How does land management affect biodiversity

Yidi

2020/7/20

## 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")) %>%  
   
 # calculate the total abundance within each study  
 group\_by(SS) %>%  
 mutate(MaxAbundance = ifelse(Diversity\_metric\_type == "Abundance",  
 max(Total\_abundance),  
 NA)) %>%  
 ungroup() %>%  
   
 # now calculate the rescaled abundance (abundance divided by the maximum within each study)  
 mutate(RescaledAbundance = ifelse(Diversity\_metric\_type == "Abundance",  
 Total\_abundance/MaxAbundance,  
 NA))

## 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, Total\_abundance, Predominant\_land\_use)

Step 2: starting/maximal model

transforming RescaledAbundance

model\_data <- mutate(model\_data,   
 logAbundance = log(RescaledAbundance + 1),  
 sqrtAbundance = sqrt(RescaledAbundance))  
m1 <- lmer(sqrtAbundance ~ Predominant\_land\_use + (1|SS) + (1|SSB) + (1|Source\_ID), data = model\_data)

Step 3: Choose the random effects

m2 <- lmer(sqrtAbundance ~ Predominant\_land\_use + (1+Predominant\_land\_use|SS) + (1|SSB) + (1|Source\_ID), data = model\_data)

## boundary (singular) fit: see ?isSingular

# compare the models using Akaike's Information Criterion (AIC)  
AIC(m1,m2)

## df AIC  
## m1 12 -7093.844  
## m2 47 -8805.076

Step 4: Choose the best fixed effects structure

# have a look at the significance of the terms  
Anova(m2)

## Analysis of Deviance Table (Type II Wald chisquare tests)  
##   
## Response: sqrtAbundance  
## Chisq Df Pr(>Chisq)   
## Predominant\_land\_use 28.899 7 0.000151 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# look at the model estimates of our mimumum adequate model (which in our case is also our maximal model)  
summary(m2)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: sqrtAbundance ~ Predominant\_land\_use + (1 + Predominant\_land\_use |   
## SS) + (1 | SSB) + (1 | Source\_ID)  
## Data: model\_data  
##   
## REML criterion at convergence: -8899.1  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.0646 -0.5573 -0.0372 0.5230 5.3941   
##   
## Random effects:  
## Groups Name Variance  
## SSB (Intercept) 0.006576  
## SS (Intercept) 0.012828  
## Predominant\_land\_useYoung secondary vegetation 0.020425  
## Predominant\_land\_useIntermediate secondary vegetation 0.014057  
## Predominant\_land\_useMature secondary vegetation 0.006110  
## Predominant\_land\_usePlantation forest 0.034634  
## Predominant\_land\_usePasture 0.034772  
## Predominant\_land\_useCropland 0.056313  
## Predominant\_land\_useUrban 0.021112  
## Source\_ID (Intercept) 0.021078  
## Residual 0.029339  
## Std.Dev. Corr   
## 0.08109   
## 0.11326   
## 0.14291 -0.53   
## 0.11856 -0.28 0.67   
## 0.07816 -0.23 0.60 0.38   
## 0.18610 -0.74 0.49 0.38 0.49   
## 0.18647 -0.50 0.78 0.58 0.33 0.18   
## 0.23730 -0.76 0.83 0.54 0.50 0.84 0.62   
## 0.14530 -0.57 0.48 0.87 0.17 0.57 0.47 0.58  
## 0.14518   
## 0.17129   
## Number of obs: 17593, groups: SSB, 1937; SS, 575; Source\_ID, 424  
##   
## Fixed effects:  
## Estimate Std. Error  
## (Intercept) 0.662301 0.011288  
## Predominant\_land\_useYoung secondary vegetation -0.019496 0.014988  
## Predominant\_land\_useIntermediate secondary vegetation -0.007078 0.014632  
## Predominant\_land\_useMature secondary vegetation 0.020541 0.015555  
## Predominant\_land\_usePlantation forest -0.043130 0.016715  
## Predominant\_land\_usePasture -0.051397 0.017056  
## Predominant\_land\_useCropland -0.075356 0.019964  
## Predominant\_land\_useUrban -0.063232 0.024440  
## t value  
## (Intercept) 58.672  
## Predominant\_land\_useYoung secondary vegetation -1.301  
## Predominant\_land\_useIntermediate secondary vegetation -0.484  
## Predominant\_land\_useMature secondary vegetation 1.321  
## Predominant\_land\_usePlantation forest -2.580  
## Predominant\_land\_usePasture -3.013  
## Predominant\_land\_useCropland -3.775  
## Predominant\_land\_useUrban -2.587  
##   
## Correlation of Fixed Effects:  
## (Intr) P\_\_Ysv P\_\_Isv P\_\_Msv Pr\_\_Pf Prd\_\_P Prd\_\_C  
## Prdmnn\_\_Ysv -0.353   
## Prdmnn\_\_Isv -0.287 0.383   
## Prdmnn\_\_Msv -0.244 0.272 0.215   
## Prdmnnt\_\_Pf -0.373 0.241 0.225 0.198   
## Prdmnnt\_l\_P -0.348 0.407 0.304 0.179 0.172   
## Prdmnnt\_l\_C -0.387 0.384 0.250 0.184 0.420 0.350   
## Prdmnnt\_l\_U -0.249 0.170 0.207 0.114 0.174 0.162 0.183  
## convergence code: 0  
## boundary (singular) fit: see ?isSingular

## Plot the results

PlotGLMERFactor(model = m2,  
 data = model\_data,  
 responseVar = "sqrt(Rescaled Abundance)",  
 xtext.srt = 20,  
 seMultiplier = 1.96,  
 logLink="n",  
 catEffects = c("Predominant\_land\_use"))

