

# Species richness model

Daniela Linero

23/6/2020

```
# Drop sites that don't have richness measures or land-use data
diversity_all <- drop_na(diversity_all,
                           Species_richness,
                           Predominant_land_use) %>%
  droplevels()

# Call the function that merges land-uses and intensities
source("./R/02_Statistical_Analysis_merge_LandUses_Intensities.R")
```

```
# Create the vectors that hold the land-uses that we want to keep with
# different use intensities
land_uses_separate <- c("Primary", "Cropland",
                        "ISV", "Plantation forest")

# Create a vector with the land-uses where we want to merge the light
# and intense use intensities
land_uses_light_intense <- c("Cropland", "ISV", "Plantation forest")

diversity_all <-
  Merge_landUses_and_intensities(dataset = diversity_all,
                                   index = 1,
                                   land_uses_separate_intensities = land_uses_separate,
                                   land_uses_merge_light_intense = land_uses_light_intense,
                                   reference = "Primary Minimal use")

# Check number of sites
table(diversity_all$LandUse.1, diversity_all$Kingdom)
```

```
##                                     Animalia Plantae nePlantae
## Primary Minimal use                  451    480    399
## Cropland Light-intense use           87     60     60
## Cropland Minimal use                 62     34     34
## ISV Light-intense use                68     88     87
## ISV Minimal use                     97     75     71
## MSV All                            44     82     35
## Pasture All                         177    407    407
## Plantation forest Light-intense use 296    389    318
## Plantation forest Minimal use        33     232    232
## Primary Intense use                  47     102    102
```

```

## Primary Light use           195      114      113
## YSV All                   155      149      148

```

1. After filtering for studies that evaluated more than 1 species, I have enough sites to model:

- Primary can be divided into the three level intensities in all cases
- Cropland can be divided in minimal use and light/intense use in animal, endo plant and non endo plant
- ISV can be divided in minimal use and light/intense use in animal, endo plant and non endo plant
- MSV has to be merged
- Pasture has to be merged
- Plantation forest can be divided in minimal use and light/intense use in animal, endo plant and non endo plant
- YSV has to be merged

```

# Get the function
source("https://highstat.com/Books/Book2/HighstatLibV10.R")

# Calculate the VIF
corvif(diversity_all[ , c("LandUse.1", "Kingdom")])

```

```

##
##
## Variance inflation factors
##
##          GVIF Df GVIF^(1/2Df)
## LandUse.1 1.063321 11     1.002795
## Kingdom    1.063321  2     1.015468

```

2. Choose random effect structure

```

m1 <- glmer(Species_richness ~ LandUse.1 + Kingdom + LandUse.1:Kingdom +
             (1|SS) + (1|SSB), data = diversity_all, family = poisson,
             # I'm increasing the number of iterations as the model
             # initially didn't converge
             control = glmerControl(optimizer = "bobyqa",
                                     optCtrl = list(maxfun = 20000)))

```

I will test for overdispersion using Ben Bolker's function

```

overdisp_fun <- function(model) {
  rdf <- df.residual(model)
  rp <- residuals(model, type="pearson")
  Pearson.chisq <- sum(rp^2)
  prat <- Pearson.chisq/rdf
  pval <- pchisq(Pearson.chisq, df=rdf, lower.tail=FALSE)
  c(chisq=Pearson.chisq, ratio=prat, rdf=rdf, p=pval)
}

overdisp_fun(m1)

```

```
##      chisq      ratio      rdf      p
## 7.203923e+03 1.222662e+00 5.892000e+03 6.903134e-30
```

Add an observation level random effect to account for overdispersion

```
m1_q <- glmer(Species_richness ~ LandUse.1 + Kingdom +
  LandUse.1:Kingdom +
  (1|SS) + (1|SSB) + (1|SSBS), data = diversity_all,
  family = poisson,
  control = glmerControl(optimizer = "bobyqa",
    optCtrl = list(maxfun = 20000)))
```

```
m2_q <- glmer(Species_richness ~ LandUse.1 + Kingdom +
  LandUse.1:Kingdom +
  (1|Source_ID) + (1|SS) + (1|SSB) + (1|SSBS),
  data = diversity_all, family = poisson,
  control = glmerControl(optimizer = "bobyqa",
    optCtrl = list(maxfun = 20000)))
```

Compare the models that converged

```
AIC(m1_q, m2_q)
```

```
##      df      AIC
## m1_q 39 18542.57
## m2_q 40 18511.30
```

