

# Practise\_FII\_Modelling

Patrick Alexander Walkden

25/02/2021

## Functional Intactness Index Model

To model FII I am using a two stage modeling framework. The first model will take site functional diversity, measured as Rao's Q, as function of land-use type and intensity. The second model will focus on functional similarity between sites of primary minimal habitat and all other land use types. Together this is will get a measure of overall functional diversity and the proportion of that function that is retained from primary minimal sites, which once projected on to current maps of the modeled pressures and multiplied together will (hopefully) give a measure of functional intactness.

### 1.1 Modelling: Functional Diversity (PGLMM)

Before modeling I considered whether sites that are more phylogenetically related to each other may respond similarly to landuse change and therefore may introduce non-independence into the modeling that would need to accounted for. To do so I first calculated pairwise phylogenetic similarity of sites with unifracs (the percentage of shared branch length between two communities). Using a clustering algorithm within `hclust` that does not assume an ultrametric structure we are able to see whether the sites are clustered phylogenetically and to what extent they form a heirachial structure. A high amount of clustering would indicate that there are sites that are a lot more similar to each other than expected if they are to be independent and therefore could show correlated responses. Low clustering would show that sites aren't that phylogenetically similar to each other anyway and therefore it is not a large problem to consider.

Secondly, I checked whether the responses to landuse within study showed phylogenetic non-independence within a model that does not consider phylogenetic relatedness. I did this by calculating the differences between random slopes within studies and then identifying whether studies that are more phylogenetically similar to each other (as calculated by unifracs) have significantly more similar slopes than the global distribution of slopes. Doing this iteratively from the closest related study to the most distant I can plot the probability of rejecting the null hypothesis of that there is no differences between the similarity of slopes. A clear trend from rejecting the null hypothesis in the most related studies to accepting the null hypothesis in the most distantly related studies will be indicative of non-independence.

I am using community phylogenetic generalized linear mixed effects models (PGLMM) using the package `phyr`. `Phyr` allows for bayesian methods of model fitting using INLA, which is advised with larger datasets. I also considered the UN-subregion, human population density ( $\ln + 1$  transformed) and density of roads at both a 1km and 50km radius surrounding the site. I included study and study block as additional random effects and tested for the inclusion of random slopes of the other pressures within study by comparing the DIC of the maximal models. The fixed effect structure was then determined through backwards stepwise model simplification by looking at whether the removal of fixed effects significantly reduced the DIC of the model.

```
rm(list = ls())
require(phytools)
require(picante)
require(phyr)
require(tidyverse)
```

```
require(raster)
require(ggirdges)
require(motmot)
require(lme4)
require(poolr)
require(car)
```

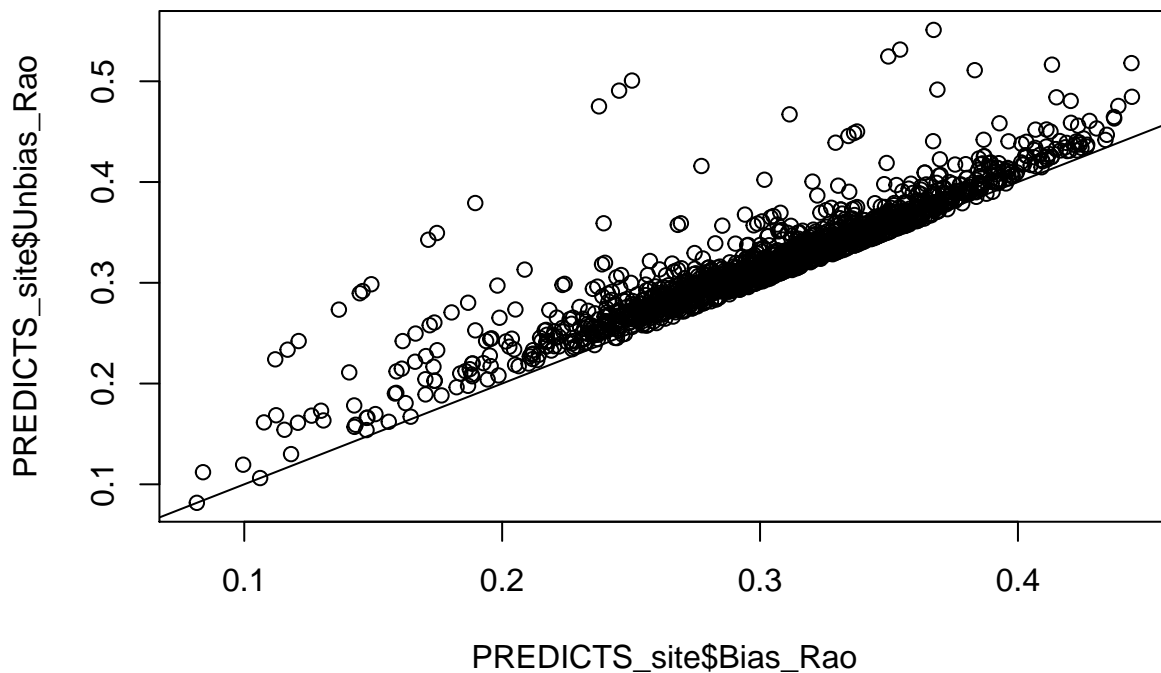
Load in the datasets from the previously calculated metrics.

```
## Load in the datasets
```

```
PREDICTS_site <- readRDS("../Functional_Intactness_Index/Outputs/PREDICTS_Site_Rao.rds")
PREDICTS_abundance <- readRDS("../Functional_Intactness_Index/Outputs/abundance_data.rds")
```

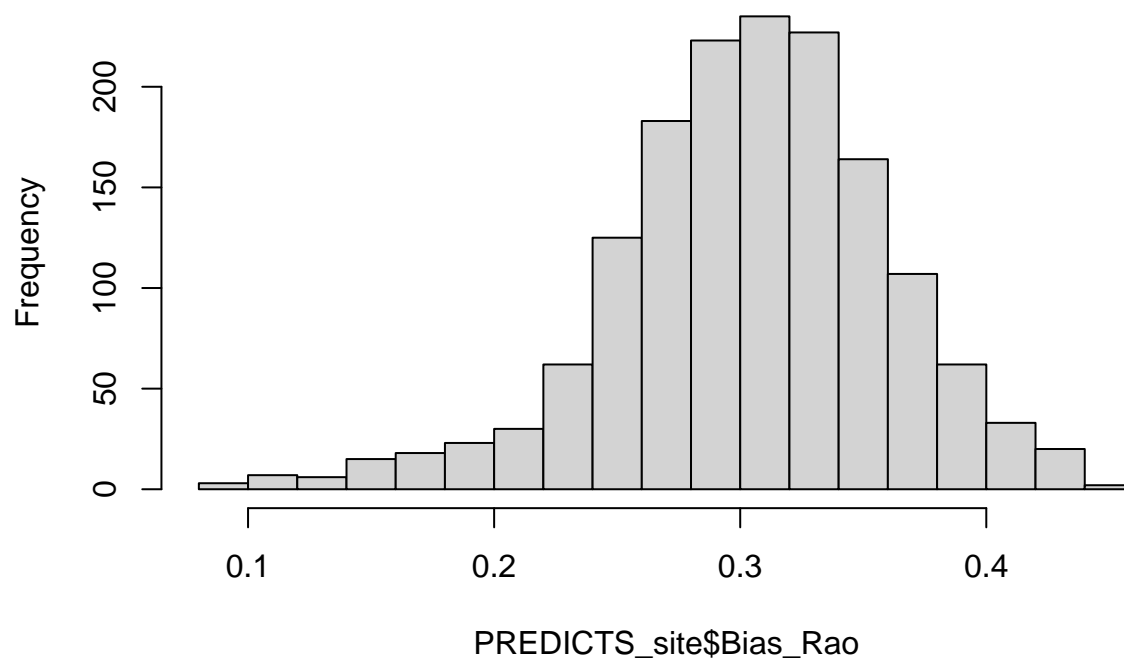
I calculated Rao's Q based on pairwise functional distances of species within a site weighted by their relative abundance. Chen et al, 2018 highlighted that the Rao estimator might be inherently bias underestimating the true value of functional diversity and proposed an unbiased metric. I have calculated both.

```
plot(PREDICTS_site$Unbias_Rao ~ PREDICTS_site$Bias_Rao)
abline(a=0, b=1)
```



