Factor-weighted estimate of extinction risk from climate change

March 25, 2024

## Load libraries and data

knitr::opts\_chunk$set(echo = TRUE, cache.lazy = FALSE, cache = TRUE)   
rm(list = ls())  
 root.dir = "C:/Users/mcu08001/Documents/1New Research/CC MetaRisk2/Analysis"  
library(MCMCglmm); library(coda); library(ggplot2); library(rstan); library(bayesplot); library(shinystan); library(loo); library(rstanarm);  
  
dataP<-read.table("Metarisk2 aggthres 5.txt",header=T); #newest data   
dataP2<-dataP[is.finite(dataP$Pre.Ind.Rise),]; attach(dataP2) # need to eliminate NA s for pre-industrial rise or stat programs crash  
  
#Bayesian stan model proportional and weighted  
#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;  
dataP2$percent2 <- percent2;  
  
data.use<-dataP2  
  
#other quantities  
P.Ind<-seq(from = 0, to = 5.5,by = .1) #prediction interval and frequency  
P.Ind2<-seq(from = 0, to = 5.5,by = .01) #prediction interval and frequency needed for Fig 1b  
N = length(data.use$percent2)  
n.Study <- length(unique(data.use$Study)) #number of studies  
betamat <- (model.matrix(~Pre.Ind.Rise:Region,data=data.use))

## Load analyses

#load baseline overall analysis  
load("2pre\_lowb.rds") #baseline  
base<-mod;  
  
#Calculate estimates; note original is 1 in matrix  
posteriorb=as.data.frame(base);  
mu.b<-posteriorb[["mu"]]  
beta.b1<-posteriorb[["beta"]]  
params.to.monitor2=c("mu","beta")#  
sumx = summary(base,probs=c(.025,0.975), digits=4, pars=params.to.monitor2)   
sumx$summary

## mean se\_mean sd 2.5% 97.5% n\_eff Rhat  
## mu -4.5273890 0.003521920 0.10455352 -4.7373329 -4.3237051 881.289 1.004510  
## beta 0.6457854 0.000897318 0.02389391 0.5995963 0.6929307 709.058 1.003601

### Load continent analysis

# dataP<-read.table("MetaRisk2 aggthres 5.txt",header=T);   
# dataP2<-dataP[is.finite(dataP$Pre.Ind.Rise),]; attach(dataP2) # need to eliminate NA s for pre-industrial rise or stat programs crash  
#   
# #betareg requires no 0s or 1s  
# koffset = 0.001 #the k that gives the best posterior predictive check  
# percent2 <- adj.percent  
# percent2[adj.percent == 0] = koffset;  
# percent2[adj.percent == 1] = 1 - koffset;  
# dataP2$percent2 <- percent2;  
#   
# data.use<-dataP2  
#   
# 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  
#   
# #create model matrix for coefficients  
# betamat <- (model.matrix(~Pre.Ind.Rise:Region,data=data.use))   
# betamat.noint <- betamat[,2:ncol(betamat)] #remove intercept and model separately  
# stan.data<-list(N = N, percent = data.use$percent2, betamat = betamat.noint, phi = phi, S = n.Study, P = ncol(betamat.noint), Study = Studyint)  
#   
# params.to.monitor=c("mu","beta","y\_rep","stu","sigma\_stu", "eta","log\_lik")  
#   
# init.fn<- function (chain\_id) {  
# list(beta = c(-5, rep(0.5,ncol(betamat))))  
# }  
  
# mod=stan(file="MetaRisk2 RSTAN quad.stan",data=stan.data,pars=params.to.monitor,  
# chains = 3, warmup=18000, cores=3,iter=21000,  
# init = init.fn, save\_warmup = FALSE, control=list(adapt\_delta = 0.9, max\_treedepth = 20))#  
  