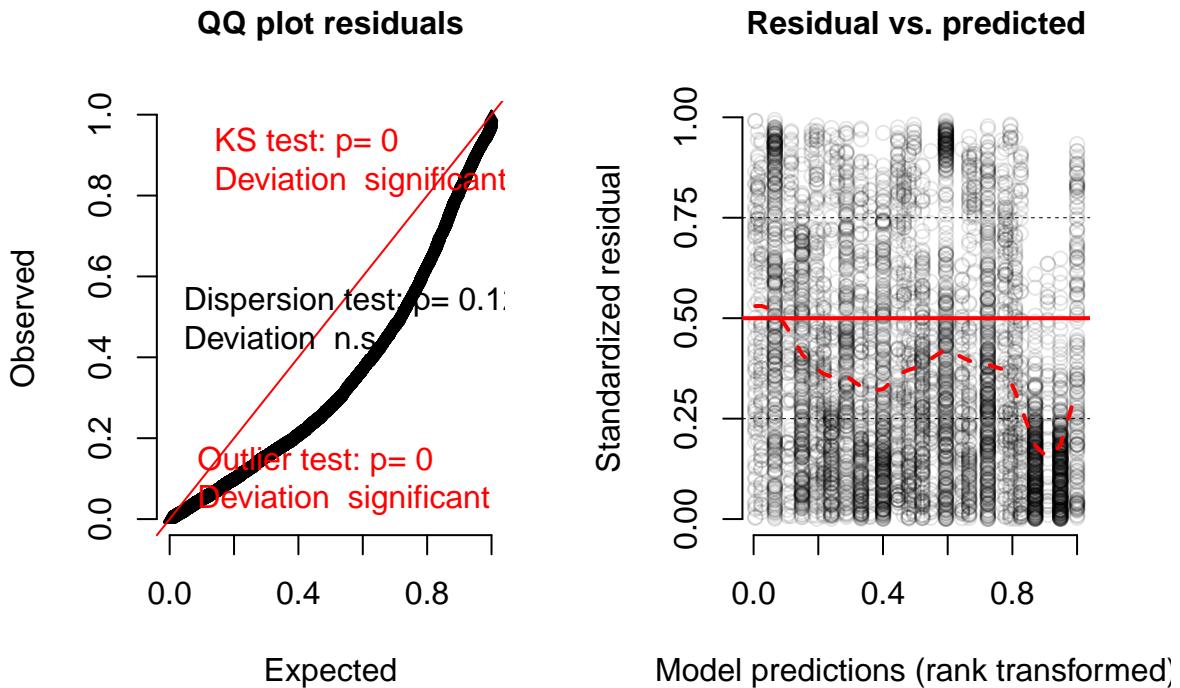
```
overdisp_fun(m2_q)
```

```
##      chisq      ratio      rdf      p
## 5583.7978938 0.9480132 5890.0000000 0.9979180
```

Plot the residuals

```
simulationOutputq <- simulateResiduals(fittedModel = m2_q, plot = TRUE)
```

## DHARMA residual diagnostics



3. See what combination of land-use and land-use intensities is better

The difference between the richness and the abundance model, is that in the abundance model cropland has all the use intensities merged (because cropland has 19, 0 and 60 sites for minimal, light and intense uses in both endozoochorous plants and plants not dispersed by animals)

First I am going to assess the models were cropland is divided into minimal and light/intense uses

For Richness:

```
#AIC of the richness model with different intensities for cropland
AIC(m2_q)
```

```
## [1] 18511.3
```

For abundance:

```
# Import table that has the site metrics calculated for the
# abundance model (that means including animals and both types of plants)

diversity_all_abundance <- readRDS("./output/cleaned_data/02_Statistical_Analysis_Abundance_Site_metrics")

# This function creates new columns with the land-uses and use
# intensities that we want, then it then merges both columns
# into a LanUse column
diversity_all_abundance <- Merge_landUses_and_intensities(diversity_all_abundance,
1,
```

```

        land_uses_separate,

# Transform rescaled abundance
diversity_all_abundance <- mutate(diversity_all_abundance,
                                    logAbundance = log(RescaledAbundance + 1),
                                    sqrtAbundance = sqrt(RescaledAbundance)
)

# Model abundance with those land-use classes
m2_a <- lmer(logAbundance ~ LandUse.1 + Kingdom +
              LandUse.1:Kingdom +
              (1|SS) + (1|SSB) + (1|Source_ID),
              data = diversity_all_abundance)
AIC(m2_a)

## [1] -3443.747

```

I will now assess the models were cropland is not divided into different intensities

For Richness:

```

# Create the vectors that hold the land-uses that we want to keep with different use intensities
land_uses_separate2 <- c("Primary", "ISV", "Plantation forest")
# Create a vector with the land-uses that we want to merge
land_uses_light_intense2 <- c("ISV", "Plantation forest")

# This function creates new columns with the land-uses and use
# intensities that we want, then it then merges both columns
# into a LanUse column
diversity_all <- Merge_landUses_and_intensities(diversity_all,
                                                2,
                                                land_uses_separate2,
                                                land_uses_light_intense2,
                                                "Primary Minimal use")