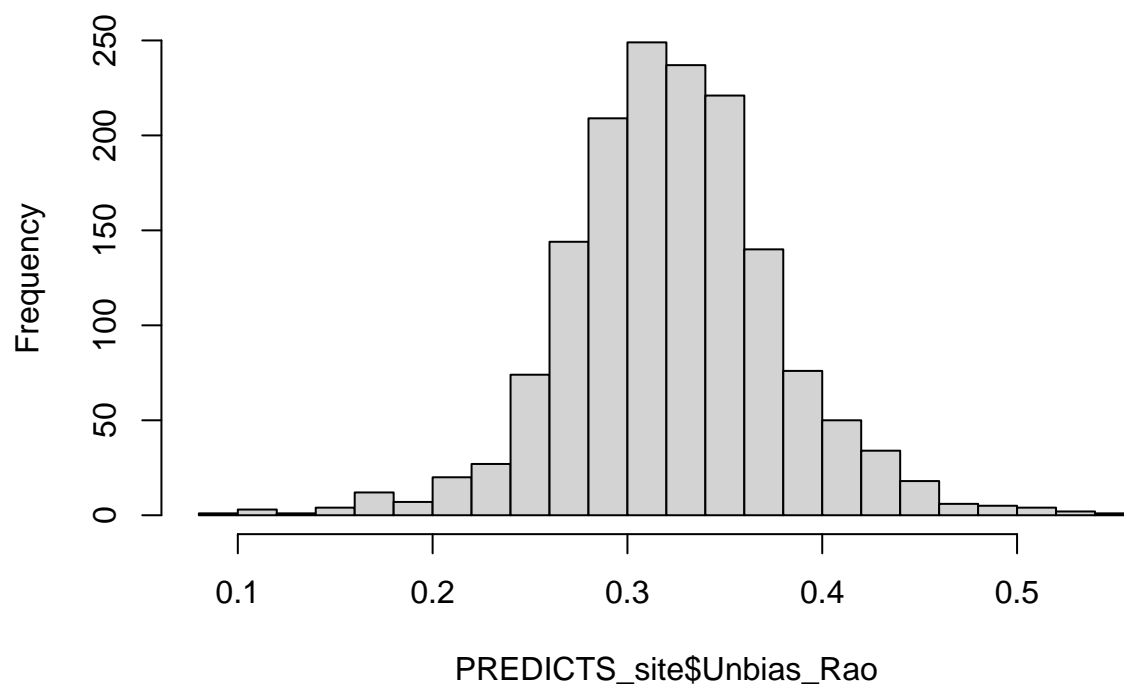
```
hist(PREDICTS_site$Bias_Rao, breaks = 20)
```

**Histogram of PREDICTS\_site\$Bias\_Rao**



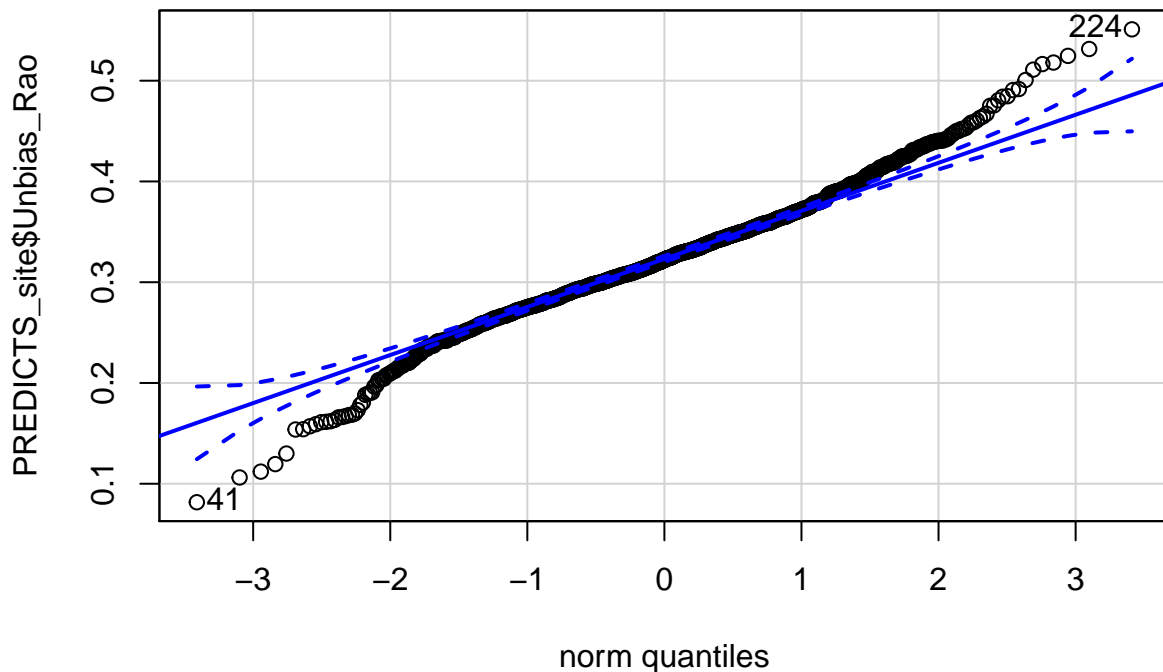
```
hist(PREDICTS_site$Unbias_Rao, breaks = 20)
```

**Histogram of PREDICTS\_site\$Unbias\_Rao**



### Outlier check

```
qqPlot(PREDICTS_site$Unbias_Rao)
```



```
## [1] 41 224
```

QQ plot has identified DI1\_2010\_\_Milder 2 41 and GN1\_2007\_\_VanBael 1 9 64 as potential outliers. Looking at the makeup of the sites it seems that DI1\_2010\_\_Milder 2 41 is completely dominated by a single species *Dives dives* and from that I noticed DI1\_2010\_\_Milder 2 48 is also dominated by *Bubulcus ibis* as well which is really skewing the calculation of the estimator. I looked at the original study published by Milder et al 2010 and I think there was an input error as Table 3 doesn't indicate the number of individuals being that high. So I don't know whether to exclude the data points or go back and recalculate the FD for the sites while correcting for sites that are dominated by a single species (possibly wrongly). For now I'll drop them.