  
#load continent analysis  
load("2geog\_pre.rds")   
cont<-mod;  
posterior=as.data.frame(cont);  
cats <- c("Africa", "Asia", "New Zealand/Australia", "Europe", "North America", "South America")  
  
params.to.monitor2=c("beta")#params.to.monitor2=c("mu","beta")#  
sumx = summary(cont,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] 0.6358940 0.0015289254 0.03823974 0.5599226 0.7086737 625.5433  
## beta[2] 0.6103164 0.0018982498 0.04013769 0.5298153 0.6878735 447.0931  
## beta[3] 0.9432687 0.0012838904 0.03658037 0.8716737 1.0160333 811.7836  
## beta[4] 0.7122551 0.0011325771 0.04256791 0.6288605 0.7943699 1412.6324  
## beta[5] 0.5124336 0.0014818006 0.06289632 0.3880057 0.6344465 1801.6520  
## beta[6] 0.7120774 0.0009844191 0.04498400 0.6253443 0.8004154 2088.1227  
## beta[7] 0.9551339 0.0015785504 0.06446730 0.8272487 1.0821722 1667.8698  
## beta[8] 0.7141942 0.0016215450 0.03484533 0.6456767 0.7823601 461.7758  
## Rhat  
## beta[1] 1.000514  
## beta[2] 1.003200  
## beta[3] 1.001252  
## beta[4] 1.004649  
## beta[5] 1.001665  
## beta[6] 1.001767  
## beta[7] 1.002239  
## beta[8] 1.013693

#Calculate estimates; note original is 1 in matrix  
mu<-posterior[["mu"]]  
beta.1<-posterior[["beta[1]"]]  
beta.2<-posterior[["beta[2]"]]  
beta.3<-posterior[["beta[3]"]]  
beta.4<-posterior[["beta[4]"]]  
#beta.5<-posterior[["beta[5]"]] #oceans not treated here  
beta.6<-posterior[["beta[6]"]]  
beta.7<-posterior[["beta[7]"]] #no beta 8 which is world (mixed)  
  
beta.cat<-cbind(beta.1,beta.2,beta.3,beta.4,beta.6,beta.7)

### Perform weighting

#Preliminary weighting based on trees (Gatti el. 2022, using observed https://www.pnas.org/doi/epdf/10.1073/pnas.2115329119)  
tree.rich <- c(10441,7035,6680,7035,8646,27186) #modeled eurasia so split species evenly between Asia and Europe; used n species from Oceania for Australia/New Zealand  
vert.rich <- c(4646 ,6597,1973,550, 4085,5620)#amphibians, mammals, reptiles from IUCN ranges  
# africa asia australia oceania south\_america europe north\_america  
# 1132 1603 222 73 2369 94 1180  
# 1364 1926 271 113 1202 260 923  
# 2150 3068 961 333 2049 196 1982  
# 4646 6597 1454 519 5620 550 4085  
cont.rich = tree.rich + vert.rich;  
rel.rich <- cont.rich/sum(cont.rich);  
rel.cont.richm = matrix(rep(rel.rich,9000), nrow = 9000, byrow = T)  
  
cont.rich.db <- data.frame(  
 Continents = cats,  
 Total.Rich = cont.rich,  
 Rel.Rich = rel.rich  
)  
  
knitr::kable(cont.rich.db, caption = "Table 1: Continental richness", format = "markdown", font = 8)

Table 1: Continental richness

| Continents | Total.Rich | Rel.Rich |
| --- | --- | --- |
| Africa | 15087 | 0.1667182 |
| Asia | 13632 | 0.1506398 |
| New Zealand/Australia | 8653 | 0.0956196 |
| Europe | 7585 | 0.0838177 |
| North America | 12731 | 0.1406834 |
| South America | 32806 | 0.3625213 |
|  |  |  |

beta.cont = beta.cat \* rel.cont.richm #weighted betas  
beta.cont.w = rowSums(beta.cont) #sum to get weighted coefficient (works because all relative to 1)  
  
diff.betas <- quantile(invlogit(beta.cont.w) - invlogit(beta.b1), probs = c(0.025, 0.5, 0.975))  
diff.mus <- quantile(invlogit(mu) - invlogit(mu.b), probs = c(0.025, 0.5, 0.975))  
cat("Difference in betas =",diff.betas)

