Hierarchical species response curves in package eHOF

Florian Jansen

April 15, 2015

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 1.8-6.  For overview type 'help("mgcv-package")'.
## Loading required package:  lattice
## This is eHOF 1.5.8
```

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

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] Id: eHOF.Rnw processed with eHOF 1.5.8 in R version 3.1.3 (2015-03-09) on April 15, 2015

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')


##
##  The following columns contain no data and are omitted:
##
##  The following numeric columns contain only 0 values and are omitted:
##
## The following numeric fields contain 0 values:
## Please check if these are really meant as 0 or if they are erroneously assigned because
of DBase restrictions.
## If so, use something like:
## site$Column_name[site$Column_name==0] <- NA

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')


##
##  The following columns contain no data and are omitted:
##
##  The following numeric columns contain only 0 values and are omitted:
##
## The following numeric fields contain 0 values:
## Please check if these are really meant as 0 or if they are erroneously assigned because
of DBase restrictions.
## If so, use something like:
## site$Column_name[site$Column_name==0] <- NA

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 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

## Bootstrap results missing.   Using selectMethod "pick.model" instead.

## Deviances:
##      STELPAU CAREACU CAREPRA CARDPRA GALUPAL POA TRI ELYMREP ALOPPRA
## I     29.824 498.281  44.733  42.574  52.425  85.751 736.321 453.553
## II    26.791 405.125  27.974  40.257  40.908  85.356 682.456 424.848
## III   21.720 308.033  19.055  37.312  32.680  73.589 376.393 276.331
## IV    15.838 218.694  22.687  27.862  30.152  61.504 250.278 209.237
## V     15.838 218.407  22.486  23.300  29.873  60.913 240.531 208.363
## VI    15.838 218.694  20.981  27.862  30.152  61.504 250.278 209.237
## VII   15.838 218.694  21.081  15.720  30.152  61.505 238.389 208.959
##      PHALARU POA PAL POA #PR RANCREP TARA/AN
## I    728.928 641.987 268.681 630.012  26.993
## II   709.313 641.482 247.349 603.353  26.300
## III  648.947 512.485 200.537 509.846  20.081
## IV   658.211 239.162 164.189 473.560  14.901
## V    658.096 167.636 160.915 435.924  14.841
## VI   658.211 239.162 164.189 473.560  14.901
## VII  658.211 239.162 164.189 473.560  14.906
##
## Suggested best models (AICc, pick.model):
## STELPAU CAREACU CAREPRA CARDPRA GALUPAL POA TRI ELYMREP ALOPPRA PHALARU
##      IV      IV     III     VII      IV      IV     VII      IV     III
## POA PAL POA #PR RANCREP TARA/AN
##       V       V       V      IV
```