With GN1\_2007\_\_VanBael 1 9 64 I don't know whether to drop it may be a true value of biological relevance.

```
PREDICTS_site <- PREDICTS_site[-c(41,48),]
rownames(PREDICTS_site) <- seq(1,nrow(PREDICTS_site),1)
```

Next we'd like to have a look at how the distribution of land use types and intensities we have across our modeled sites.

```
table(PREDICTS_site$LandUse, PREDICTS_site$Use_intensity)
```

```
##
##           Minimal use Light use Intense use
## Cropland           33         38          0
## Pasture            40         53         63
## Plantation forest  119         36          0
## Primary            417        335         49
## Secondary Vegetation 127         46         18
## Urban             113         56          0
```

Although Plantation forest exists as a land use class within PREDICTS it is rarely incorporated in global land-use layers therefore I included Plantation minimally used habitat in secondary vegetation light use and Plantation Light use in secondary intense use following De Palma et al, in Review. Because we are collapsing some land use classes and intensities in on each other I have calculated another factor or LUI

*#### collapse Plantation forest into secondary vegetation and relevel the LUI variable to have Primary*

```
PREDICTS_site$SSBS <- factor(PREDICTS_site$SSBS) %>% droplevels()

PREDICTS_site <- PREDICTS_site %>% dplyr::mutate(LandUse_Intensity = ifelse(grepl(LandUse_Intensity, pattern = "Secondary Vegetation_Light use",
paste(LandUse_Intensity)),
LandUse_Intensity = ifelse(grepl(LandUse_Intensity, pattern = "Secondary Vegetation_Intense use",
paste(LandUse_Intensity)),
LandUse_Intensity = relevel(factor(LandUse_Intensity),
ref = "Primary_Minimal use"))

table(PREDICTS_site$LandUse_Intensity)
```

```
##
##           Primary_Minimal use           Cropland_Light use
##                   417                   38
##           Cropland_Minimal use           Pasture_Intense use
##                   33                   63
##           Pasture_Light use           Pasture_Minimal use
##                   53                   40
##           Primary_Intense use           Primary_Light use
##                   49                   335
## Secondary Vegetation_Intense use Secondary Vegetation_Light use
##                   54                   165
## Secondary Vegetation_Minimal use           Urban_Light use
##                   127                   56
##           Urban_Minimal use
##                   113
```

*##### Going to rename for some ease of outputs*

```
PREDICTS_site <- PREDICTS_site %>% dplyr::rename(LUI = LandUse_Intensity)
levels(PREDICTS_site$LUI) <- c("PriMin", "CrpLig", "CrpMin", "PasIn", "PasLig",
"PasMin", "PriIn", "PriLig", "SecIn", "SecLig", "SecMin", "UrbLig", "UrbMin")
```

*##### Then if the modelling of LUI is not sufficient there is also a LandUse factor that has plan*

```
PREDICTS_site <- PREDICTS_site %>% dplyr::mutate(LandUse = ifelse(LUI == "PriMin", "PriMin", paste(LandUse, "Plantation forest", "Secondary Vegetation_Light use", "Secondary Vegetation_Intense use", "Pasture_Light use", "Pasture_Minimal use", "Cropland_Light use", "Urban_Light use", "Urban_Minimal use", sep = " "))
LandUse = ifelse(LandUse == "Plantation forest", "Secondary Vegetation_Light use", "Secondary Vegetation_Intense use", "Pasture_Light use", "Pasture_Minimal use", "Cropland_Light use", "Urban_Light use", "Urban_Minimal use", sep = " ")
LandUse = relevel(factor(LandUse), ref = "PriMin"))
```

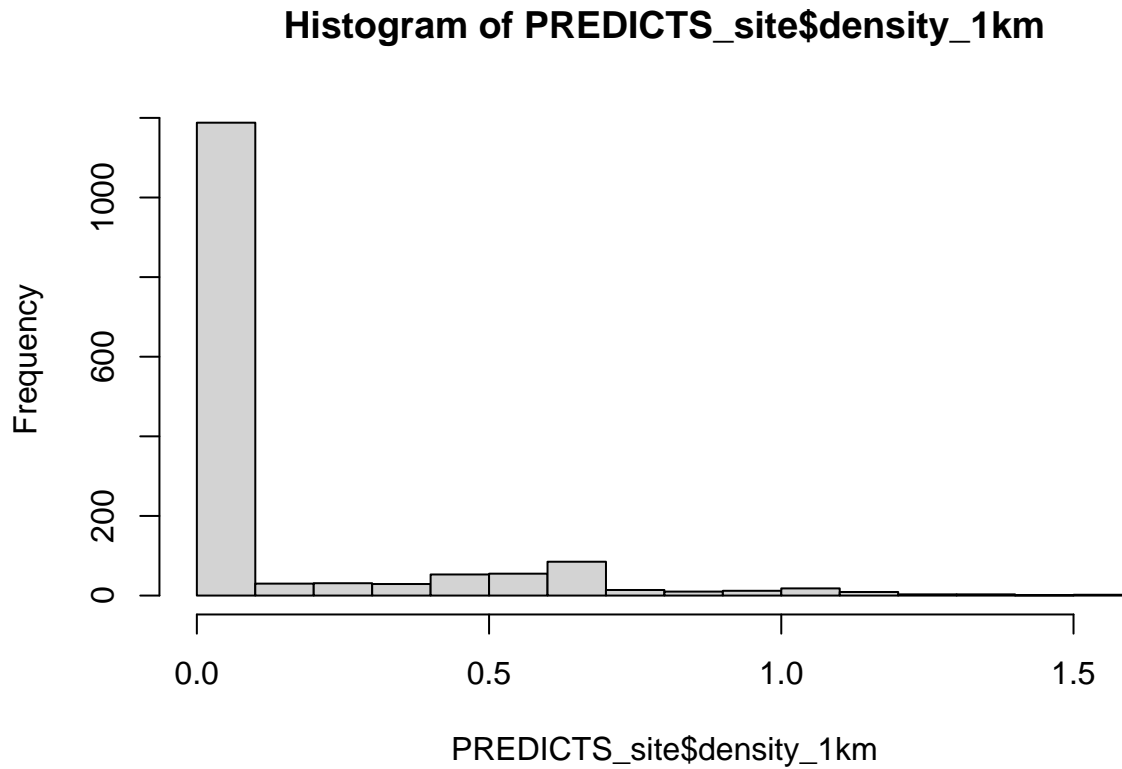
```
table(PREDICTS_site$LUI)
```

```
##
## PriMin CrpLig CrpMin PasIn PasLig PasMin PriIn PriLig SecIn SecLig SecMin
## 417 38 33 63 53 40 49 335 54 165 127
## UrbLig UrbMin
## 56 113
```

