

Hierarchical species response curves in package eHOF

Florian Jansen

January 10, 2022

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 1.8-38. For overview type 'help("mgcv-package")'.
## Loading required package: lattice
## This is eHOF 1.10
```

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

¹Id: eHOF.Rnw processed with eHOF 1.10 in R version 3.6.3 (2020-02-29) on January 10, 2022

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, Olf 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, Olf 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)
```

```

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')

```

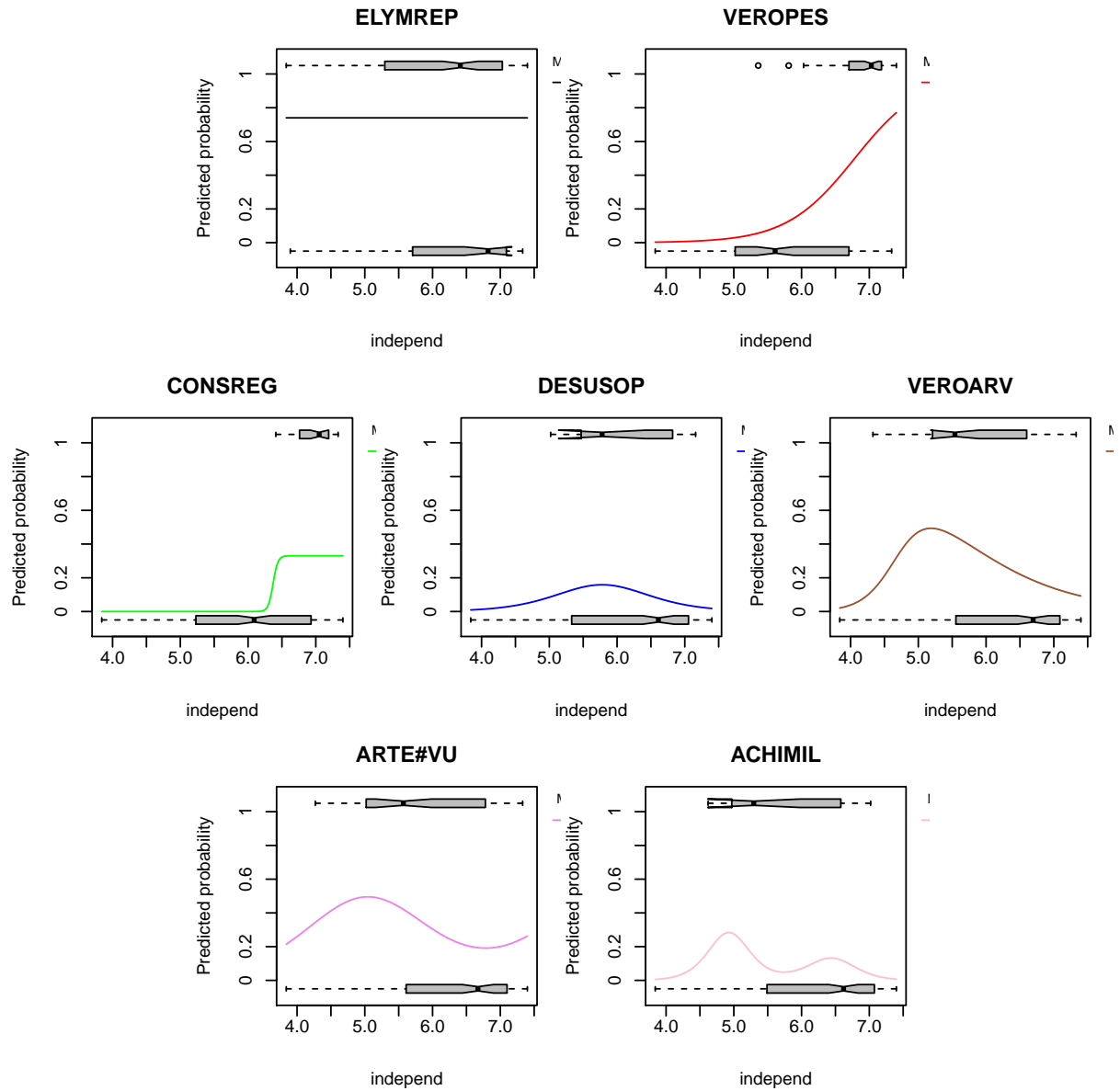


Figure 1: The layout plot shows the seven model types of enhanced Hierarchical logistic regression modelling.

```

mods

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

## Deviances:
##      11573      1230      14151      15      15133      2564      26611      336      4206
## I      29.824 498.281 44.733 736.321 42.574 52.425 85.751 453.553 728.928
## II     26.791 405.125 27.974 682.456 40.257 40.908 85.356 424.848 709.313
## III    21.720 308.033 19.055 376.393 37.312 32.680 73.589 276.331 648.947
## IV     15.838 218.694 22.687 250.278 27.862 30.152 61.504 209.237 658.211
## V      15.838 218.688 22.486 243.203 23.300 30.109 60.913 209.113 658.149
## VI     15.838 218.694 20.981 250.278 27.862 30.152 61.504 209.237 658.211
## VII    15.838 218.694 21.081 238.389 20.165 30.152 61.505 208.959 658.211
##      4364      4366      4768      7434
## I     641.987 268.681 630.012 26.993
## II     641.482 247.349 603.353 26.300
## III    512.485 204.852 509.846 20.081
## IV     239.364 164.189 473.560 14.901
## V      167.636 162.663 473.197 14.841
## VI     239.467 164.189 473.560 14.901
## VII    239.364 164.189 473.560 14.906
##
## Suggested best models (AICc, pick.model):
## 11573 1230 14151      15 15133 2564 26611      336 4206 4364 4366 4768 7434
##      IV      IV      III      VII      VII      IV      IV      IV      III      V      IV      IV      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.

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 data software 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.

```
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.

