

Extinction risk from climate change: taxonomy

March 25th, 2024

Load libraries and data

```
rm(list = ls())
root.dir = "C:/Users/mcu08001/Documents/1New Research/CC MetaRisk2/Analysis"

#Load Libraries
library(coda); library(ggplot2); library(rstan); library(bayesplot); library(
shinystan); library(loo); library(rstanarm); library(dplyr); library(ggpubr)
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = FALSE)

#Load data
dataP<-read.table("MetaRisk2 aggthres 5.txt",header=T);
attach(dataP)
#note that do not need to eliminate NAs like for pre-industrial

#betareg requires no 0s or 1s
koffset = 0.001 #the k that gives the best posterior predictive check
percent2 <- adj.percent
percent2[adj.percent == 0] = koffset;
percent2[adj.percent == 1] = 1 - koffset;
dataP$percent2 <- percent2;

data.use<-dataP
N = length(data.use$percent2)
n.Study <- length(unique(data.use$Study)) #number of studies
Studyint<-as.integer(unclass(factor(data.use$Study)))
phi = data.use$Total.N
```

Test for effect of Taxonomic group

Here I tested if extinction predictions change based on taxonomic group.

```
#create model matrix for coefficients
betamat <- (model.matrix(~Taxa,data=data.use)) #create model matrix 1 - amphi
bians, 2 - birds, 3 - fish, 4 - insects, 5 - inverts, 6 - mammals, 7 - mixed,
8 - plants, 9 - reptiles

stan.data<-list(N = N, percent = data.use$percent2, betamat = betamat, phi =
phi, S = n.Study, P = ncol(betamat), Study = Studyint)
```

```

params.to.monitor=c("beta","y_rep","stu","sigma_stu", "eta","log_lik")

init.beta=rep(0,ncol(betamat)-1)