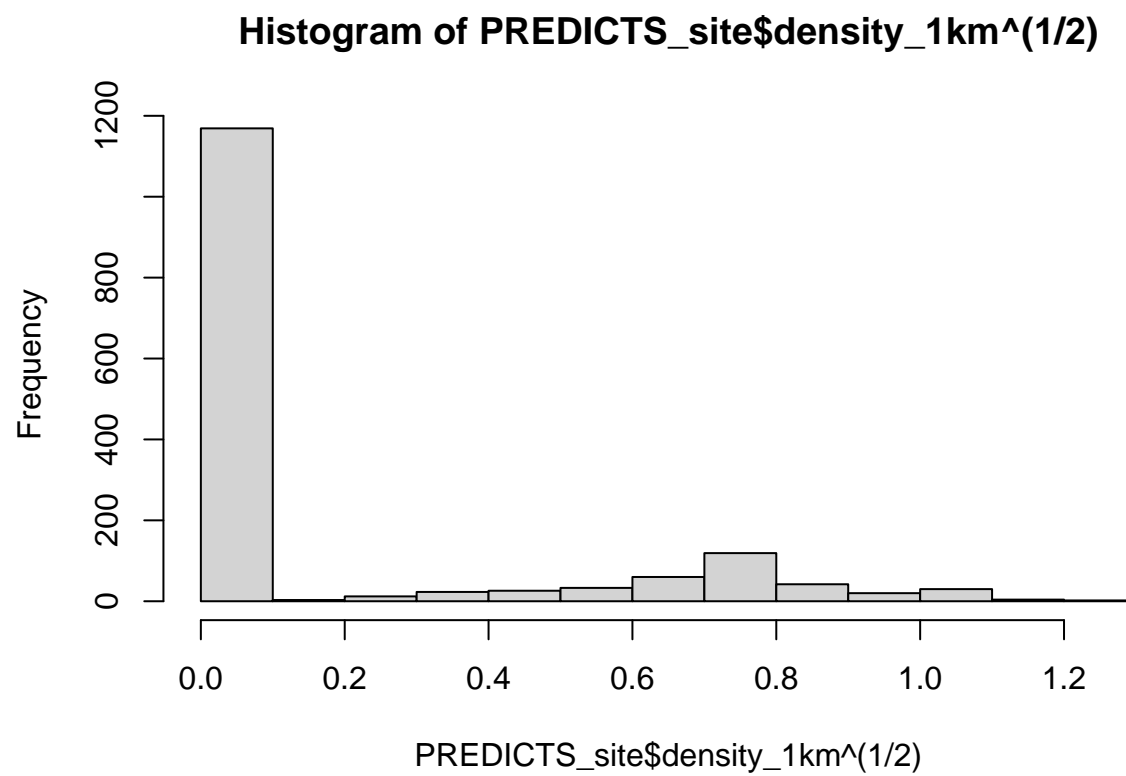
## Variable transformations

The other continuous variables that I will be including in my model will be human population density, which is already log transformed, and the density of roads at 1km and 50km radius from sites.

```
hist(PREDICTS_site$density_1km)
```



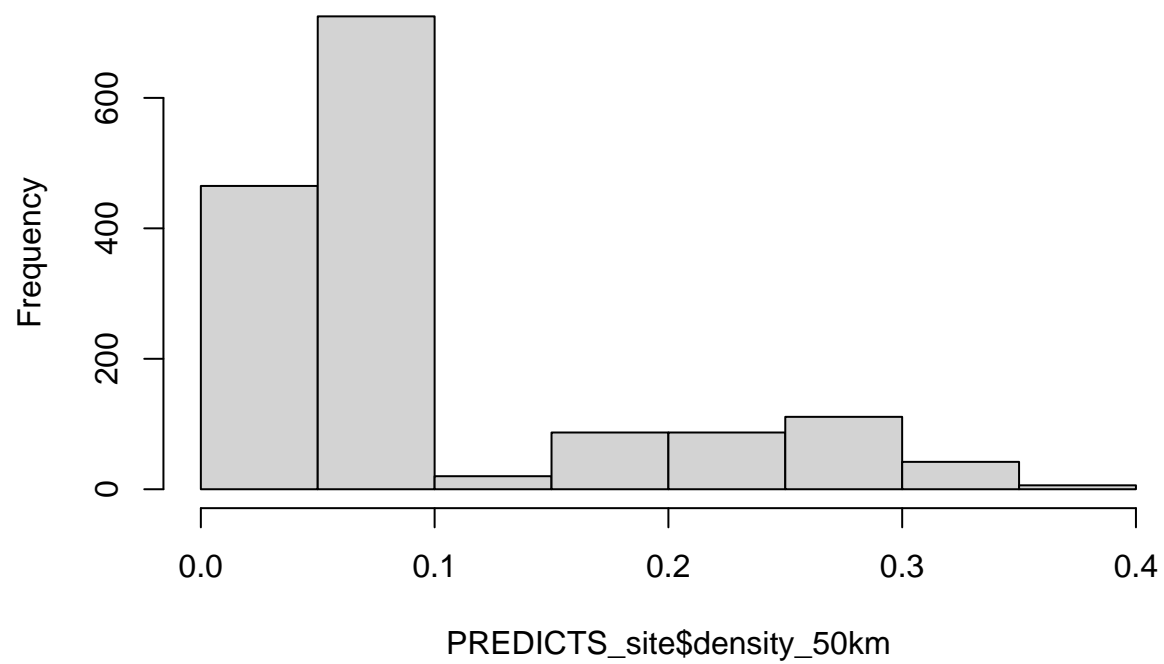
```
hist(PREDICTS_site$density_1km^(1/2))
```



```
hist(PREDICTS_site$density_50km)
```

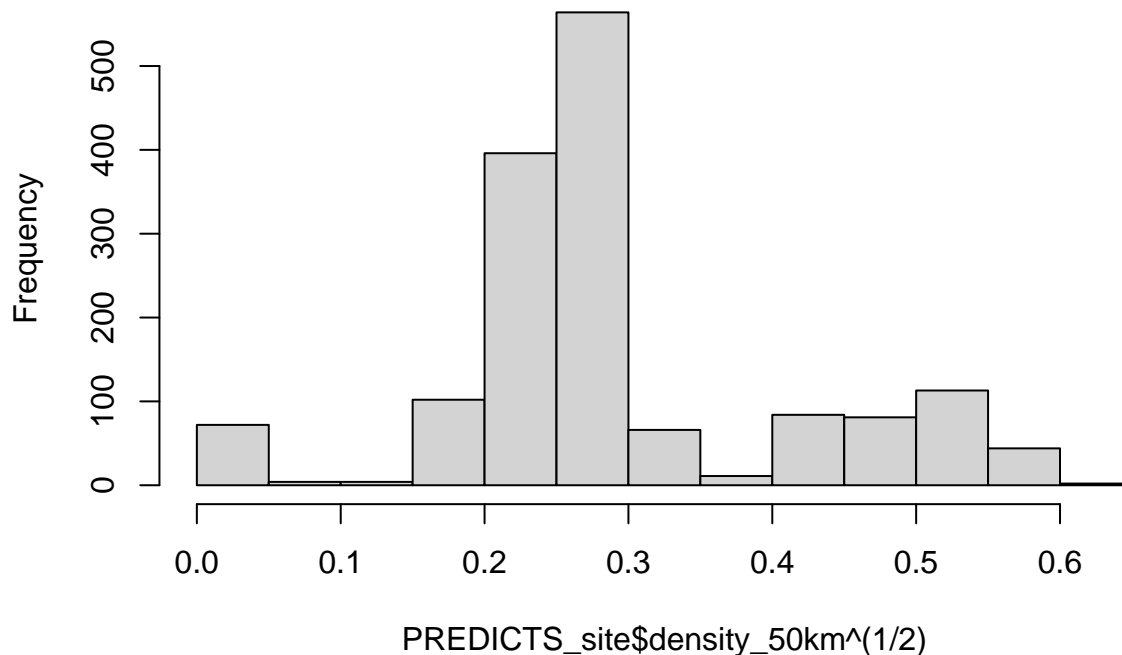


**Histogram of PREDICTS\_site\$density\_50km**



```
hist(PREDICTS_site$density_50km^(1/2))
```

## Histogram of PREDICTS\_site\$density\_50km<sup>(1/2)</sup>



### Colinearity

Next I am going to check for colinearity in the predictor variables, but beforehand I apply the square root transformation to the density of roads and scale all variables to reduce possible colinearity issues.

```
PREDICTS_site$logHPD <- scale(PREDICTS_site$logHPD)
PREDICTS_site$RD1k <- scale(PREDICTS_site$density_1km^(1/2))
PREDICTS_site$RD50k <- scale(PREDICTS_site$density_50km^(1/2))

#source("https://highstat.com/Books/Book2/HighstatLibV10.R")
#corvif(PREDICTS_site[,c("LUI", "logHPD", "RD1k", "RD50k", "UN_subregion")])

### a lot of colinearity try removing UN-subregion
#corvif(PREDICTS_site[,c("LUI", "logHPD", "RD1k", "RD50k")])

## removing UN_subregion has reduced the colinearity to acceptable levels
```

