasets, it seems to be advisable not to use an untrammelled modelling technique like Generalised Additive Models (GAM) but to restrict ourselves to a more conservative set of model types which can be interpreted afterwards.

2.1 Modeltypes

Huisman, Olff and Fresco suggested five response shapes Huisman et al. [1993]. Additional to these we added two bimodal model types so that we get 7 hierarchical model types in total (see Fig. 1).

The seven models are of increasing complexity. Maybe the most important model type is number I: a flat response, that means there is no significant trend along the gradient for that species. It is our null hypothesis and ensures that only species with a clear response will be modelled with one of the further model types. Shape II is monotone sigmoid with a top at one end of the gradient, III is monotone sigmoid with a plateau below the maximal upper abundance value. Model type IV is the canonical form

```
data(acre)
sel <- c("ELYMREP", "VEROPES", "CONSREG", "DESUSOP", "VEROARV", "ARTE#VU", "ACHIMIL")
mo <- HOF(acre[match(sel, names(acre))], acre.env$PH_KCL, M = 1, bootstrap = NULL)
par(mar = c(4, 4, 1, 1) + 0.1)
autolayout(7)
par(mar = c(4, 4, 1, 1) + 0.1)
for (i in 1:7) plot(mo[[i]], model = eHOF.modelnames[i], marginal = "n")
```
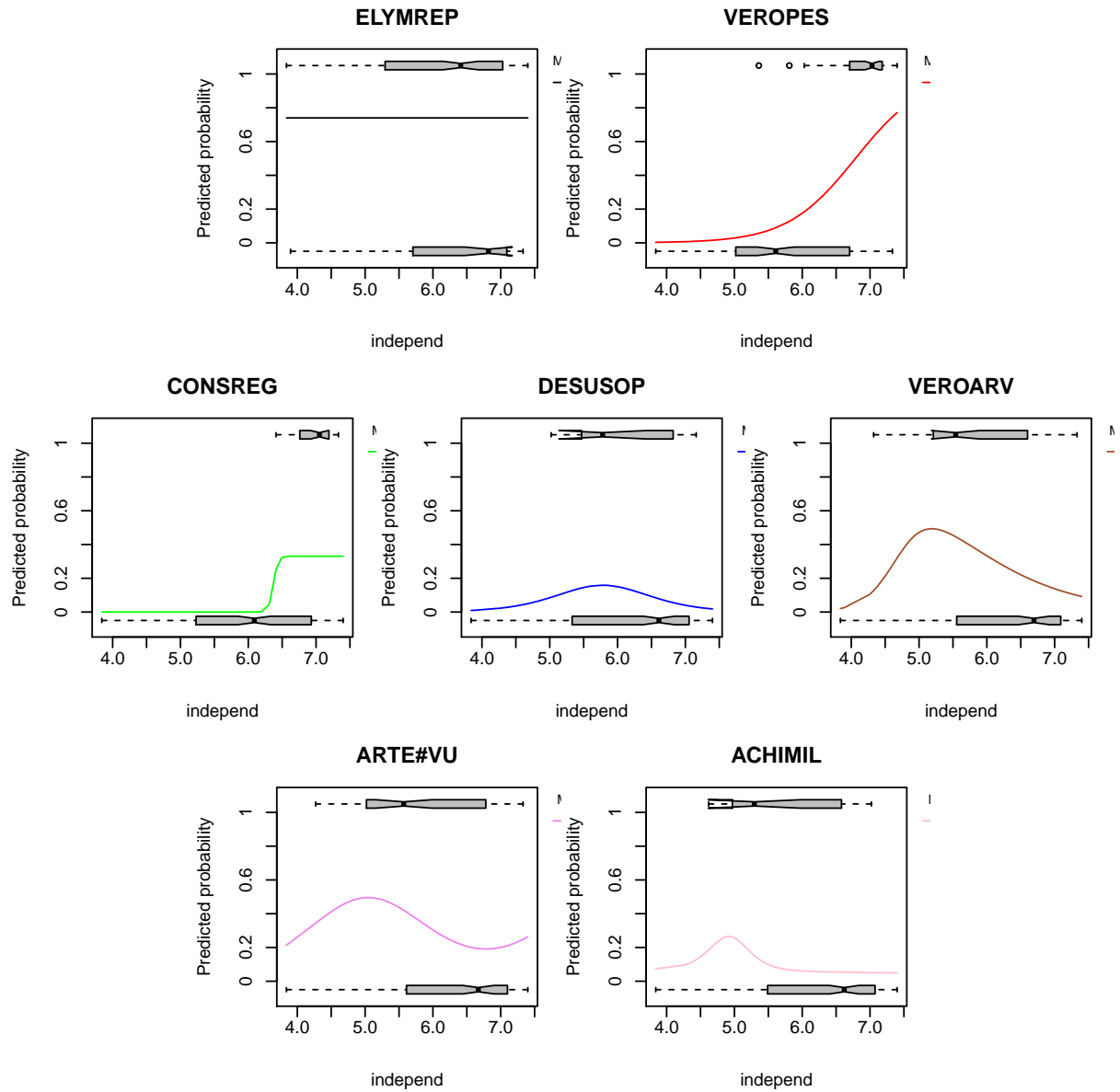