init.fn<- function (chain_id) {
  list(beta = c(-2.5,init.beta))
}

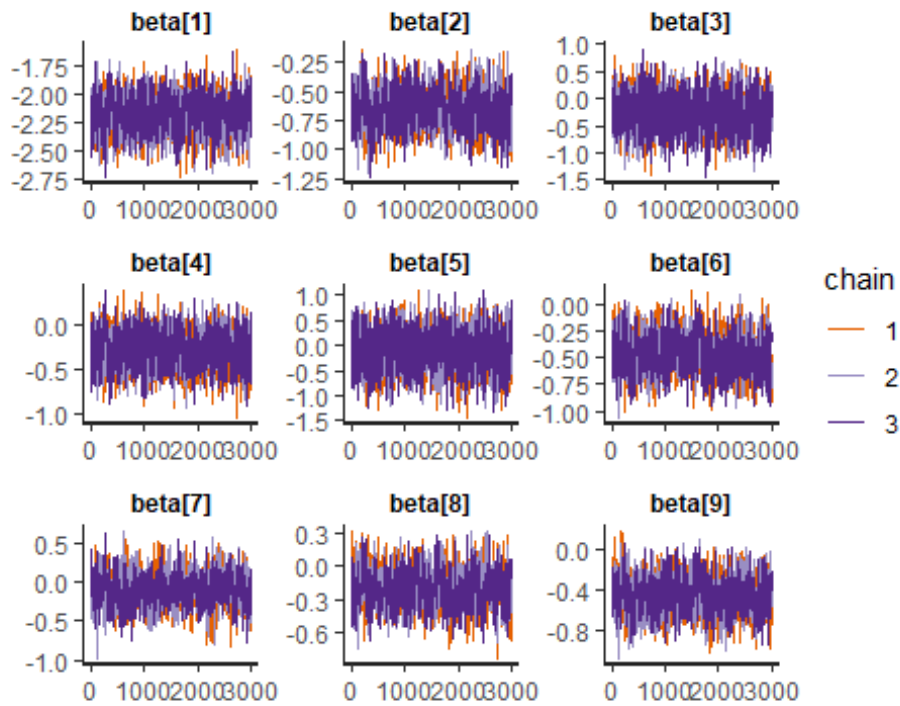
# mod=stan(file="MetaRisk2 RSTAN beta mat.stan",data=stan.data,pars=params.to
.monitor,
#       chains = 3, warmup=18000, cores=8,iter=21000,
#       init = init.fn, save_warmup = FALSE, control=list(adapt_delta = 0.
9, max_treedepth = 15))

load("2taxaL4.rds")
params.to.monitor2=c("beta")#
sumx = summary(mod,probs=c(.025,0.975), digits=4, pars=params.to.monitor2)
sumx$summary

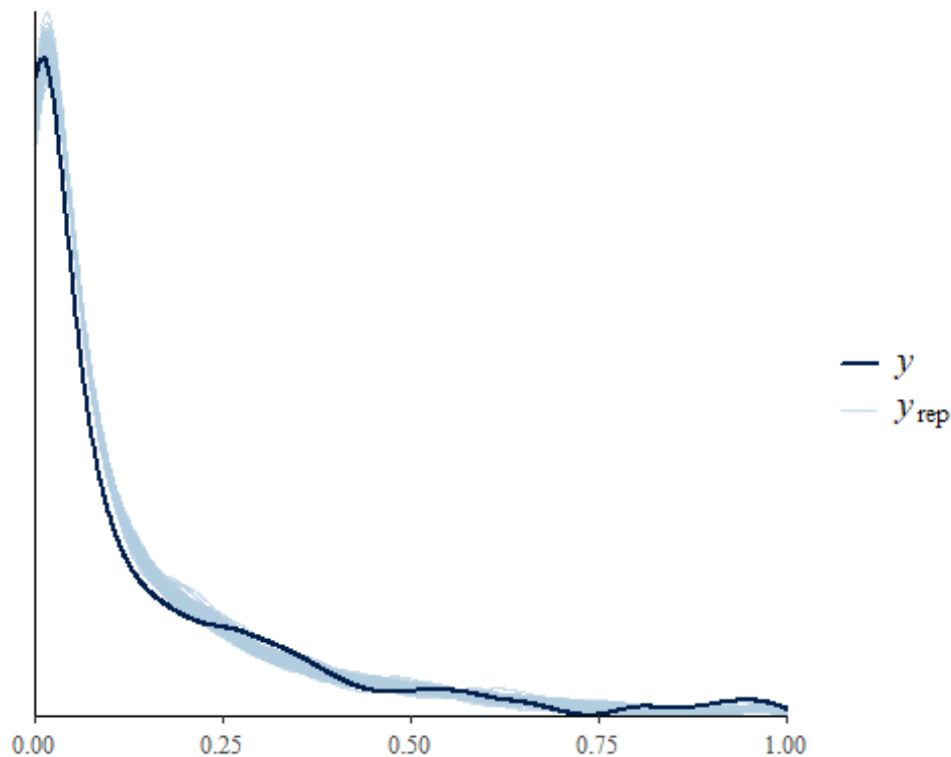
##              mean      se_mean      sd      2.5%      97.5%      n_eff
## beta[1] -2.1891429 0.005818926 0.1557564 -2.4958500 -1.88330293  716.4838
## beta[2] -0.6556791 0.006546553 0.1613068 -0.9715787 -0.33512185  607.1279
## beta[3] -0.2668427 0.008240859 0.3137951 -0.8736203  0.35900645 1449.9316
## beta[4] -0.2846334 0.006594635 0.1889663 -0.6608612  0.07566196  821.0837
## beta[5] -0.1053025 0.008989957 0.3444496 -0.7952416  0.55825747 1468.0339
## beta[6] -0.4646970 0.006658980 0.1693717 -0.7953425 -0.13244809  646.9432
## beta[7] -0.1009387 0.009502242 0.2084218 -0.4976557  0.31057629  481.0987
## beta[8] -0.1996163 0.007506970 0.1654578 -0.5173072  0.12368858  485.7861
## beta[9] -0.4526508 0.006383483 0.1710563 -0.7927669 -0.11490334  718.0633
##              Rhat
## beta[1] 1.002969
## beta[2] 1.003046
## beta[3] 1.000259
## beta[4] 1.002847
## beta[5] 1.000720
## beta[6] 1.001662
## beta[7] 1.005918
## beta[8] 1.009752
## beta[9] 1.001104

#checks
traceplot(mod,pars=params.to.monitor2,inc_warmup=FALSE)

