output2-sr-2

Yidi

2020/8/19

## Load the packages

Load some packages for manipulating and modelling the data

library(maps)  
library(devtools)  
library(predictsFunctions)  
library(StatisticalModels)  
library(raster)  
library(dplyr)  
library(tidyr)  
library(lme4)  
library(car)  
library(DHARMa)  
library(MuMIn)  
library(Hmisc)

## Read in and process the diversity data

diversity <- readRDS("/Users/dd/Desktop/PREDICTS data/database.rds")  
diversity <- mutate(diversity,   
 Measurement = Effort\_corrected\_measurement,  
 Sampling\_effort = Rescaled\_sampling\_effort)

An optional step to merge any sites that are within the same land-use type and that have identical coordinates, start and end dates.

diversity <- MergeSites(diversity, silent = TRUE)

## Calculate the diversity metrics

sites <- diversity %>%  
   
 # add Diversity\_metric\_is\_valid column  
 mutate(Diversity\_metric\_is\_valid = TRUE) %>%  
   
 # calculate SiteMetrics   
 SiteMetrics(extra.cols = c("SSB", "SSBS", "Predominant\_land\_use"))

## Computing site metrics for 2906994 measurements  
## The data contain 480 sources, 666 studies and 22678 sites  
## Computing site-level values  
## Computing total abundance  
## Computing species richness  
## Assembling site-level values

sites <- sites %>%  
   
 mutate(  
   
 # collapse primary forest and non-forest together into primary vegetation as these aren't well distinguished  
 Predominant\_land\_use = recode\_factor(Predominant\_land\_use,   
 "Primary forest" = "Primary vegetation",   
 "Primary non-forest" = "Primary vegetation"),  
   
 # indeterminate secondary veg and cannot decide get NA  
 Predominant\_land\_use = na\_if(Predominant\_land\_use, "Secondary vegetation (indeterminate age)"),  
 Predominant\_land\_use = na\_if(Predominant\_land\_use, "Cannot decide"),  
   
 # set reference levels  
 Predominant\_land\_use = factor(Predominant\_land\_use),  
 Predominant\_land\_use = relevel(Predominant\_land\_use, ref = "Primary vegetation")  
 )

## Model site-level diversity

Step 1: complete cases

model\_data <- drop\_na(sites, Species\_richness, Predominant\_land\_use)

Step 2: starting/maximal model

model\_data <- mutate(model\_data,   
 logSpecies\_richness = log(Species\_richness + 1),  
 sqrtSpecies\_richness = sqrt(Species\_richness))  
m3 <- glmer(Species\_richness ~ Predominant\_land\_use   
 + (1|SS) + (1|SSB) + (1|SSBS), data = model\_data, family = poisson)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.00341117 (tol = 0.002, component 1)

summary(m3)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: poisson ( log )  
## Formula: Species\_richness ~ Predominant\_land\_use + (1 | SS) + (1 | SSB) +   
## (1 | SSBS)  
## Data: model\_data  
##   
## AIC BIC logLik deviance df.resid   
## 119126.3 119213.5 -59552.2 119104.3 20424   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.4796 -0.4380 -0.0128 0.3308 5.8389   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## SSBS (Intercept) 0.08515 0.2918   
## SSB (Intercept) 0.04306 0.2075   
## SS (Intercept) 1.49492 1.2227   
## Number of obs: 20435, groups: SSBS, 20435; SSB, 2182; SS, 647  
##   
## Fixed effects:  
## Estimate Std. Error  
## (Intercept) 2.61836 0.04942  
## Predominant\_land\_useYoung secondary vegetation -0.16034 0.01703  
## Predominant\_land\_useIntermediate secondary vegetation -0.15706 0.01638  
## Predominant\_land\_useMature secondary vegetation -0.01303 0.02170  
## Predominant\_land\_usePlantation forest -0.26639 0.01560  
## Predominant\_land\_usePasture -0.18713 0.01369  
## Predominant\_land\_useCropland -0.27403 0.01618  
## Predominant\_land\_useUrban -0.28894 0.03325  
## z value Pr(>|z|)   
## (Intercept) 52.981 <2e-16 \*\*\*  
## Predominant\_land\_useYoung secondary vegetation -9.417 <2e-16 \*\*\*  
## Predominant\_land\_useIntermediate secondary vegetation -9.591 <2e-16 \*\*\*  
## Predominant\_land\_useMature secondary vegetation -0.600 0.548   
## Predominant\_land\_usePlantation forest -17.077 <2e-16 \*\*\*  
## Predominant\_land\_usePasture -13.666 <2e-16 \*\*\*  
## Predominant\_land\_useCropland -16.940 <2e-16 \*\*\*  
## Predominant\_land\_useUrban -8.689 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) P\_\_Ysv P\_\_Isv P\_\_Msv Pr\_\_Pf Prd\_\_P Prd\_\_C  
## Prdmnn\_\_Ysv -0.061   
## Prdmnn\_\_Isv -0.069 0.259   
## Prdmnn\_\_Msv -0.063 0.183 0.235   
## Prdmnnt\_\_Pf -0.069 0.142 0.202 0.141   
## Prdmnnt\_l\_P -0.074 0.212 0.260 0.172 0.164   
## Prdmnnt\_l\_C -0.077 0.194 0.222 0.154 0.207 0.334   
## Prdmnnt\_l\_U -0.052 0.067 0.089 0.108 0.073 0.082 0.095  
## convergence code: 0  
## Model failed to converge with max|grad| = 0.00341117 (tol = 0.002, component 1)