### Phylogenetic Similarity

Finally I calculate the pairwise phylogenetic similarity between sites, as mentioned before I will be using 1-unifrac.

```
#### Load in Phylogeny - just tae the first one when running the inital models

all_bird_tree <- read.tree("../Datasets/AllBirdsHackett1.tre")
```

```

all_bird_tree <- all_bird_tree[[1]]

#####
#### Unifrac for all sites among studies ####
#####

create_vcv <- function(data, level) {

  data <- droplevels(data)
  species <- sub(unique(data$Jetz_Name), pattern = " ", replacement = "_")
  drop.species <- all_bird_tree$tip.label[which(!(all_bird_tree$tip.label %in% species))]

  overall_tree <- drop.tip(all_bird_tree, drop.species)

  ID <- as.character(unique(PREDICTS_site[, level]))

  comm_data <- t(species)
  colnames(comm_data) <- species
  comm_data <- data.frame(comm_data[-1,])

  for(i in 1:length(ID)){

    ID_data <- data.frame(data[data[,level] == ID[i], c("Jetz_Name", "Effort_Corrected_Measurement")])

    for(spp in species){
      comm_data[i, paste(spp)] <- ifelse(any(ID_data[ID_data$Jetz_Name == sub(spp, pattern = "_", replacement = " ")]),
      ID_data[which(ID_data$Jetz_Name == sub(spp, pattern = "_", replacement = " ")), "Effort_Corrected_Measurement"], 0)
    }

    rownames(comm_data)[i] <- ID[i]
  }

  for(i in 1:ncol(comm_data)){
    comm_data[,i] <- as.numeric(comm_data[,i])
  }

  comm_data <- as.matrix(comm_data)

  suppressWarnings(memory.limit(120000))

  vcv <- 1 - as.matrix(unifrac(comm = comm_data, tree = overall_tree))

```

```

    return(vcv)
}

#site_vcv <- among_site_unifrac(PREDICTS_abundance, level = "SSBS")

#write_rds(file = "../Functional_Intactness_Index/Outputs/site_vcv.rds", site_vcv)

```

## Checking for phylogenetic signal in functional diversity and the responses of sites to land use change

```

##### Identify whether there is phylogenetic signal of in the responses of sites of LU change #####
##### first we are going to create a cluster dendrogram based on the phylogenetic distances between sites
##### Using an non-ultmetric tree we will be able to see just how heirarchial the resulting tree is, and
##### in the modelling.

PREDICTS_site <- PREDICTS_site %>% droplevels()

site_vcv <- readRDS("Outputs/site_vcv.rds")

studies <- as.character(unique(PREDICTS_site$SS))

## get the first site from each study as it is very difficult to observe the full tree with all sites

first_site <- c()
for(i in 1:length(studies)){
  first <- which(grepl(studies[i], colnames(site_vcv)))[1]
  first_site <- c(first_site, first)
}

first_sites <- site_vcv[first_site, first_site]

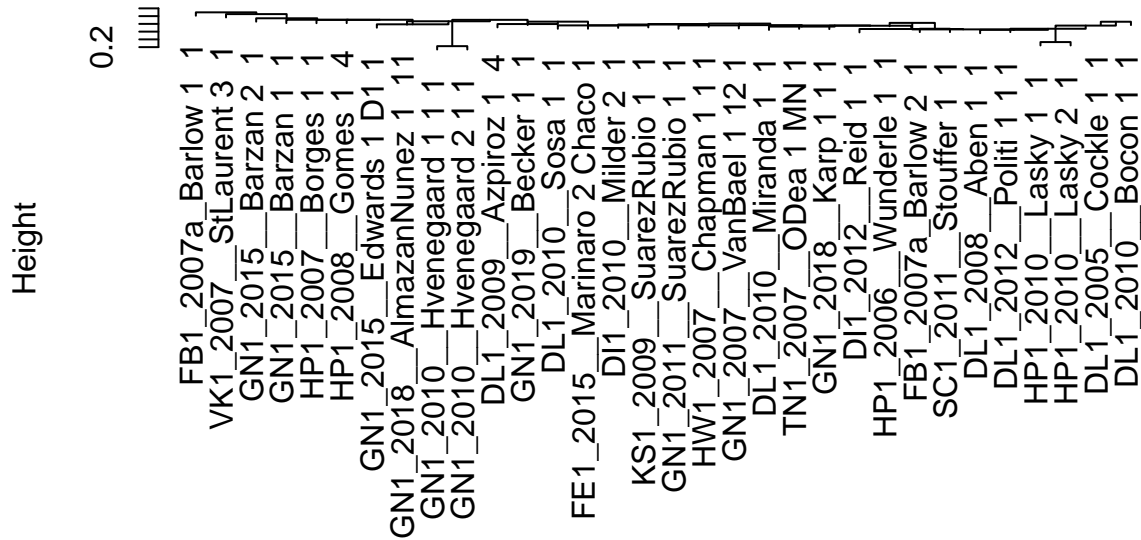
### performing a clustering algorithm based on the distances between sites - first convert back to diss

site_dendro <- hclust(as.dist(1 - first_sites), method = "median")

plot(site_dendro)

```

## Cluster Dendrogram



```
as.dist(1 - first_sites)
hclust (*, "median")
```

### Can actually see that there is not a large amount of heirarchical structure within the sites with on

The two Lasky studies are related to each other because they are studies at the same sites but between the wet and the dry season - should I then collapse them in on each other? The Hvenegaard studies are again at the same sites but conducted at a 50m radius and unlimited radius again should probably remove or collapse the studies into each other?

#### This initially does not suggest that there is a large amount of phylogenetic similarity between sites  
 #### correlated responses to land-use change - but another check I am going to conduct is to see whether  
 #### slopes of landuse within studies are more similar in studies that are more phylogenetically similar  
 #### of slope differences.