```

```

# Model richness with those land-use classes
m2_q2 <- glmer(Species_richness ~ LandUse.2 + Kingdom +
                  LandUse.2:Kingdom +
                  (1|Source_ID) + (1|SS) + (1|SSB) + (1|SSBS),
                  data = diversity_all,
                  family = poisson,
                  control = glmerControl(optimizer = "bobyqa",
                                         optCtrl = list(maxfun = 20000)))

AIC(m2_q2)

```

```
## [1] 18893.27
```

For abundance:

```

diversity_all_abundance <-
  Merge_landUses_and_intensities(diversity_all_abundance,
    2,
    land_uses_separate2,
    land_uses_light_intense2,
    "Primary Minimal use")

# Model abundance with those land-use classes
m2_a2 <- lmer(logAbundance ~ LandUse.2 + Kingdom + LandUse.2:Kingdom +
  (1|SS) + (1|SSB) + (1|Source_ID),
  data = diversity_all_abundance)

AIC(m2_a2)

```

## [1] -3538.588

Compare AIC

```

# Compare AICs

# For the first combination
AIC(m2_q) + AIC(m2_a)

```

## [1] 15067.55

```

# For the second combination
AIC(m2_q2) + AIC(m2_a2)

```

## [1] 15354.68

According to these results I should keep cropland with different land-use intensities

4. Compare with model that has all the land-use intensities merged

```

# Complete cases: get the rows that have information for the levels
# that we choose in the more complex model
diversity_all <- drop_na(diversity_all,
  Species_richness, LandUse.1) %>% droplevels()

# Merge all the intensity levels for all land-use categories

# Create the vectors that hold the land-uses that we want to keep
# with different use intensities
land_uses_separate3 <- "NA"
# Create a vector with the land-uses that we want to merge
land_uses_light_intense3 <- "NA"

# This function creates new columns with the land-uses
# and use intensities that we want, then it then merges
# both columns into a LandUse column

```

```


diversity_all <- Merge_landUses_and_intensities(diversity_all,
  3,
  land_uses_separate3,
  land_uses_light_intense3,
  "Primary All")



# Model richness with those land-use classes



m2_q_3 <- glmer(Species_richness ~ LandUse.3 + Kingdom +
  LandUse.3:Kingdom +
  (1|Source_ID) + (1|SS) + (1|SSB) + (1|SSBS),
  data = diversity_all, family = poisson,
  control = glmerControl(optimizer = "bobyqa",
  optCtrl = list(maxfun = 20000)))


```

```
anova(m2_q, m2_q_3)
```

```

## Data: diversity_all
## Models:
## m2_q_3: Species_richness ~ LandUse.3 + Kingdom + LandUse.3:Kingdom +
## m2_q_3: (1 | Source_ID) + (1 | SS) + (1 | SSB) + (1 | SSBS)
## m2_q: Species_richness ~ LandUse.1 + Kingdom + LandUse.1:Kingdom +
## m2_q: (1 | Source_ID) + (1 | SS) + (1 | SSB) + (1 | SSBS)
##     npar   AIC   BIC  logLik deviance Chisq Df Pr(>Chisq)
## m2_q_3  25 18799 18966 -9374.4    18749
## m2_q     40 18511 18779 -9215.6    18431 317.53 15 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

### 5. Plant contrast

Create null model

```
# For the null model I am going to replace the nePlantae for Plantae
diversity_all <- diversity_all %>%
```

```


mutate(
  # Create a new kingdom column as the copy of the kingdom column
  # we used in the first model
  Kingdom.1 = paste(Kingdom),



# Replace nePlantae for Plantae
Kingdom.1 = recode_factor(Kingdom.1, "nePlantae" = "Plantae"),



# set reference level
Kingdom.1 = factor(Kingdom.1),
Kingdom.1 = relevel(Kingdom.1, ref = "Animalia"))


```

```
m2_q_4 <- glmer(Species_richness ~ LandUse.1 + Kingdom.1 +
  LandUse.1:Kingdom.1 +
  (1|Source_ID) + (1|SS) + (1|SSB) + (1|SSBS),
  data = diversity_all, family = poisson,
```

```

control = glmerControl(optimizer = "bobyqa",
                      optCtrl = list(maxfun = 20000)))

anova(m2_q_4, m2_q, test = "F")