Figure 1: The layout plot shows the seven model types of enhanced Hierarchical logistic regression modelling.

of species response, a unimodal symmetric model, V is a unimodal skewed model and model types VI and VII have two optima, VI with tops being equal.

2.2 How to use function HOF?

For all species in a vegetation data frame above a specific frequency threshold (10 by default) all six model shapes are modelled and stored in a HOF object. The most appropriate model type is evaluated only at the moment this object is printed, summarised or plotted.

```
mods <- HOF(veg, site$MGL, M = 100, family = poisson, bootstrap = NULL)
```

```
mods

## Warning:  No bootstrap results in object.  Using selectMethod "raw" instead.

## Deviances:
##      STELPAU CAREACU CAREPRA CARDPRA GALUPAL POA TRI ELYMREP ALOPPRA
## I       29.8   498.3    44.7    42.6    52.4   85.8   736.3   453.6
## II      26.8   405.1    28.0    40.3    40.9   85.4   682.5   424.8
## III     21.7   308.0    19.1    37.3    32.7   73.6   376.4   276.3
## IV      15.8   218.7    22.7    27.9    30.2   61.5   250.3   209.2
## V       15.8   218.4    22.5    23.3    29.9   60.9   240.5   208.4
## VI      15.8   218.7    22.7    27.9    30.2   61.5   250.3   209.2
## VII     15.8   218.7    22.7    27.9    30.2   61.5   250.3   209.2
##      PHALARU POA PAL POA #PR RANCREP
## I      728.9   642.0   268.7   630.0 27.0
## II     709.3   641.5   247.3   603.4 26.3
## III    648.9   512.5   200.5   509.8 20.1
## IV     658.2   239.2   164.2   473.6 14.9
## V      658.1   167.6   160.9   435.9 14.8
## VI     658.2   239.2   164.2   473.6 14.9
## VII    658.2   239.2   164.2   473.6 14.9
##
## Suggested best models (AICc, raw):
## STELPAU CAREACU CAREPRA CARDPRA GALUPAL POA TRI ELYMREP ALOPPRA PHALARU
##      IV      IV     III       V      IV      IV       V      IV     III
## POA PAL POA #PR RANCREP    <NA>
##       V       V       V      IV
```

   Printing the output of a HOF objects with more than one species will give a matrix of deviances for all modelled species along all model types and thereafter the model type suggested as most appropriate by selected test criteria (Aikaike Information Criterion corrected for small sample size by default).

   Depending on the chosen performance measure HOF modelling will lead to different species response shapes.

3 Model parameters

Restricting models to predefined shapes offers the possibility to derive ecologically interpretable model characteristics like "the optimum", species niches etc. (see Fig. 4). Package HOF contains functions to compute parameters for the different model types.