#### First calculate study level phylogenetic similarity

```
Study_sim <- create_vcv(PREDICTS_abundance, level = "SS")
```

### Next run a model without accounting for phylogenetic signal of sites and extract the random slopes

#1. fit model with no phylogeny in error term

```
test_mod <- lmer(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +
  (1|SS) + (1|SSB) + (1 + LUI|SS), data = PREDICTS_site)
```

```

## boundary (singular) fit: see ?isSingular
#### extract the random effects

Randoms <- lme4::ranef(test_mod)

study_slope <- Randoms$SS

significance <- c()

for(position in 1:c(length(studies)-1)){

  position_diff <- c()
  relatives <- c()

  for(study in studies){

    position_rel <- colnames(Study_sim)[order(Study_sim[study,],decreasing = TRUE)[position + 1]]
    rel_diff <- abs(study_slope[study,] - study_slope[position_rel,])
    rownames(rel_diff) <- paste(position_rel)

    position_diff <- rbind(position_diff, rel_diff)
    relatives <- rbind(relatives, data.frame(stud1 = study, stud2 = position_rel) )

  }

  stud_com <- data.frame(gtools::combinations(v = studies, r = 2, n = c(length(studies)-1)))

  drops <- c()
  for(k in 1:nrow(relatives)){
    drop_row <- which(stud_com$X1 == relatives[k,1] & stud_com$X2 == relatives[k,2] | stud_com$X1 == relatives[k,2] & stud_com$X2 == relatives[k,1])
    drops <- c(drops, drop_row)
  }

  stud_com <- stud_com[-unique(drops),]

  global_differences <- c()
  for(i in 1:NROW(stud_com)){
    abs_diff <- abs(study_slope[stud_com[i,1],] - study_slope[stud_com[i,2],])
    rownames(abs_diff) <- paste(stud_com[i,2])

    global_differences <- rbind(global_differences, abs_diff)
  }

  p_val <- wilcox.test(position_diff[,2], global_differences[,2], alternative = "less")
  p_val <- data.frame(Position = position, significance = p_val$p.value)

  significance <- rbind(significance, p_val)
}

```

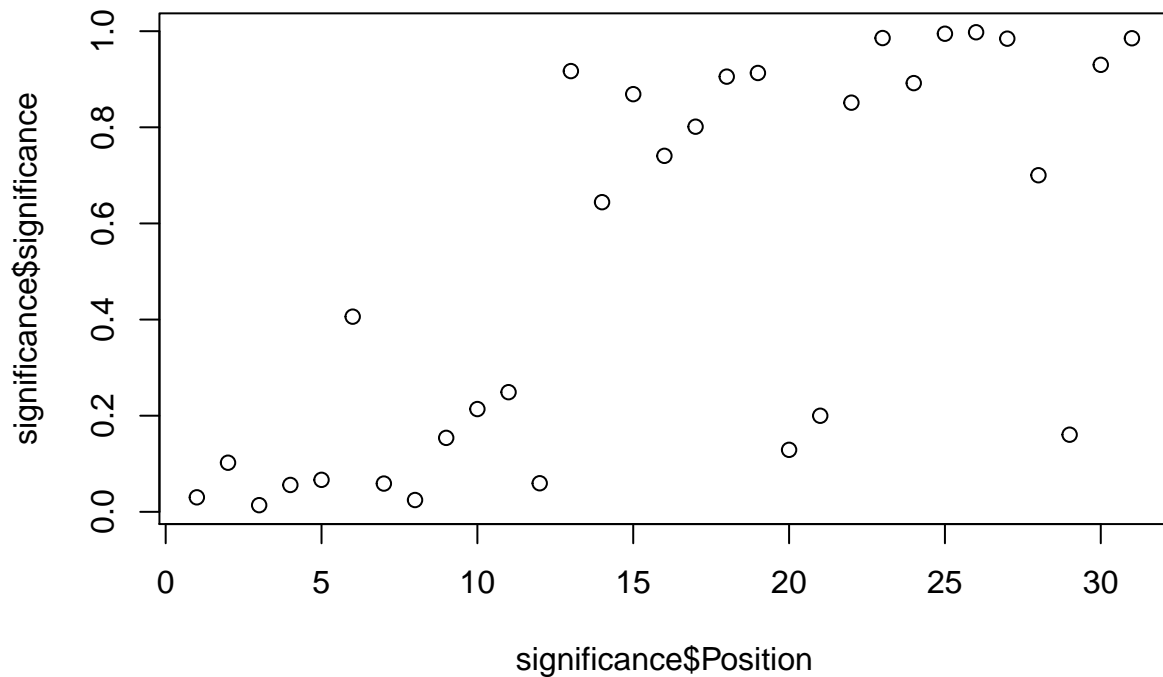
```

}

#### plot doesn't clearly indicate that the responses of sites within studies are significantly more si
#### similar sites

plot(significance$significance ~ significance$Position)

```



```

### fishers method - combined p values testing whether the hypothesis test (are the difference in slopes
### to each other in studies that are more phylogenetically related?)

```

```

### No they are not

```

```

poolr::fisher(significance$significance)

```

```

## combined p-values with:      Fisher's method
## number of p-values combined: 31
## test statistic:              76.692 ~ chi-square(62)
## adjustment:                  none
## combined p-value:            0.0992

```

There does seem to be a slight association between the probability of rejecting the null hypothesis and the phylogenetic similarity of the comparison study so that when comparing the differences in slopes in more phylogenetically related studies there is a greater probability of the differences in slopes being significantly more similar to each compared to the global distribution of differences in slopes. However, when performing a Fishers combined probability there is not a significant difference so we can accept the null hypothesis. — Possibly if I drop the two related studies *lasky* and *hvengaard*... Removing one of each made the p-value

from fishers combined increase a good amount.

Therefore, that does not indicate to me at the moment that I need to account for phylogenetic similarity of sites within the modelling framework, as there is not a clear hierarchy of sites with greater phylogenetic similarity to each other and they are not significantly ore similar in their responses to ;and use change therefore I will continue on with GLMMs, however, I will continue to use the `phyr` package as it will allow me to fit models using INLA and bayesian techniques, which helps with some convergence issues.

## INLA/GLMM

With that we now have everything ready to begin modeling. Random effects that I will be including in the models will be study, to account for between study differences in location, method and sampling effort and study block, to account for spatial configuration of sites. Additionally, I will be testing for random slopes for each of the predictor variables within study by identifying which random effect structure has the lowest DIC in the maximal model.

```
#####  
# With this we can proceed with modelling using GLMMs as opposed to PGLMMs  
#####  
PREDICTS_site <- PREDICTS_site %>% droplevels()  
  
Rao_Model_1b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +  
                                     LUI:logHPD + LUI:RD1k + LUI:RD50k +  
                                     (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)  
  
#### lets see what random effect structure gives the best AIC  
  
### adding random slopes 1) LUI within study  
  
Rao_Model_2b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +  
                                     LUI:logHPD + LUI:RD1k + LUI:RD50k +  
                                     (1|SS) + (1|SSB) + (LUI|SS), data = PREDICTS_site, bayes = TRUE)  
  
#2) logHPD within study  
  
Rao_Model_3b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +  
                                     LUI:logHPD + LUI:RD1k + LUI:RD50k +  
                                     (1|SS) + (1|SSB) + (logHPD|SS), data = PREDICTS_site, bayes = TRUE)  
  
#3) Road Density_1km within study  
  
Rao_Model_4b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +  
                                     LUI:logHPD + LUI:RD1k + LUI:RD50k +
```



```

      (1|SS) + (1|SSB) + (RD1k|SS), data = PREDICTS_site, bayes = TRUE)

# 4) RD50km within study

Rao_Model_5b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +
  LUI:logHPD + LUI:RD1k + LUI:RD50k +
  (1|SS) + (1|SSB) + (RD50k|SS), data = PREDICTS_site, bayes = TRUE)

Random_DIC <- data.frame(mod1 = Rao_Model_1b$DIC, mod2 = Rao_Model_2b$DIC, mod3 = Rao_Model_3b$DIC,
  mod4 = Rao_Model_4b$DIC, mod5 = Rao_Model_5b$DIC)

Random_DIC

##      mod1      mod2      mod3      mod4      mod5
## 1 -5379.99 -5374.548 -5379.487 -5374.497 -5379.808

The model with the lowest DIC is with no random slope therefore model 1 will be carried forward.

Next to identify the optimal fixed effect structure I removed both interactions in turn to see whether that
reduced the DIC of the model. Typically a reduction >2 will provide support for the simplified model.

#### Remove each interaction to see if this improves the model at all.

#LUI:RD50k

Rao_Model_6b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +
  LUI:logHPD + LUI:RD1k +
  (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)

#LUI:RD1k

Rao_Model_7b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +
  LUI:logHPD + LUI:RD50k +
  (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)

# LUI:logHPD

Rao_Model_8b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k + RD50k +
  LUI:RD50k + LUI:RD1k +
  (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)

Model_differences <- data.frame(RD50k = Rao_Model_6b$DIC - Rao_Model_1b$DIC,
  RD1k = Rao_Model_7b$DIC - Rao_Model_1b$DIC,
  logHPD = Rao_Model_8b$DIC - Rao_Model_1b$DIC)