```



```
pp_check(
  stan.data$percent,
  rstan::extract(mod, par = 'y_rep')$y_rep[1:100, ],
  fun = 'dens_overlay'
)
```



```
#calculate loo
# log_lik_1 <- extract_log_lik(mod, merge_chains = FALSE)
# r_eff <- relative_eff(exp(log_lik_1), cores = 6)
# loo.mod <- loo(log_lik_1, r_eff = r_eff, cores = 6)
loo.mod #
```

```
##
## Computed from 9000 by 3235 log-likelihood matrix.
##
##           Estimate      SE
## elpd_loo    7614.6   96.7
## p_loo       1983.3   25.1
## looic       -15229.3 193.5
## -----
## MCSE of elpd_loo is NA.
##
## Pareto k diagnostic values:
##               Count Pct.    Min. ESS
## (-Inf, 0.7]  (good)   1441  44.5%   120
##  (0.7, 1]   (bad)    1531  47.3%   <NA>
##  (1, Inf)   (very bad)  263   8.1%   <NA>
## See help('pareto-k-diagnostic') for details.
```

```
#create data frame of looics from two models
load("2taxaL4.rds")
loo.mod2=loo.mod # rename loo.mod so can load n
mod2 = mod
```

```
load("2all_interc.rds")

table.data<-data.frame(
  Model = c("Intercept-only model","Model including taxonomy"),
  L00ic = c(loo.mod$estimates[3],loo.mod2$estimates[3]),
  SE = c(loo.mod$estimates[6],loo.mod2$estimates[6])
)
knitr::kable(table.data, caption = "Table 1: Comparisons of L00ic between baseline and model with taxonomy", format = "markdown")
```

Table 1: Comparisons of L00ic between baseline and model with taxonomy

Model	L00ic	SE
Intercept-only model	-15213.77	193.3243
Model including taxonomy	-15229.28	193.4593

```
Looic.diff = loo.mod2$estimates[3] - loo.mod$estimates[3]
cat("Difference in L00ic =", Looic.diff)

## Difference in L00ic = -15.51152

#Intercept-only model
load("2all_interc.rds") #intercept only model
mod.int <- mod
loo.int <- loo.mod

#use common definition for global median for graph, or else it varies a bit
posterior2=as.data.frame(mod.int)
grand.mean = posterior2[["mu"]]
grand.mean.pred <- invlogit(quantile(grand.mean, probs = c(0.025, 0.5, 0.975)
))

load("2taxaL4.rds")
modx = mod
# (Intercept) TaxaBirds TaxaFish TaxaInsects TaxaInvertebrates TaxaMammals TaxaMixed TaxaPlants TaxaReptiles
cats <- c("Amphibians", "Birds", "Fish", "Insects", "Invertebrates", "Mammals", "Mixed", "Plants", "Reptiles")
#Calculate estimates; note original is 1 in matrix
posterior=as.data.frame(modx);

n.total <- nrow(data.use) #total N

#absolute risks
beta.0<-posterior[["beta[1]"]]
beta.1<-posterior[["beta[1]"]]+posterior[["beta[2]"]]
beta.2<-posterior[["beta[1]"]]+posterior[["beta[3]"]]
```

```

beta.3<-posterior[["beta[1]"]]+posterior[["beta[4]"]]
beta.4<-posterior[["beta[1]"]]+posterior[["beta[5]"]]
beta.5<-posterior[["beta[1]"]]+posterior[["beta[6]"]]
beta.6<-posterior[["beta[1]"]]+posterior[["beta[7]"]]
beta.7<-posterior[["beta[1]"]]+posterior[["beta[8]"]]
beta.8<-posterior[["beta[1]"]]+posterior[["beta[9]"]]

beta.cat<-cbind(beta.0,beta.1,beta.2,beta.3,beta.4,beta.5,beta.6,beta.7,beta.
8)
#calculate median and credible intervals
pred.cat = invlogit(apply(beta.cat, 2, quantile, probs = c(0.025, 0.5, 0.975)
,na.rm=TRUE))
pred.cat.df <- data.frame(x = cats,
                        mean = pred.cat[2,],
                        low = pred.cat[1,],
                        hi = pred.cat[3,])

#arrange in increasing order#
pred.cat.df$cats = with (pred.cat.df, reorder(cats, mean))

#rbeta.cat = invlogit(beta.cat)-invlogit(grand.mean)
rbeta.cat = invlogit(beta.cat)-grand.mean.pred[2] #use median rather than eac
h sample (compare vs. one value)
rpred.cat = (apply(rbeta.cat, 2, quantile, probs = c(0.025, 0.5, 0.975),na.rm
=TRUE))
rpred.cat.df <- data.frame(x = cats,
                        mean = rpred.cat[2,],
                        low = rpred.cat[1,],
                        hi = rpred.cat[3,])

#Differences from zero
glob.mean.over <- rep("n",nrow(rpred.cat.df))
glob.mean.over[rpred.cat.df$mean > 0 & rpred.cat.df$low > 0] = "y" #decided t
o use grand mean vs. all samples (vs. median value)
glob.mean.over[rpred.cat.df$mean < 0 & rpred.cat.df$hi < 0] = "y"
rpred.cat.df$glob.mean.over <- glob.mean.over

#arrange in increasing order#
rpred.cat.df$cats = with (rpred.cat.df, reorder(cats, mean))

#show values
pred.cat.r <- pred.cat.df[,1:4]
pred.cat.r[,2:4] <- round((pred.cat.r[,2:4]), digits = 4)
knitr::kable(pred.cat.r, caption = "Table 1: Extinction risk by taxa, with 95
% credible intervals", format = "markdown", font = 7)