3.1 Applications

Many applications in applied and theoretical ecology are possible with such modelling framework and with the derived model parameters.

```
mods.sq <- HOF(veg.sqrt, site$MGL, M = 10, family = poisson, freq.limit = 10,
    bootstrap = NULL)
plot(mods.sq)
```
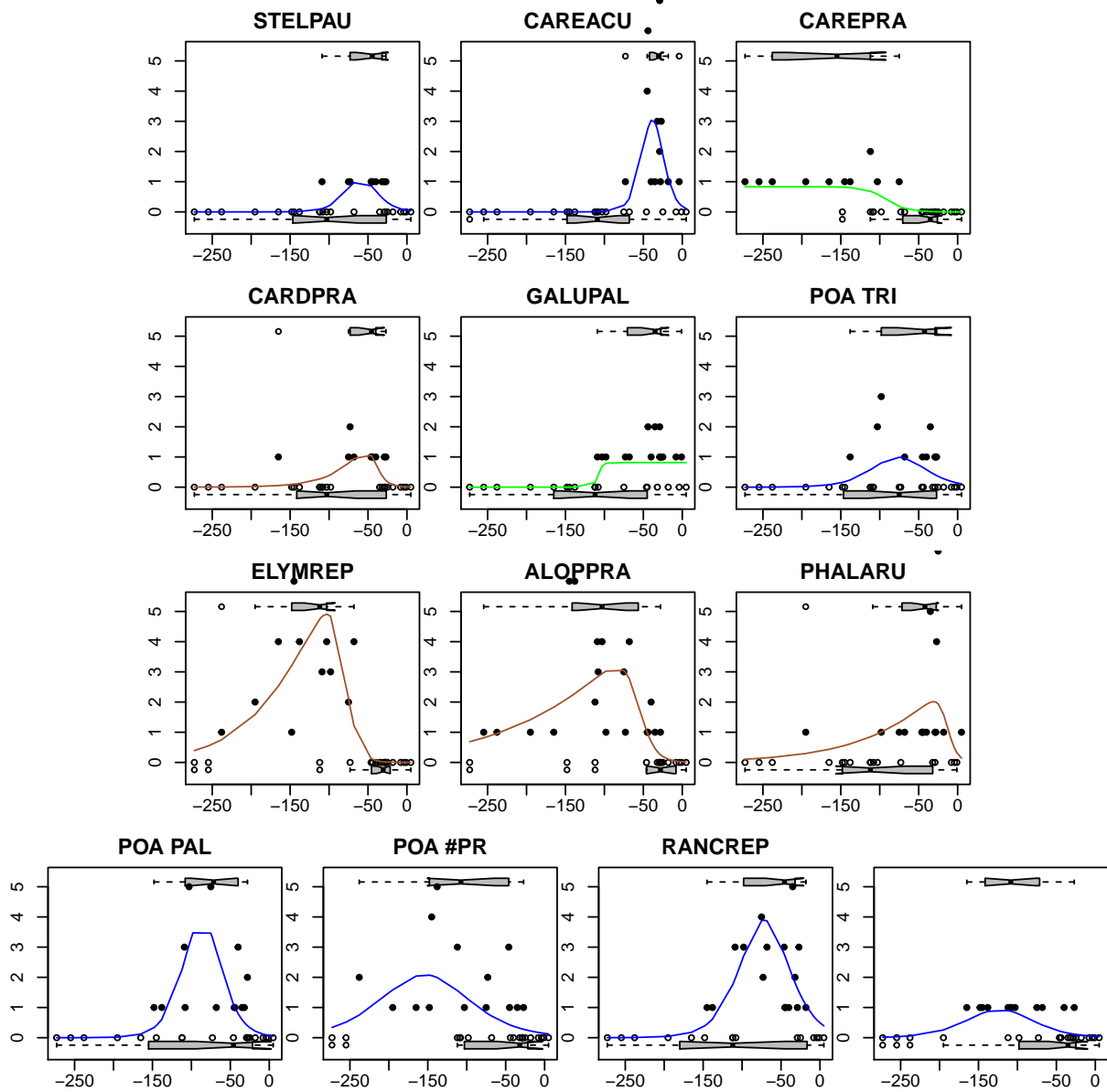