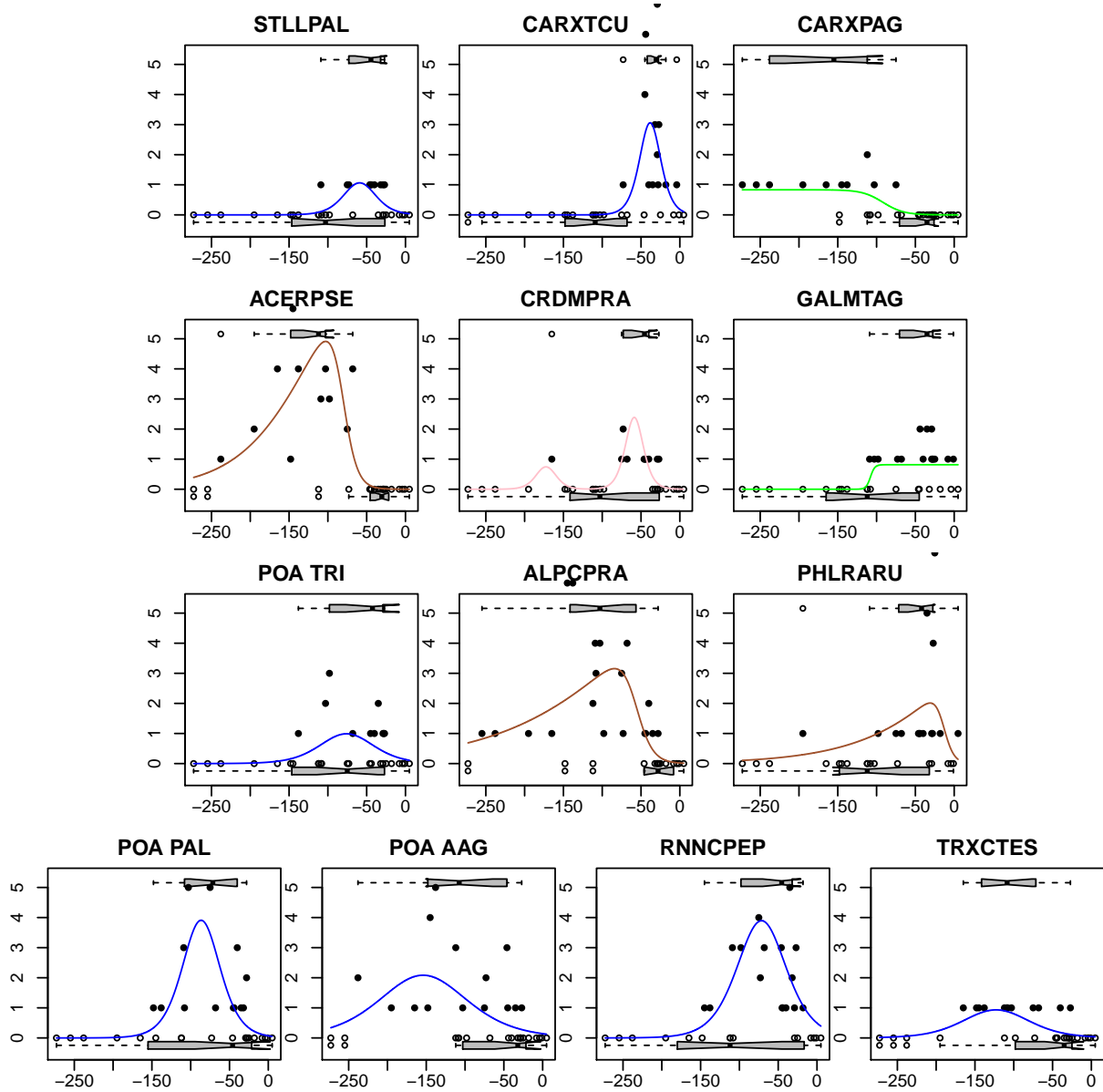


Figure 2: Most adequate model types for all species of the Elbaue dataset with at least 10 occurrences, square-root cover performance.