```

Table 2: Extinction risk by taxa, with 95% credible intervals

	x	mean	low	hi
beta.0	Amphibians	0.1009	0.0761	0.1320
beta.1	Birds	0.0550	0.0446	0.0674
beta.2	Fish	0.0787	0.0471	0.1324
beta.3	Insects	0.0777	0.0604	0.0999
beta.4	Invertebrates	0.0921	0.0515	0.1564
beta.5	Mammals	0.0656	0.0528	0.0817
beta.6	Mixed	0.0917	0.0669	0.1264
beta.7	Plants	0.0841	0.0709	0.0991
beta.8	Reptiles	0.0666	0.0525	0.0834

```
rpred.cat.r <- rpred.cat.df[,1:4]
rpred.cat.r[,2:4] <- round((rpred.cat.r[,2:4]), digits = 4)
knitr::kable(rpred.cat.r, caption = "Table 2: Relative extinction risk by tax
a, with 95% credible intervals", format = "markdown", font = 7)
```

Table 3: Relative extinction risk by taxa, with 95% credible intervals

	x	mean	low	hi
beta.0	Amphibians	0.0250	0.0002	0.0560
beta.1	Birds	-0.0210	-0.0314	-0.0086
beta.2	Fish	0.0027	-0.0288	0.0564
beta.3	Insects	0.0018	-0.0156	0.0239
beta.4	Invertebrates	0.0161	-0.0244	0.0804
beta.5	Mammals	-0.0103	-0.0231	0.0057
beta.6	Mixed	0.0157	-0.0091	0.0504
beta.7	Plants	0.0081	-0.0051	0.0231
beta.8	Reptiles	-0.0094	-0.0235	0.0074

#Figures

```
Fig1a<-ggplot(data = pred.cat.df)+
  geom_vline(xintercept=grand.mean.pred[2]) +
  geom_errorbar(aes(y = cats, xmin = low, xmax = hi), width = 0) +
  geom_point(stat = "identity", aes(y = cats, x = mean), color = "#416788", s
ize = 3, shape = 15) +
  xlab("Pre-industrial rise \n in temperature (C)") + xlim(c(0,.5)) +
```

```

theme_classic()+
  theme(axis.title.y = element_blank(),axis.title=element_text(size=14),axis.
text = element_text(size=12))+
  guides(size=F)

```

Warning: The ``<scale>`` argument of ``guides()`` cannot be `FALSE`. Use "none" instead as
 ## of ggplot2 3.3.4.
 ## This warning is displayed once every 8 hours.
 ## Call ``lifecycle::last_lifecycle_warnings()`` to see where this warning was
 ## generated.