Figure 2: Most adequate model types for all species of the Elbaue dataset with at least 10 occurrences, square-root cover performance.

```
mods.pa <- HOF(veg.pa, site$MGL, M = 1, bootstrap = NULL)
plot(mods.pa)
```
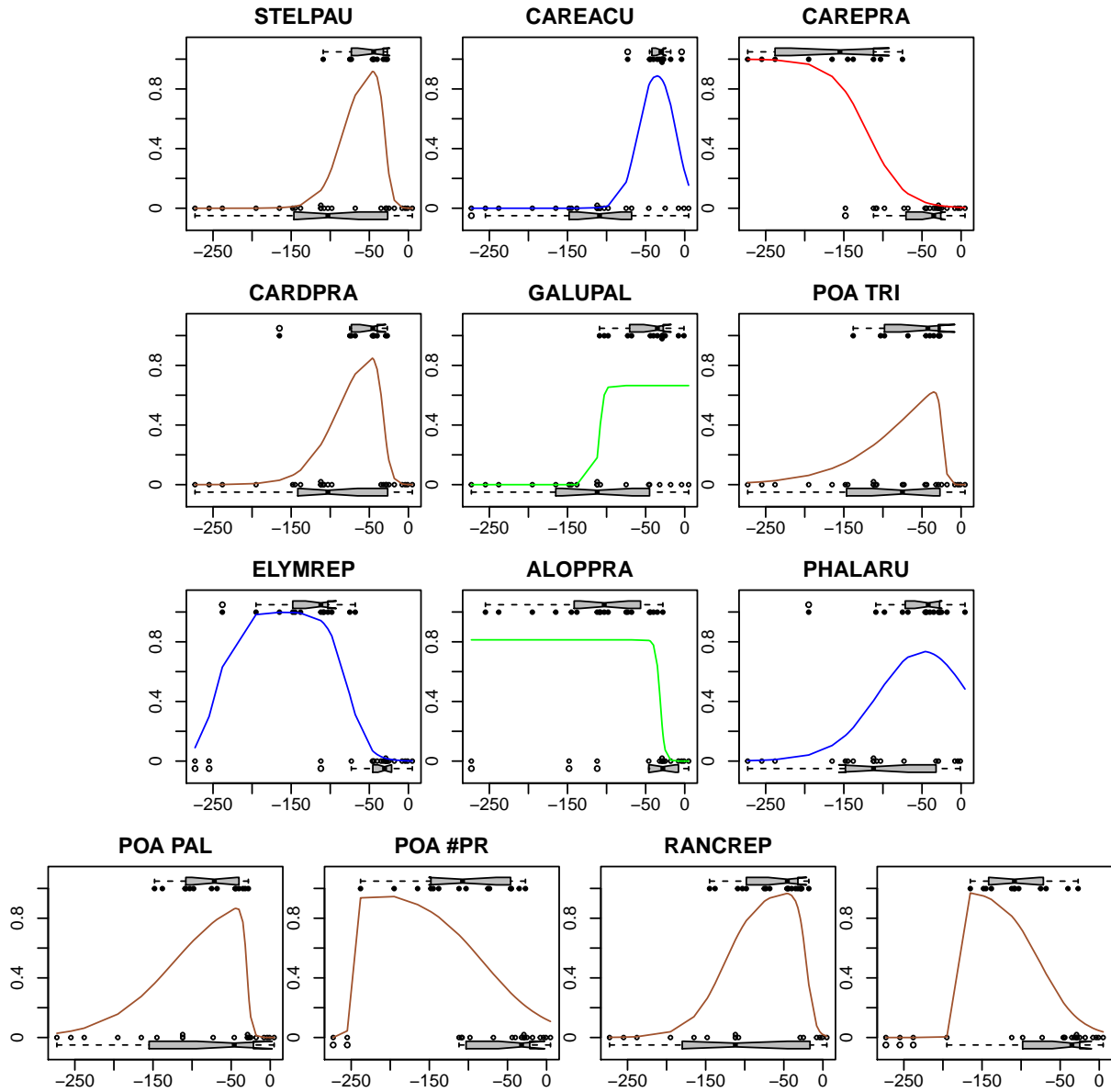

Figure 3: Most adequate model types for all species of the elbaue dataset with at least 10 occurrences, presence-absence information.

```
plot(mods.pa[["RANCREP"]], para = TRUE, onlybest = FALSE)

## Bootselect or IC.weight method only possible after bootstrapping. Choosing raw model inste
```
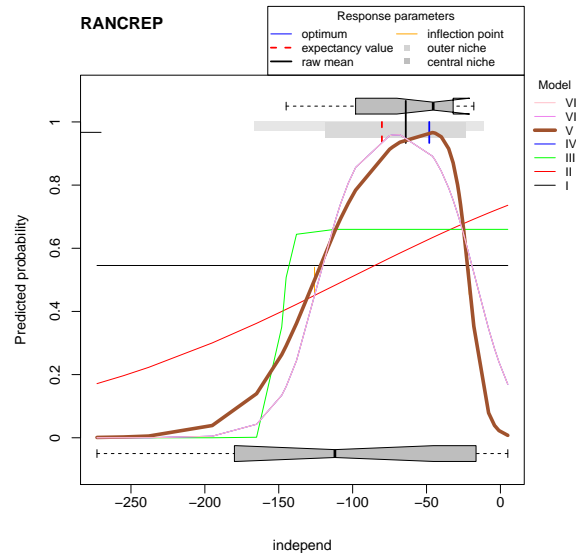