## Difference in betas = 0.01525872 0.03270342 0.04953143

print("")

## [1] ""

cat("Difference in intercepts =", diff.mus)

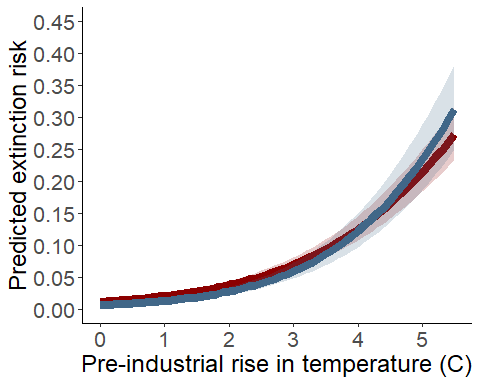
## Difference in intercepts = -0.00765267 -0.00492765 -0.002364435

## Results

The model weighted by continental richness was characterized by a larger slope and smaller intercept.

#Calculate estimates  
pred.reg = sapply(1:length(mu.b), FUN = function(x) {mu.b[x] + beta.b1[x]\*P.Ind}) #mu[x]  
pred.reg.quant = invlogit(apply(pred.reg, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df <- data.frame(x = P.Ind,  
 mean\_line = pred.reg.quant[2,],  
 low\_line = pred.reg.quant[1,],  
 hi\_line = pred.reg.quant[3,])  
  
pred.regw = sapply(1:length(beta.1), FUN = function(x) {mu[x] + beta.cont.w[x]\*P.Ind}) #mu[x]  
pred.reg.quantw = invlogit(apply(pred.regw, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df$mean\_linew = pred.reg.quantw[2,]  
pred.reg.df$low\_linew = pred.reg.quantw[1,]  
pred.reg.df$hi\_linew = pred.reg.quantw[3,]  
  
#----finer data ------#  
pred.reg2 = sapply(1:length(mu.b), FUN = function(x) {mu.b[x] + beta.b1[x]\*P.Ind2}) #mu[x]  
pred.reg.quant2 = invlogit(apply(pred.reg2, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df2 <- data.frame(x = P.Ind2,  
 mean\_line = pred.reg.quant2[2,],  
 low\_line = pred.reg.quant2[1,],  
 hi\_line = pred.reg.quant2[3,])  
  
   
pred.regw2 = sapply(1:length(beta.1), FUN = function(x) {mu[x] + beta.cont.w[x]\*P.Ind2}) #mu[x]  
pred.reg.quantw2 = invlogit(apply(pred.regw2, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df2$mean\_linew = pred.reg.quantw2[2,]  
pred.reg.df2$low\_linew = pred.reg.quantw2[1,]  
pred.reg.df2$hi\_linew = pred.reg.quantw2[3,]

Fig1<-ggplot(data = pred.reg.df)+  
 geom\_ribbon(data = pred.reg.df, aes(x=P.Ind,ymin=low\_line,ymax=hi\_line),alpha=.2,fill="darkred") +  
 geom\_line(data = pred.reg.df, aes(x=P.Ind,y=mean\_line),size=3,color="darkred") +  
 geom\_ribbon(data = pred.reg.df, aes(x=P.Ind,ymin=low\_linew,ymax=hi\_linew),alpha=.2,fill="#416788") +  
 geom\_line(data = pred.reg.df, aes(x=P.Ind,y=mean\_linew),size=3,color="#416788") +  
 xlab("Pre-industrial rise in temperature (C)") + ylab("Predicted extinction risk") +  
 theme\_classic() + scale\_x\_continuous(breaks = seq(0,5,1), limits = c(0,5.5)) + scale\_y\_continuous(breaks = seq(0,.45,0.05), limits = c(0,.45)) +  
 theme(axis.title=element\_text(size=18),title=element\_text(size=20),axis.text = element\_text(size=16))+  
 guides(size=F)  
Fig1



#ggsave("Fig contwtd preind.png",width=8,height=6,unit="in",dpi="print")

### Despite differences in intercepts and slopes, the two estimates overlap acros the range of values and 95% credible intervals incorporate the alternative estimate. The bias-corrected estimate has larger credible intervals, likely reflecting the higher uncertainties in more diverse regions such as South America.