Fig1a

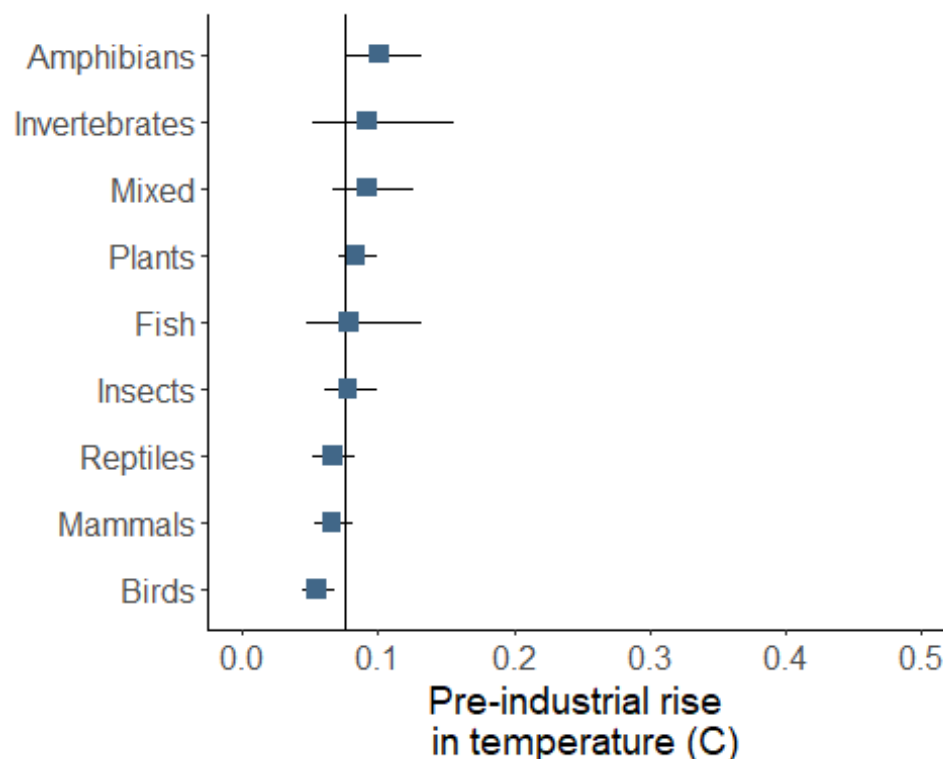


Fig. 1. Predicted extinction risk based on taxonomy

```

Fig1b<-ggplot(data = rpred.cat.df)+
  geom_vline(xintercept=0) +
  geom_errorbar(aes(y = cats, xmin = low, xmax = hi), width = 0) +
  geom_point(stat = "identity", aes(y = cats, x = mean, color = glob.mean.ove
r), size = 3, shape = 15) +
  scale_color_manual(values=c('grey', '#E98a15'))+
  xlab("Percent difference \n from global mean") + xlim(c(-.1,.5)) +
  theme_classic()+
  theme(axis.title.y = element_blank(),axis.ticks.y = element_blank(),axis.li
ne.y = element_blank(),axis.text.y = element_blank(),
        axis.title=element_text(size=14),axis.text = element_text(size=12),le
gend.position = "none")+