```

#### Model DIC is reduced in all but most when removing the interaction between LUI:RD1k so model 7b is

## RD1k fixed

```
Rao_Model_9b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD50k +  
                                     LUI:logHPD + LUI:RD50k +  
                                     (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

# LUI:RD50k

```
Rao_Model_10b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD50k + RD1k +  
                                       LUI:logHPD +  
                                       (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

# LUI:logHPD

```
Rao_Model_11b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD50k + RD1k +  
                                       LUI:RD50k +  
                                       (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

```
Model_differences <- data.frame(RD1k = Rao_Model_9b$DIC - Rao_Model_7b$DIC,  
                               RD50k = Rao_Model_10b$DIC - Rao_Model_7b$DIC,  
                               logHPD = Rao_Model_11b$DIC - Rao_Model_7b$DIC)
```

Model\_differences

```
##      RD1k      RD50k      logHPD  
## 1 4.880973 -12.76837 -6.536527
```

#### Model DIC is reduced in all but most when removing the interaction between LUI:RD50k so model 10b

## RD50k fixed

```
Rao_Model_12b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k +  
                                       LUI:logHPD +  
                                       (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

### RD1k fixed

```
Rao_Model_13b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD50k +  
                                       LUI:logHPD +  
                                       (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

```
#### LUI:logHPD
```

```
Rao_Model_14b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD50k + RD1k +  
                                     (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

```
Model_differences <- data.frame(RD50k = Rao_Model_12b$DIC - Rao_Model_10b$DIC,  
                                RD1k = Rao_Model_13b$DIC - Rao_Model_10b$DIC,  
                                logHPD = Rao_Model_14b$DIC - Rao_Model_10b$DIC)
```

```
Model_differences
```

```
##      RD50k      RD1k      logHPD  
## 1 0.1060146 8.254223 -7.334714
```

```
#### Model DIC is reduced wehn removing the interaction between LUI:logHPD and is increased in the other
```

```
#### RD50k
```

```
Rao_Model_15b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD1k +  
                                     (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

```
#### RD1k
```

```
Rao_Model_16b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + logHPD + RD50k +  
                                     (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

```
### logHPD
```

```
Rao_Model_17b <- phyr::communityPGLMM(Unbias_Rao ~ LUI + RD1k + RD50k +  
                                     (1|SS) + (1|SSB), data = PREDICTS_site, bayes = TRUE)
```

```
Model_differences <- data.frame(RD50k = Rao_Model_15b$DIC - Rao_Model_14b$DIC,  
                                RD1k = Rao_Model_16b$DIC - Rao_Model_14b$DIC,  
                                logHPD = Rao_Model_17b$DIC - Rao_Model_14b$DIC)
```

```
####
```

```
Model_differences
```

```
##      RD50k      RD1k      logHPD
```

```
## 1 -0.9806204 6.684843 -1.206768
```

## Final Model

After going through model selection the best model seems to be the one with LUI, logHPD and RD1k and RD50k as Fixed effects and no interactions between LUI and any of the other pressures with the best model being Rao\_Model\_14b.

*## funtion to visualise the residuals from the PGLMM and check whether they are normally distributed*