Printing the output of a HOF object 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.

```
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)+.1)
autolayout(7)
par(mar=c(4,4,1,1)+.1)
for(i in 1:7) plot(mo[[i]], model = eHOF.modelnames[i], marginal ='n')
```
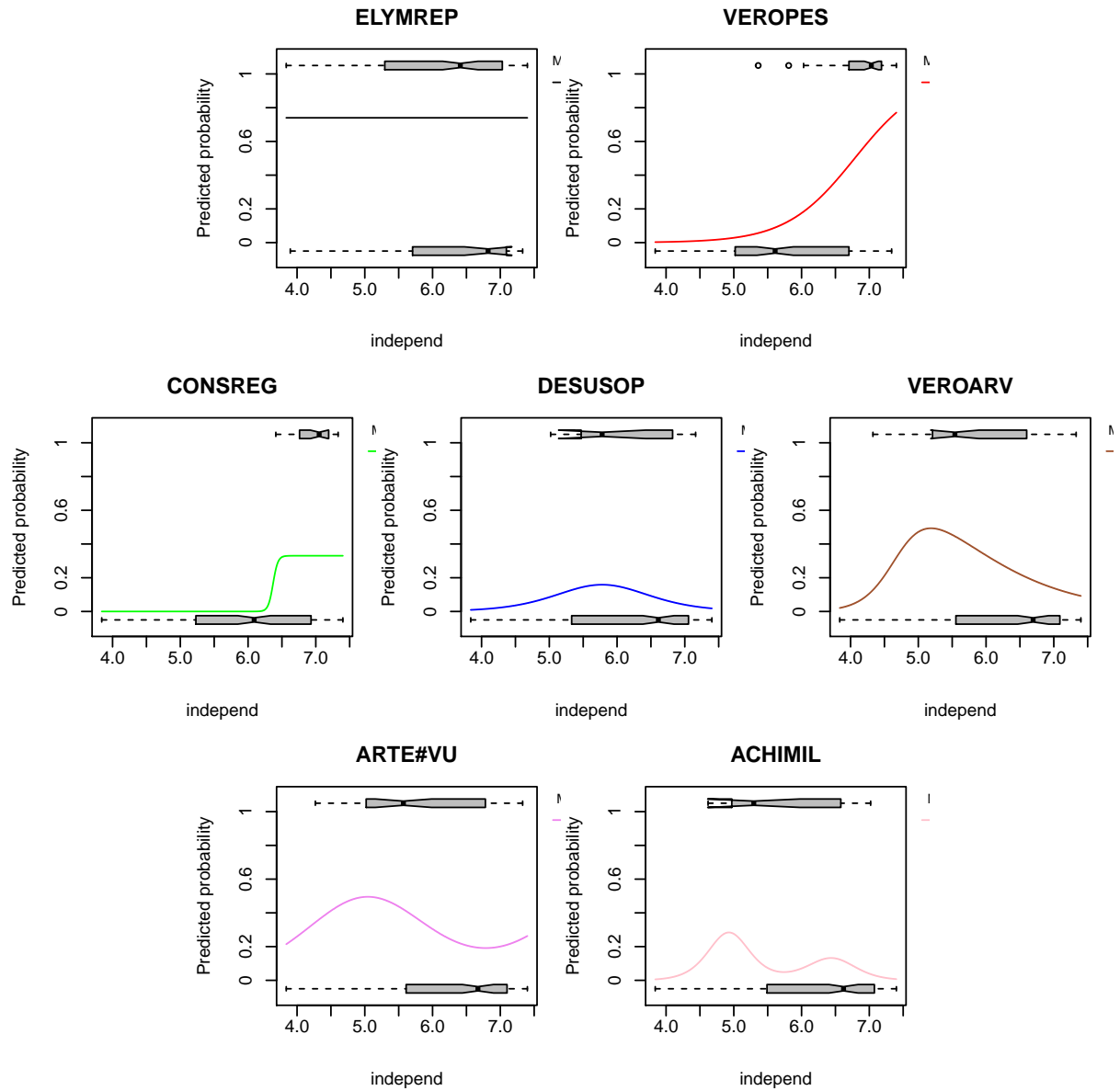