```



```
guides(size=F)  
Fig1b
```

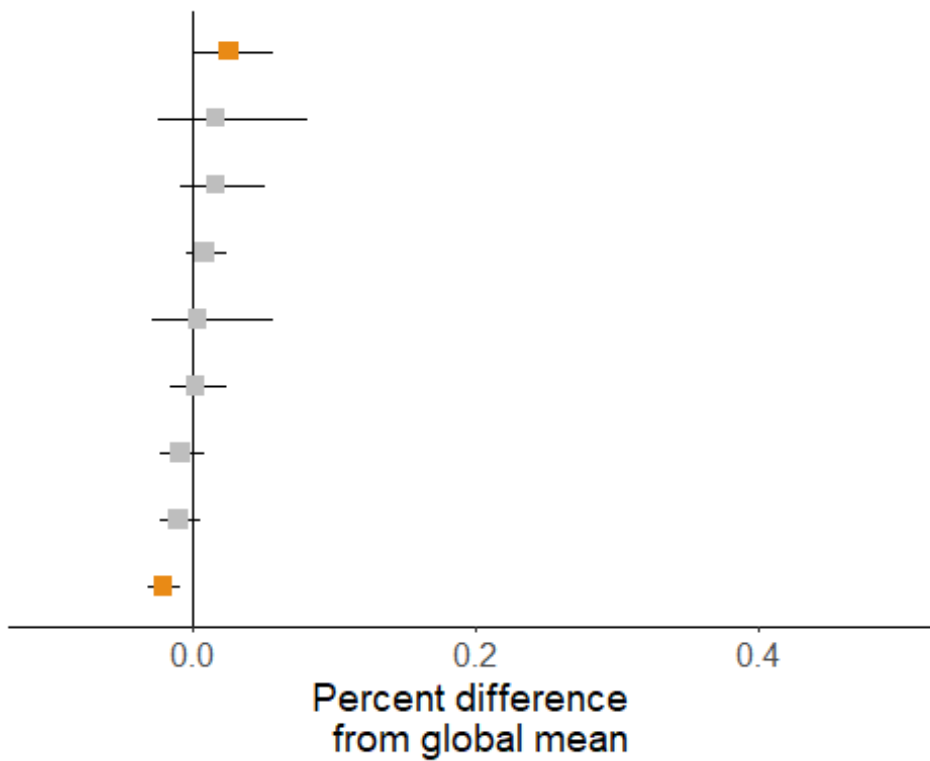


Fig. 2. Predicted extinction risk based on taxonomy

```
ggarrange(Fig1a, NULL, Fig1b, ncol=3, widths = c(4,.4, 2))
```

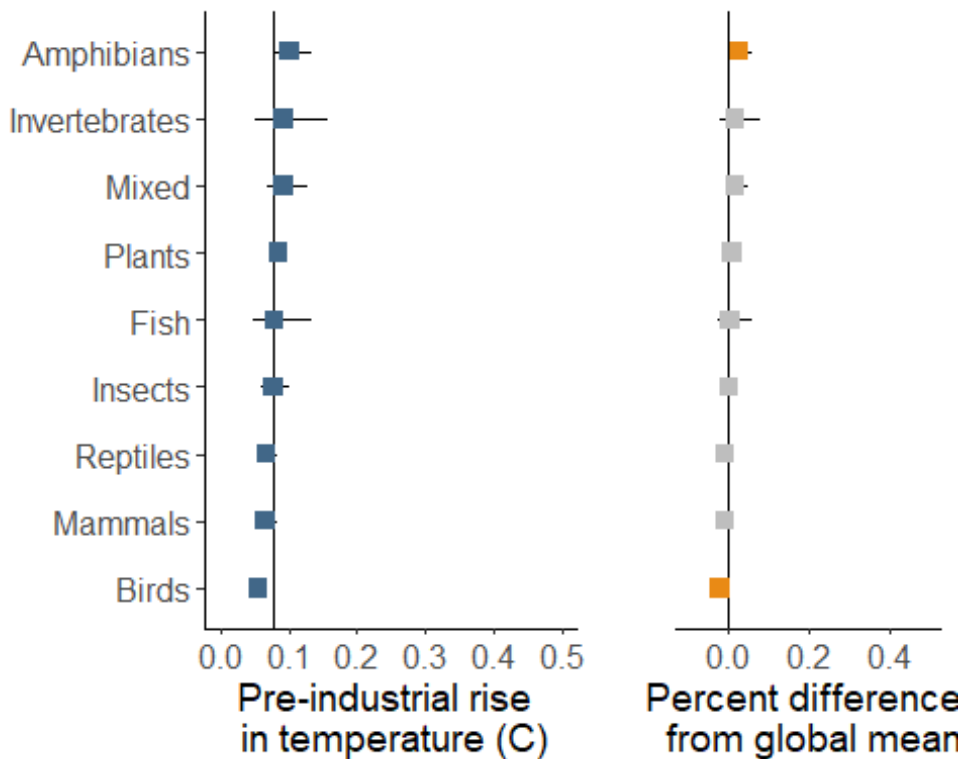


Fig. 3. Predicted extinction risk based on taxonomy

```
#ggsave("Metarisk2_taxa.png",width=4,height=3,unit="in",dpi="print")
```