## Now look at taxonomic richness

# #create model matrix for coefficients  
# betamat <- (model.matrix(~Pre.Ind.Rise:Taxa,data=data.use))   
# betamat.noint <- betamat[,2:ncol(betamat)] #remove intercept and model separately  
# stan.data<-list(N = N, percent = data.use$percent2, betamat = betamat.noint, phi = phi, S = n.Study, P = ncol(betamat.noint), Study = Studyint)  
#   
# params.to.monitor=c("mu","beta","y\_rep","stu","sigma\_stu", "eta","log\_lik")  
#   
# init.fn<- function (chain\_id) {  
# list(beta = c(-5, rep(0.5,ncol(betamat))))  
# }  
#   
# mod=stan(file="MetaRisk2 RSTAN quad.stan",data=stan.data,pars=params.to.monitor,  
# chains = 3, warmup=18000, cores=3,iter=21000,  
# init = init.fn, save\_warmup = FALSE, control=list(adapt\_delta = 0.9, max\_treedepth = 20))#  
# save(mod, loo.mod, file="2taxa\_pre.rds")  
#load taxa analysis  
load("2taxa\_pre2.rds")   
tax<-mod;  
tposterior=as.data.frame(tax);  
tcats <- c("Amphibians", "Birds", "Fish", "Insects", "Invertebrates", "Mammals", "Plants", "Reptiles")  
  
params.to.monitor2=c("mu","beta")#  
sumx = summary(tax,probs=c(.025,0.975), digits=4, pars=params.to.monitor2)   
sumx$summary

## mean se\_mean sd 2.5% 97.5% n\_eff  
## mu -5.2161475 0.004949360 0.12552569 -5.4670615 -4.9803349 643.2313  
## beta[1] 0.8543885 0.001253540 0.05009772 0.7590125 0.9514321 1597.2011  
## beta[2] 0.6364515 0.001384346 0.03256647 0.5718572 0.7012766 553.4169  
## beta[3] 0.8564201 0.001925321 0.09343761 0.6732510 1.0407770 2355.2497  
## beta[4] 0.8127695 0.001025904 0.04394258 0.7255237 0.8981847 1834.6688  
## beta[5] 0.8237923 0.002049476 0.11011514 0.6060028 1.0377418 2886.7454  
## beta[6] 0.6609862 0.001214824 0.03527579 0.5910316 0.7292577 843.1922  
## beta[7] 0.6857753 0.001763475 0.04904259 0.5900207 0.7817384 773.4067  
## beta[8] 0.7736272 0.001335976 0.03000699 0.7162204 0.8337531 504.4836  
## beta[9] 0.6551610 0.001457175 0.03827581 0.5794529 0.7313315 689.9625  
## Rhat  
## mu 1.004307  
## beta[1] 1.004477  
## beta[2] 1.004324  
## beta[3] 1.001962  
## beta[4] 1.003112  
## beta[5] 1.001021  
## beta[6] 1.004522  
## beta[7] 1.003184  
## beta[8] 1.005233  
## beta[9] 1.004897

#Calculate estimates; note original is 1 in matrix  
tmu<-tposterior[["mu"]]  
tbeta.1<-tposterior[["beta[1]"]] #amphibs  
tbeta.2<-tposterior[["beta[2]"]] #Birds  
tbeta.3<-tposterior[["beta[3]"]] #Fish  
tbeta.4<-tposterior[["beta[4]"]] #Insects  
tbeta.5<-tposterior[["beta[5]"]] #Inverts  
tbeta.6<-tposterior[["beta[6]"]] #Mammals  
tbeta.7<-tposterior[["beta[7]"]] #Mixed  
tbeta.8<-tposterior[["beta[8]"]] #Plants  
tbeta.9<-tposterior[["beta[9]"]] #Reptiles  
  
tbeta.cat<-cbind(tbeta.1,tbeta.2,tbeta.3,tbeta.4,tbeta.5,tbeta.6,tbeta.8,tbeta.9) # missing tbeta.7 = mixed taxa, not informative