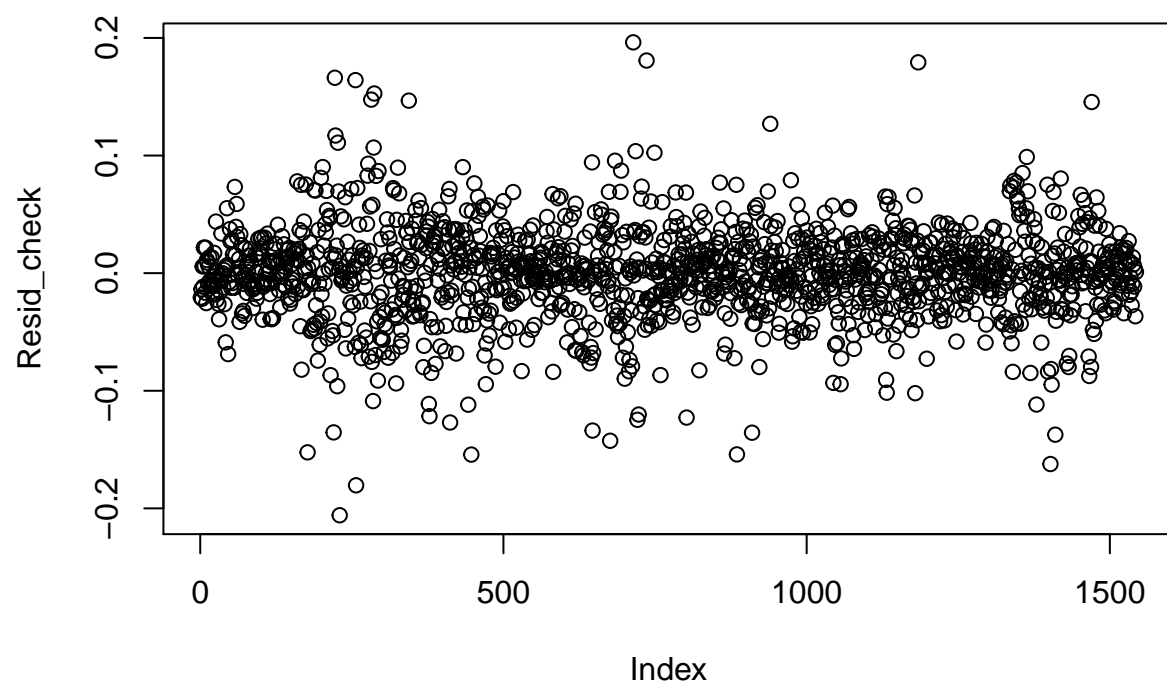
```
residuals.communityPGLMM <- function(
  object,
  type = if(object$family %in% c("binomial","poisson")) "deviance" else "response",
  scaled = FALSE, ...){
  if(object$family == "gaussian"){
    y <- object$Y
    mu <- pglmm_predicted_values(object)$Y_hat
    res <- switch(type,
                  deviance = stop("no deviance residuals for gaussian model", call. = FALSE),
                  response = y - mu
    )
    if(scaled) res/sqrt(object$s2resid)
  }

  if(object$family %in% c("binomial","poisson")){
    y <- as.numeric(object$Y)
    mu <- unname(object$mu[, 1])
    if(object$family == "binomial") dres <- sqrt(binomial()$dev.resids(y, mu, 1))
    if(object$family == "poisson") dres <- sqrt(poisson()$dev.resids(y, mu, 1))
    res <- switch(type,
                  deviance = {
                    dres
                    ifelse(y > mu, dres, - dres)
                  },
                  response = y - mu
    )
  }
  if(object$family %nin% c("gaussian", "binomial", "poisson"))
    stop("no residual methods for family other than gaussian, binomial and poisson, yet", call. = FALSE)

  unname(res)
}

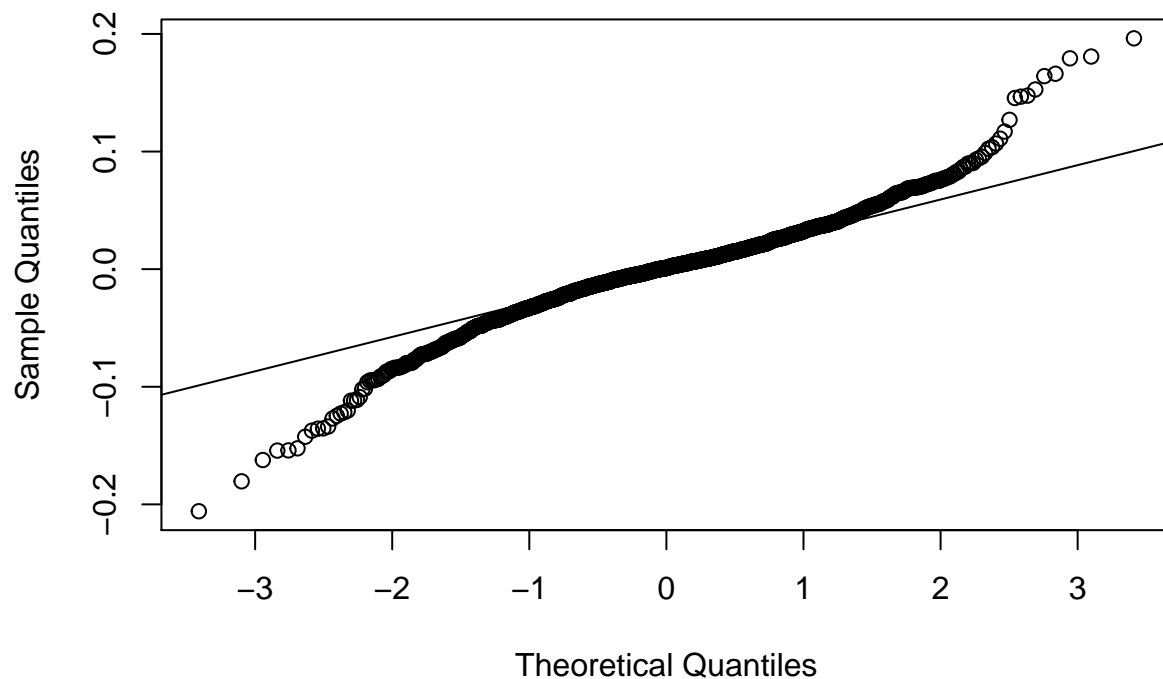
Resid_check <- residuals.communityPGLMM(Rao_Model_14b)

plot(Resid_check)
```



```
qqnorm(Resid_check)
qqline(Resid_check)
```

## Normal Q-Q Plot



```
write_rds(file = "Outputs/Rao_model.rds", Rao_Model_14b)
```

Rao\_Model\_14b

```
summary(Rao_Model_14b)
```

```
## Linear mixed model fit by Bayesian INLA
##
## Call:Unbias_Rao ~ LUI + logHPD + RD50k + RD1k
##
## marginal logLik          DIC          WAIC
##           2535          -5413          -5413
##
## Random effects:
##           Variance Std.Dev lower.CI upper.CI
## 1|SS      0.0009970 0.03158 5.972e-04 0.0018470
## 1|SSB      0.0001636 0.01279 9.555e-05 0.0002925
## residual 0.0016447 0.04055 1.530e-03 0.0017724
##
## Fixed effects:
##           Value      lower.CI upper.CI
## (Intercept) 0.33607000 0.32830099 0.3439
## LUICrpLig    -0.04767673 -0.06344903 -0.0319
## LUICrpMin    -0.02747767 -0.04456758 -0.0104
## LUIPasIn     -0.02290568 -0.04039390 -0.0054
## LUIPasLig    -0.01145595 -0.02471468 0.0018
```

## LUIPasMin	-0.03942799	-0.05430201	-0.0246
## LUIPriIn	0.00467118	-0.00902397	0.0184
## LUIPriLig	-0.00406234	-0.01273282	0.0045
## LUISecIn	0.00050143	-0.01470222	0.0157
## LUISecLig	-0.01734341	-0.02874447	-0.0060
## LUISecMin	-0.01618459	-0.02659290	-0.0058
## LUIUrbLig	-0.00502072	-0.03161615	0.0215
## LUIUrbMin	-0.01177387	-0.02352796	-0.0001
## logHPD	0.00631699	-0.00111954	0.0136
## RD50k	-0.00275997	-0.01217165	0.0067
## RD1k	0.00453902	0.00173988	0.0073

The summary of the final Rao model shows that Functional diversity, compared to primary minimal habitat, is reduced across LUIs except Primary intense use and Secondary intense use where it increases slightly. Significant reductions, as determined by the credible intervals of the parameter distribution not crossing zero, were shown in cropland light and minimal, pasture intense and minimal, secondary light and minimal and urban minimal habitats. In response to the other pressures FD increases with greater logHPD and road density at 1km radius, while FD decreases with Road density at 50km.

Running the model selection framework with the “biased” Rao yielded similar results however, the best model also included a random slope of LUI within study which resulted in FD swapping from observed declines in Primary light use, and urban light use to positive affecting functional diversity.

## 1.2 Modelling: Functional Similarity/Overlap (GLMM)

To assess the impact of landuse change and other pressures on functional similarity I used GLMMS. The other pressures include human population density, density of roads @ 1 and 50km radii, and geographic and environmental distances. Additionally, since the functional similarity of sites may be influenced by the absolute level of pressure experienced at that site but also the difference in pressure experience between the sites, I included the pressure at site 2 (the site being compared to Primary minimal land use) and the difference in pressures between site 1 and 2.

To determine the optimal random effect structure I used the aforementioned selection process picking the structure with the lowest AIC. To determine the best fixed effect structure is more difficult however, as the comparisons are not-independent of each other as a single primary minimal site will be compared to multiple of other sites, therefore traditional likelihood ratio tests will not be suitable. Again following De Palma et al. In review, I permuted the data 1000 times by randomly shuffling the functional similarity scores within studies and refitted the full and simplified model to each permuted dataset to generate a distribution of Likelihood ratio scores. Comparing the likelihood ratio score between models using the observed data to this distribution, a simplified model would be supported if the score is lower than the 95th percentile of the distribution as it has not significantly reduced the explanatory power of the model.

Model 1 seems to have the most support so we will carry on without any random slopes. Next stage is to perform backwards stepwise model simplification based on the likelihood ratio distribution test on 1000 permuted datasets.

Finally testing the removal of all other fixed effects and interactions in the model I have found that **Model\_16** is the best model where these removals result in a model with significantly less explanatory power based on the permuted datasets.

Looking more closely at the model output we can see that the similarity of communities compared to that of primary minimally used sites