Conclusion

Taxonomic groups vary in risk, with the highest risks for plants and the lowest risks for birds, reptiles and mammals.

```
#catalog sample sizes
N.1.st<-length(unique(dataP$Study[dataP$Taxa == "Birds"]))
N.1.mod<-length((dataP$Study[dataP$Taxa == "Birds"]))

N.2.st<-length(unique(dataP$Study[dataP$Taxa == "Fish"]))
N.2.mod<-length((dataP$Study[dataP$Taxa == "Fish"]))

N.3.st<-length(unique(dataP$Study[dataP$Taxa == "Insects"]))
N.3.mod<-length((dataP$Study[dataP$Taxa == "Insects"]))

N.4.st<-length(unique(dataP$Study[dataP$Taxa == "Invertebrates"]))
N.4.mod<-length((dataP$Study[dataP$Taxa == "Invertebrates"]))

N.5.st<-length(unique(dataP$Study[dataP$Taxa == "Mammals"]))
N.5.mod<-length((dataP$Study[dataP$Taxa == "Mammals"]))

N.6.st<-length(unique(dataP$Study[dataP$Taxa == "Mixed"]))
N.6.mod<-length((dataP$Study[dataP$Taxa == "Mixed"]))
```

```

N.7.st<-length(unique(dataP$Study[dataP$Taxa == "Plants"]))
N.7.mod<-length((dataP$Study[dataP$Taxa == "Plants"]))

N.8.st<-length(unique(dataP$Study[dataP$Taxa == "Reptiles"]))
N.8.mod<-length((dataP$Study[dataP$Taxa == "Reptiles"]))

N.9.st<-length(unique(dataP$Study[dataP$Taxa == "Amphibians"]))
N.9.mod<-length((dataP$Study[dataP$Taxa == "Amphibians"]))

table.data<-data.frame(
  Factor = c("Birds","Fish","Insects","Invertebrates","Mammals","Mixed","Plants",
    "Reptiles","Amphibians"),
  Studies = c(N.1.st,N.2.st,N.3.st,N.4.st,N.5.st,N.6.st,N.7.st,N.8.st,N.9.st)
  ,
  Models = c(N.1.mod,N.2.mod,N.3.mod,N.4.mod,N.5.mod,N.6.mod,N.7.mod,N.8.mod,
    N.9.mod)
)
knitr::kable(table.data, caption = "Table 2: Number of studies and models for
each factor", format = "markdown")

```

Table 3: Number of studies and models for each factor

Factor	Studies	Models
Birds	92	536
Fish	24	68
Insects	68	389
Invertebrates	21	57
Mammals	75	333
Mixed	14	167
Plants	199	1177
Reptiles	48	275
Amphibians	43	233

Variation explained

```

#After Gelman 2019 R2 for Bayesian
#
#Load model and beta matrix - check if mu is modeled separately
load("2taxaL4.rds")
posterior=as.data.frame(mod);
betamat <- (model.matrix(~Taxa,data=data.use))

#Variables and matrices
S = 9000; #samples

```

```

K = ncol(betamat); #factors
p.mat <- as.matrix(posterior[,1:K])
y = dataP$percent2
y.mat = t(matrix(rep(y,S), nrow = N, ncol = S))
y.mean <- mean(y)

#Calculate y.pred for fixed effects only
y.pred <- matrix(rep(NA, N*S), nrow = S, ncol = N)
theta <- y.pred
for (i in 1:N) {
  theta[,i] = invlogit(p.mat %*% betamat[i,])#rows = samples, cols = i
  y.pred[,i] = (theta[,i] * data.use$Total.N[i])/(theta[,i] * data.use$Total.N[i] + (1-theta[,i]) * data.use$Total.N[i])
}

#Calcluate residual variance
res.f = y.mat - y.pred
RSS.f = rowSums((res.f)^2)
res.v.f = 1/(N -1) * RSS.f

#Calculate fit variance
pred.v.f = 1/(N-1) * rowSums((y.pred)^2)

#Calculate R2
R2.v.f = pred.v.f/(pred.v.f + res.v.f)
cat("fixed effects R2 = ", quantile(R2.v.f,probs = c(0.025, 0.5, 0.975),na.rm = T))

## fixed effects R2 = 0.0653845 0.08645069 0.1135329

#Total model With random effects
y.pred.c <-(as.matrix(posterior[, (K+1):(N+K)])) #calculated in STAN, with all RE and weightings

#Calculate residual variance
res.c = y.mat - y.pred.c
RSS.c = rowSums((res.c)^2)
res.v.c = 1/(N-1) * RSS.c

#Calculate fit variance

pred.v.c = 1/(N-1) * rowSums(y.pred.c^2)

#Calculate full model R2
R2.v.c = pred.v.c/(pred.v.c + res.v.c)
print("
## [1] "

```

```
cat("Overall model R2 = ", quantile(R2.v.c, probs = c(0.025, 0.5, 0.975), na.rm  
= T))  
## Overall model R2 = 0.764939 0.7861843 0.8065545
```