## Data: diversity_all
## Models:
## m2_q_4: Species_richness ~ LandUse.1 + Kingdom.1 + LandUse.1:Kingdom.1 +
## m2_q_4:      (1 | Source_ID) + (1 | SS) + (1 | SSB) + (1 | SSBS)
## m2_q: Species_richness ~ LandUse.1 + Kingdom + LandUse.1:Kingdom +
## m2_q:      (1 | Source_ID) + (1 | SS) + (1 | SSB) + (1 | SSBS)
##      npar   AIC   BIC  logLik deviance Chisq Df Pr(>Chisq)
## m2_q_4  28 20100 20288 -10022.2    20044
## m2_q     40 18511 18779  -9215.6    18431 1613.1 12 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

##### 5. Results of the selected model

Estimates

```

m2_q_estimates <- read.csv("./output/summary_m2_q.csv", header = TRUE, sep = ";")
kable(m2_q_estimates, format="latex", booktabs=TRUE) %>%
  kable_styling(latex_options="scale_down")

```

X	Estimate	std.error	statistic	p.value	Actual.estimate
(Intercept)	1.16	0.15	7.66	0.000	1.16
Cropland Light-intense use	-0.29	0.10	-2.84	0.004	0.86
Cropland Minimal use	-0.22	0.09	-2.48	0.013	0.93
ISV Light-intense use	-0.09	0.13	-0.72	0.472	1.06
ISV Minimal use	0.03	0.08	0.31	0.760	1.18
MSV All	0.01	0.14	0.06	0.953	1.16
Pasture All	-0.57	0.14	-4.02	0.000	0.58
Plantation forest Light-intense use	-0.38	0.07	-5.49	0.000	0.78
Plantation forest Minimal use	-0.04	0.10	-0.40	0.692	1.12
Primary Intense use	0.23	0.13	1.75	0.081	1.39
Primary Light use	0.41	0.07	6.18	0.000	1.56
YSV All	-0.05	0.08	-0.57	0.569	1.11
KingdomPlantae	-0.07	0.20	-0.35	0.724	1.08
KingdomnePlantae	-1.16	0.21	-5.65	0.000	0.00
Cropland Light-intense use:KingdomPlantae	-0.78	0.19	-4.20	0.000	0.01
Cropland Minimal use:KingdomPlantae	-0.26	0.15	-1.69	0.092	0.60
ISV Light-intense use:KingdomPlantae	0.22	0.15	1.41	0.159	1.21
ISV Minimal use:KingdomPlantae	-0.07	0.12	-0.62	0.532	1.04
MSV All:KingdomPlantae	0.11	0.17	0.65	0.513	1.20
Pasture All:KingdomPlantae	0.87	0.19	4.58	0.000	1.38
Plantation forest Light-intense use:KingdomPlantae	0.18	0.10	1.82	0.069	0.89
Plantation forest Minimal use:KingdomPlantae	-0.20	0.13	-1.48	0.139	0.85
Primary Intense use:KingdomPlantae	-0.22	0.19	-1.16	0.246	1.10
Primary Light use:KingdomPlantae	-0.21	0.11	-1.87	0.061	1.28
YSV All:KingdomPlantae	-0.15	0.11	-1.30	0.194	0.89
Cropland Light-intense use:KingdomnePlantae	-1.45	0.35	-4.10	0.000	-1.74
Cropland Minimal use:KingdomnePlantae	0.66	0.16	4.14	0.000	0.43
ISV Light-intense use:KingdomnePlantae	-0.15	0.19	-0.77	0.443	-0.24
ISV Minimal use:KingdomnePlantae	0.03	0.14	0.20	0.842	0.05
MSV All:KingdomnePlantae	-0.38	0.20	-1.90	0.058	-0.38
Pasture All:KingdomnePlantae	2.13	0.19	11.17	0.000	1.55
Plantation forest Light-intense use:KingdomnePlantae	0.63	0.11	5.57	0.000	0.25
Plantation forest Minimal use:KingdomnePlantae	1.10	0.13	8.25	0.000	1.06
Primary Intense use:KingdomnePlantae	0.29	0.20	1.41	0.160	0.52
Primary Light use:KingdomnePlantae	0.63	0.12	5.19	0.000	1.03
YSV All:KingdomnePlantae	0.51	0.12	4.22	0.000	0.46

Plots

```
# Plot the differences between estimates
PlotErrBar_interactions_modi(model = m2_q,
                               resp = "Species richness",
                               Effect1 = "LandUse.1",
                               Effect2 = "Kingdom",
                               ylims = c(-3,2.5),
                               pointtype = c(16,17, 18),
                               blackwhite = FALSE)

# Plot the x label
text(x = c(0.7:12), y = -3, labels = c("Primary M",
                                         "Cropland LI",
                                         "Cropland M",
                                         "ISV LI",
```

```

"ISV M",
"MSV A",
"Pasture A",
"Plantation LI",
"Plantation M",
"Primary I",
"Primary L",
"YSV A"), srt = 18, cex= 0.7)

```