# considering the overdispersion  
m4 <- glmer(Species\_richness ~ Predominant\_land\_use + (1|SS)  
 + (1|SSBS), data = model\_data, family = poisson)

# have a look at the significance of the terms  
Anova(m4)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: Species\_richness  
## Chisq Df Pr(>Chisq)   
## Predominant\_land\_use 642.62 7 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(m4)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: poisson ( log )  
## Formula: Species\_richness ~ Predominant\_land\_use + (1 | SS) + (1 | SSBS)  
## Data: model\_data  
##   
## AIC BIC logLik deviance df.resid   
## 119567.7 119647.0 -59773.9 119547.7 20425   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.3635 -0.4434 -0.0125 0.3324 5.6945   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## SSBS (Intercept) 0.09582 0.3096   
## SS (Intercept) 1.52616 1.2354   
## Number of obs: 20435, groups: SSBS, 20435; SS, 647  
##   
## Fixed effects:  
## Estimate Std. Error  
## (Intercept) 2.62226 0.04939  
## Predominant\_land\_useYoung secondary vegetation -0.16082 0.01676  
## Predominant\_land\_useIntermediate secondary vegetation -0.16124 0.01640  
## Predominant\_land\_useMature secondary vegetation -0.03611 0.02112  
## Predominant\_land\_usePlantation forest -0.27673 0.01548  
## Predominant\_land\_usePasture -0.20177 0.01369  
## Predominant\_land\_useCropland -0.27490 0.01629  
## Predominant\_land\_useUrban -0.25199 0.03054  
## z value Pr(>|z|)   
## (Intercept) 53.096 <2e-16 \*\*\*  
## Predominant\_land\_useYoung secondary vegetation -9.595 <2e-16 \*\*\*  
## Predominant\_land\_useIntermediate secondary vegetation -9.831 <2e-16 \*\*\*  
## Predominant\_land\_useMature secondary vegetation -1.710 0.0873 .   
## Predominant\_land\_usePlantation forest -17.874 <2e-16 \*\*\*  
## Predominant\_land\_usePasture -14.743 <2e-16 \*\*\*  
## Predominant\_land\_useCropland -16.875 <2e-16 \*\*\*  
## Predominant\_land\_useUrban -8.251 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) P\_\_Ysv P\_\_Isv P\_\_Msv Pr\_\_Pf Prd\_\_P Prd\_\_C  
## Prdmnn\_\_Ysv -0.061   
## Prdmnn\_\_Isv -0.067 0.262   
## Prdmnn\_\_Msv -0.062 0.200 0.232   
## Prdmnnt\_\_Pf -0.067 0.137 0.191 0.134   
## Prdmnnt\_l\_P -0.072 0.208 0.250 0.182 0.154   
## Prdmnnt\_l\_C -0.076 0.188 0.215 0.157 0.198 0.327   
## Prdmnnt\_l\_U -0.050 0.065 0.086 0.100 0.079 0.085 0.106

PlotGLMERFactor(model = m4,  
 data = model\_data,  
 responseVar = "Species\_richnes",  
 xtext.srt = 20,  
 seMultiplier = 1.96,  
 logLink="n",  
 catEffects = c("Predominant\_land\_use"))