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

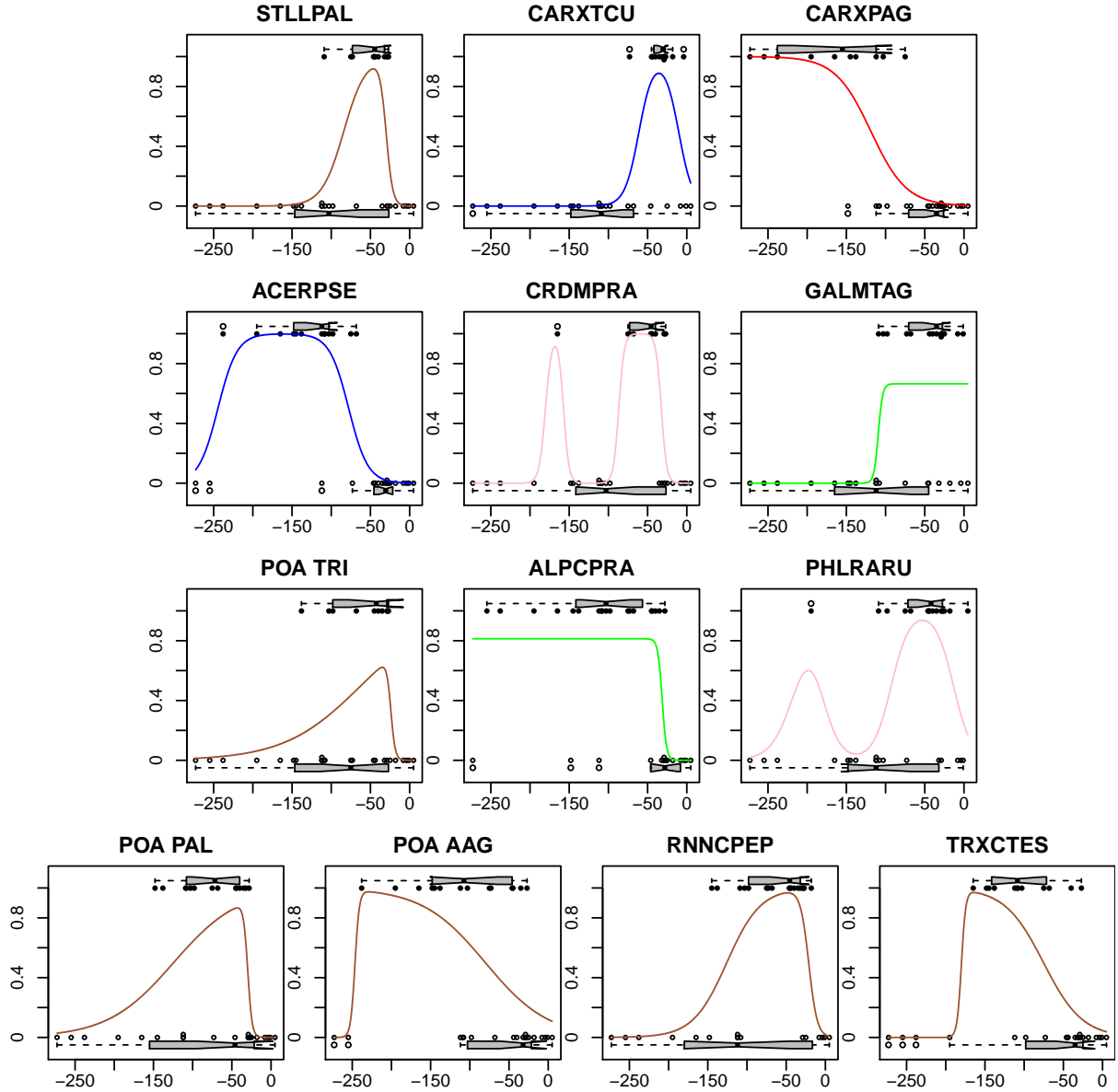


Figure 3: Most adequate model types for all species of the elbaue dataset with at least 10 occurrences, presence-absence information.

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

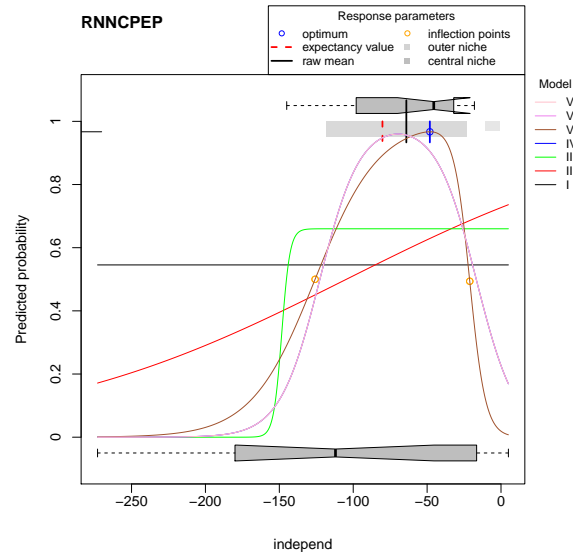
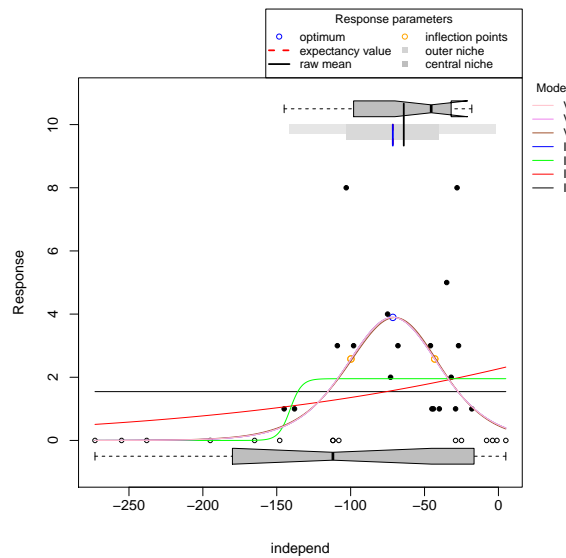


Figure 4: Hierarchical logistic regression models for *Ranunculus repens* presence/absence within the Elbaue dataset with plotted parameters of the most adequate model.

```
mod.sqrt <- HOF(veg.sqrt$RNNCPEP, site$MGL, M=10, family=poisson, bootstrap = 10)
plot(mod.sqrt, marginal='point', para=TRUE, onlybest=FALSE, newdata=seq(min(mod.sqrt$range), max(mod
```



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

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