Figure 4: Hierarchical logistic regression models for *Ranunculus repens* presence/absence within the Elbaue dataset with plotted parameters of the most adequate model.

```
mod.ELYM.sqrt <- HOF(veg.sqrt$ELYMREP, site$MGL, M = 10, family = poisson, bootstrap = 10)
plot(mod.ELYM.sqrt, marginal = "point", para = TRUE, onlybest = FALSE, newdata = seq(min(mod.ELYM.sq
    max(mod.ELYM.sqrt$range), length.out = 10000))
```
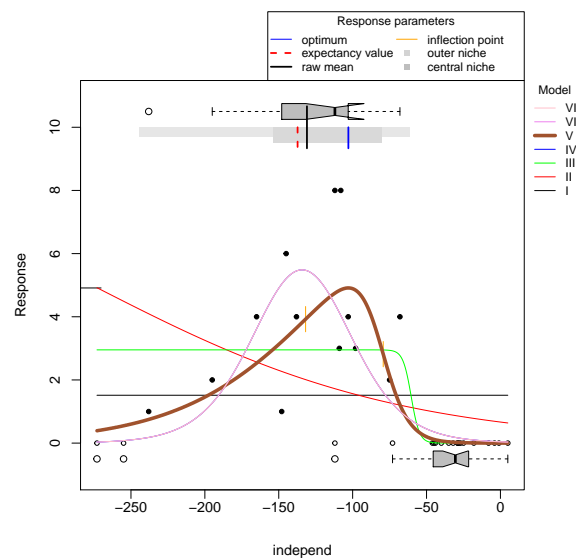

Figure 5: Hierarchical logistic regression models for *Elymus repens* square root cover within the Elbaue dataset with plotted parameters of the most adequate model.

Slopes of species response curves provide for instance an easy and direct way to model species turnover along an environmental gradient [Peper et al., 2011].

To think about the consequences of using only canonical response shapes (IV and maybe V) instead of a more complete set of model types you might want to analyse the optima of unimodal models (II,IV,V) compared to HOF modelling results.

References

Stephan M. Hennekens and Johannes Hendrikus Jacques Schaminée. Turboveg, a comprehensive data base management system for vegetation datasoftware package for input, processing, and presentation of phytosociological data. *Journal of Vegetation Science*, 12:589–591, 2001.

J Huisman, H Olff, and L F M Fresco. A Hierarchical Set of Models for Species Response Analysis. *Journal of Vegetation Science*, 4(1):37–46, 1993.

Florian Jansen and Jürgen Dengler. GermanSL - eine universelle taxonomische Referenzliste für Vegetationsdatenbanken. *Tuexenia*, 28:239–253, 2008.

Florian Jansen and Jürgen Dengler. Plant names in vegetation databases - a neglected source of bias. *Journal of Vegetation Science*, 21(6):1179–1186, Aug 2010. doi: 10.1111/j.1654-1103.2010.01209.x. URL `http://doi.wiley.com/10.1111/j.1654-1103.2010.01209.x`.

Ilona Leyer and Karsten Wesche. *Multivariate Statistik in der Ökologie*. Springer, Berlin, 2007.

Jan Peper, Florian Jansen, Dorothea Pietzsch, and Michael Manthey. Patterns of plant species turnover along grazing gradients. *Journal of Vegetation Science*, 22(3): 457–466, June 2011. ISSN 11009233. doi: 10.1111/j.1654-1103.2011.01260.x. URL `http://doi.wiley.com/10.1111/j.1654-1103.2011.01260.x`.