# "Amphibians", "Birds", "Fish", "Insects", "Invertebrates", "Mammals", "Plants", "Reptiles"  
taxa.rich <- c(6515,9990,31269,1000000,359365,5487,297857,8734)  
rel.taxa.rich <- taxa.rich/sum(taxa.rich);  
rel.taxa.richm = matrix(rep(rel.taxa.rich,9000), nrow = 9000, byrow = T)  
  
taxa.rich.db <- data.frame(  
 taxa = tcats,  
 Total.Rich = taxa.rich,  
 Rel.Rich = rel.taxa.rich  
)  
  
knitr::kable(taxa.rich.db, caption = "Table 2: Taxa richness", format = "markdown", font = 8)

Table 2: Taxa richness

| taxa | Total.Rich | Rel.Rich |
| --- | --- | --- |
| Amphibians | 6515 | 0.0037895 |
| Birds | 9990 | 0.0058108 |
| Fish | 31269 | 0.0181879 |
| Insects | 1000000 | 0.5816601 |
| Invertebrates | 359365 | 0.2090283 |
| Mammals | 5487 | 0.0031916 |
| Plants | 297857 | 0.1732515 |
| Reptiles | 8734 | 0.0050802 |
|  |  |  |

beta.tax = tbeta.cat \* rel.taxa.richm #weighted betas  
beta.tax.w = rowSums(beta.tax) #sum to get weighted coefficient (works because all relative to 1)  
  
diff.tbetas <- quantile(invlogit(beta.tax.w) - invlogit(beta.b1), probs = c(0.025, 0.5, 0.975))  
diff.tmus <- quantile(invlogit(tmu) - invlogit(mu.b), probs = c(0.025, 0.5, 0.975))  
cat("Taxa: Difference in betas =",diff.tbetas)

## Taxa: Difference in betas = 0.01572937 0.0353082 0.0551147

print(",")

## [1] ","

cat("Taxa: Difference in intercepts =", diff.tmus)

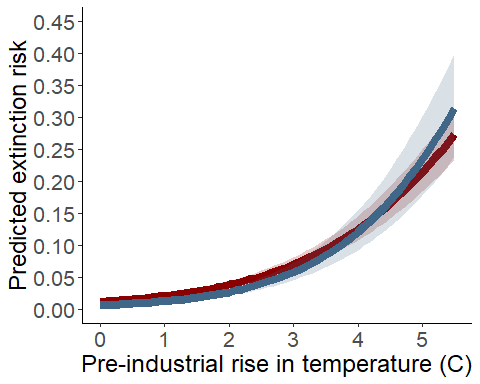
## Taxa: Difference in intercepts = -0.007936108 -0.005296114 -0.002771574

## Results

Again, the weighted estimate has higher beta but lower intercept.

tpred.regw = sapply(1:length(tbeta.1), FUN = function(x) {tmu[x] + beta.tax.w[x]\*P.Ind}) #mu[x]  
tpred.reg.quantw = invlogit(apply(tpred.regw, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df$mean\_linet = tpred.reg.quantw[2,]  
pred.reg.df$low\_linet = tpred.reg.quantw[1,]  
pred.reg.df$hi\_linet = tpred.reg.quantw[3,]  
  
tpred.regw2 = sapply(1:length(tbeta.1), FUN = function(x) {tmu[x] + beta.tax.w[x]\*P.Ind2}) #mu[x]  
tpred.reg.quantw2 = invlogit(apply(tpred.regw2, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df2$mean\_linet = tpred.reg.quantw2[2,]  
pred.reg.df2$low\_linet = tpred.reg.quantw2[1,]  
pred.reg.df2$hi\_linet = tpred.reg.quantw2[3,]