Figure 1: The layout plot shows the seven model types of enhanced Hierarchical logistic regression modelling.

```
mods.sq <- HOF(veg.sqrt, site$MGL, M=10, family= poisson, freq.limit=10, bootstrap=NULL)
plot(mods.sq)
            ## Bootselect or IC.weight method only possible after bootstrapping.
```
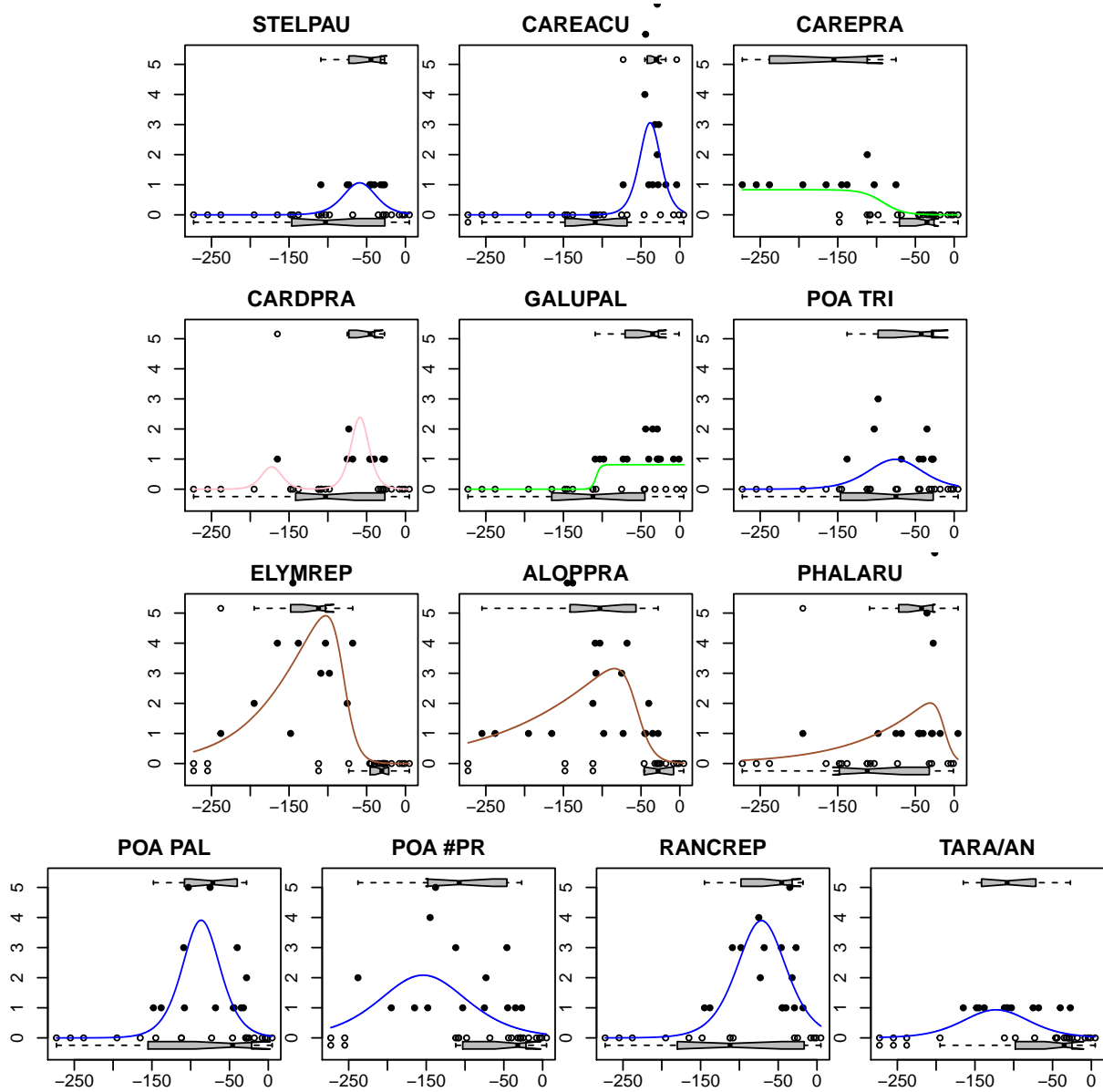

Figure 2: Most adequate model types for all species of the Elbaue dataset with at least 10 occurrences, square-root cover performance.

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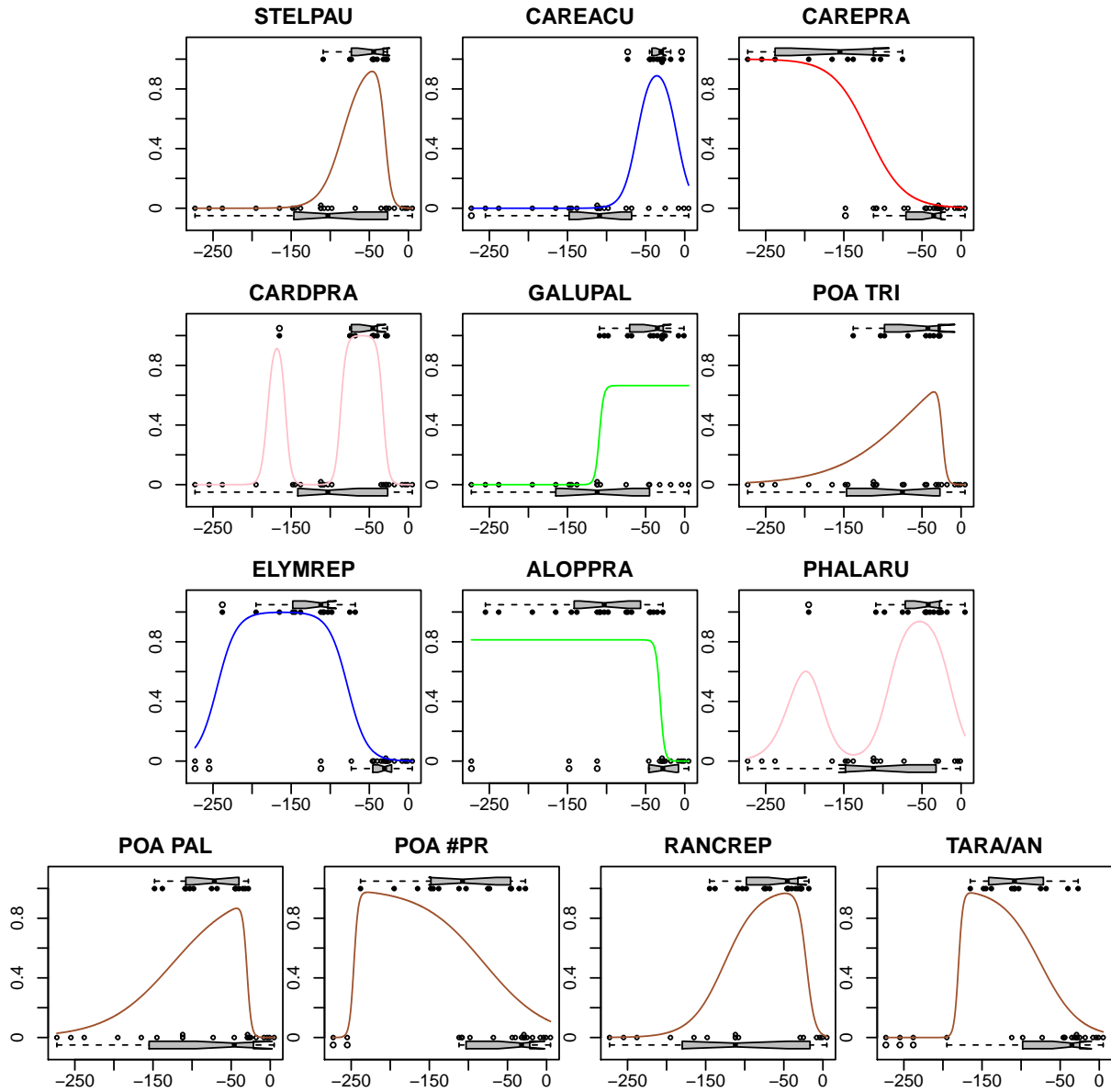