Robert Harding Whittaker. Gradient analysis of vegetation. *Biological Reviews*, 49:207–264, 1967.

```
mods.pa <- HOF(veg.pa, site$MGL, M = 1, freq.limit = 7, family = poisson, bootstrap = NULL)
p <- Para(mods.pa)
p4 <- vector("list", length(mods.pa))
for (i in 1:length(mods.pa)) p4[[i]] <- Para(mods.pa[[i]], modeltypes = "IV")
# p4
ind <- !sapply(p, function(x) x$model) %in% c("III", "VI", "VII")
p.opt <- sapply(p, function(x) x$opt)
p4.opt <- sapply(p4, function(x) x$opt)
plot(p.opt[ind], p4.opt[ind], xlab = "mean groundwater level", ylab = "only model IV")
for (i in 1:length(p.opt[!ind])) lines(unlist(p.opt[!ind][i]), rep(p4.opt[!ind][i],
    2), col = "darkgreen")
points(c(-270, -270), p4.opt[is.na(p.opt)], pch = "+")
```
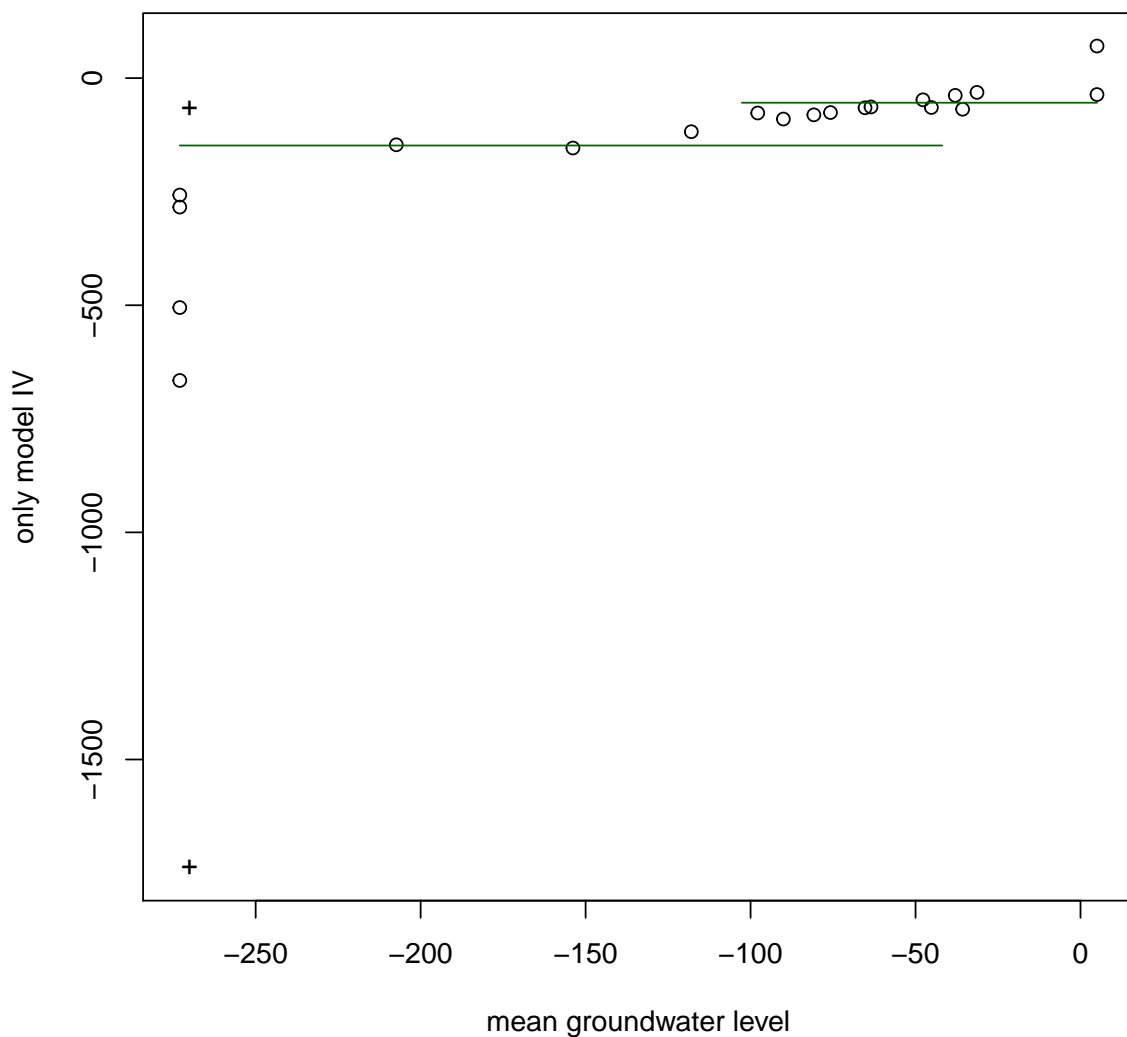

Figure 6: Scatter Plot of optima (mean groundwater level below surface in centimetre in the Elbaue dataset) with best HOF model versus response models restricted to be unimodal symmetric ('IV'). The two crosses are plotted at x=-270 but in fact they have no x values since they are recognized to be model I (no response) species by AICc selection. The green lines are representations of species without punctual optimum.