Fig2<-ggplot(data = pred.reg.df)+  
 geom\_ribbon(data = pred.reg.df, aes(x=P.Ind,ymin=low\_line,ymax=hi\_line),alpha=.2,fill="darkred") +  
 geom\_line(data = pred.reg.df, aes(x=P.Ind,y=mean\_line),size=3,color="darkred") +  
 geom\_ribbon(data = pred.reg.df, aes(x=P.Ind,ymin=low\_linet,ymax=hi\_linet),alpha=.2,fill="#416788") +  
 geom\_line(data = pred.reg.df, aes(x=P.Ind,y=mean\_linet),size=3,color="#416788") +  
 xlab("Pre-industrial rise in temperature (C)") + ylab("Predicted extinction risk") +  
 theme\_classic() + scale\_x\_continuous(breaks = seq(0,5,1), limits = c(0,5.5)) + scale\_y\_continuous(breaks = seq(0,.45,0.05), limits = c(0,.45)) +  
 theme(axis.title=element\_text(size=18),title=element\_text(size=20),axis.text = element\_text(size=16))+  
 guides(size=F)  
Fig2

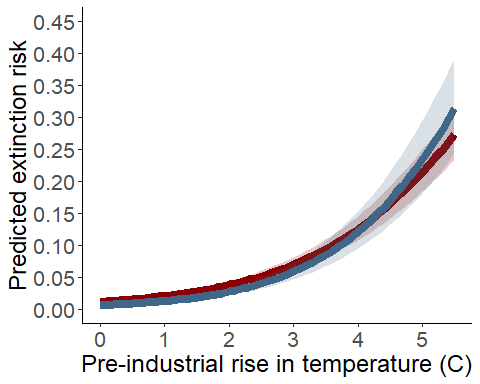


#ggsave("Fig taxawtd preind.png",width=8,height=6,unit="in",dpi="print")

## Combine the two estimates into ensemble

#Combine both estimates  
tg.predw <- cbind(pred.regw,tpred.regw)  
tg.reg.quantw = invlogit(apply(tg.predw, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df$mean\_linetg = tg.reg.quantw[2,]  
pred.reg.df$low\_linetg = tg.reg.quantw[1,]  
pred.reg.df$hi\_linetg = tg.reg.quantw[3,]  
  
tg.predw2 <- cbind(pred.regw2,tpred.regw2)  
tg.reg.quantw2 = invlogit(apply(tg.predw2, 1, quantile, probs = c(0.025, 0.5, 0.975),na.rm=TRUE))  
pred.reg.df2$mean\_linetg = tg.reg.quantw2[2,]  
pred.reg.df2$low\_linetg = tg.reg.quantw2[1,]  
pred.reg.df2$hi\_linetg = tg.reg.quantw2[3,]  
  
#save(pred.reg.df,pred.reg.df2,file = "preind Weighted estimate.rds")

Fig3<-ggplot(data = pred.reg.df)+  
 geom\_ribbon(data = pred.reg.df, aes(x=P.Ind,ymin=low\_line,ymax=hi\_line),alpha=.2,fill="darkred") +  
 geom\_line(data = pred.reg.df, aes(x=P.Ind,y=mean\_line),size=3,color="darkred") +  
 geom\_ribbon(data = pred.reg.df, aes(x=P.Ind,ymin=low\_linetg,ymax=hi\_linetg),alpha=.2,fill="#416788") +  
 geom\_line(data = pred.reg.df, aes(x=P.Ind,y=mean\_linetg),size=3,color="#416788") +  
 xlab("Pre-industrial rise in temperature (C)") + ylab("Predicted extinction risk") +  
 theme\_classic() + scale\_x\_continuous(breaks = seq(0,5,1), limits = c(0,5.5)) + scale\_y\_continuous(breaks = seq(0,.45,0.05), limits = c(0,.45)) +  
 theme(axis.title=element\_text(size=18),title=element\_text(size=20),axis.text = element\_text(size=16))+  
 guides(size=F)  
Fig3



#ggsave("Fig bothwtd preind.png",width=8,height=6,unit="in",dpi="print")