Figure 3: Most adequate model types for all species of the elbaue dataset with at least 10 occurrences, presence-absence information.

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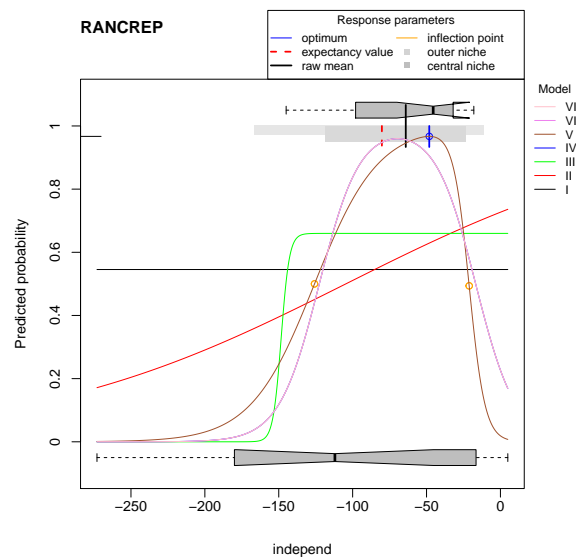

Figure 4: Hierarchical logistic regression models for *Ranunculus repens* presence/absence within the Elbaue dataset with plotted parameters of the most adequate model.
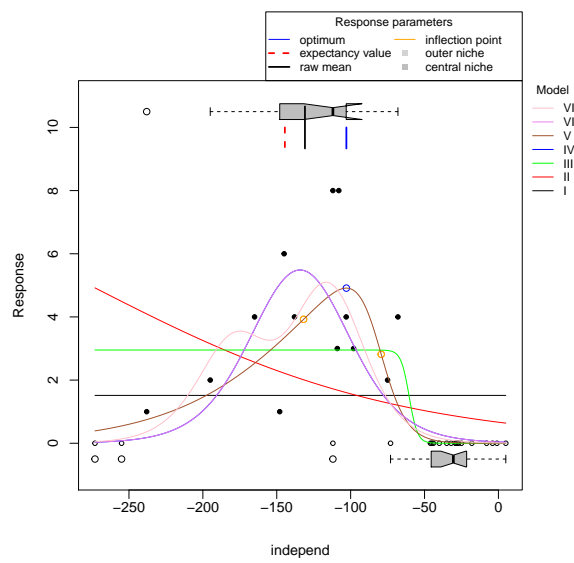
3 Model parameters

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

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.

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

```
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.sqrt$range
```



```
Para(mod.ELYM.sqrt)
```

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