Extinction risk from climate change: geography

Mar. 25, 2024

Load libraries and data

```r
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)

## The following object is masked from package:base:
##
##     version

#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;
dataP$percent2 <- percent2;
```

Intercept-only model

First need to calculate the intercept-only model (mean) to compare against categorical models.

```r
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

#create model matrix for coefficients
stan.data<-list(N = N, percent = data.use$percent2, phi = phi, S = n.Study, S
```

```
tudy = Studyint)
params.to.monitor=c("mu","y_rep","study_u", "psi","eta","log_lik")

init.fn<- function (chain_id) {
  list(mu = -2.5)
}
load("2all_interc.rds") # includes all data *dataP*
# mod=stan(file="MetaRisk2 RSTAN int only 1.stan",data=stan.data,pars=params.
to.monitor,
#         chains = 3, warmup=5000, cores=3,iter=8000, save_warmup = FALSE,
#         init = init.fn, control=list(adapt_delta = 0.9, max_treedepth = 15
))

params.to.monitor2=c("mu")#
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      Rhat
## mu -2.500338 0.002887489 0.07610981 -2.654984 -2.355588 694.7693 1.001517

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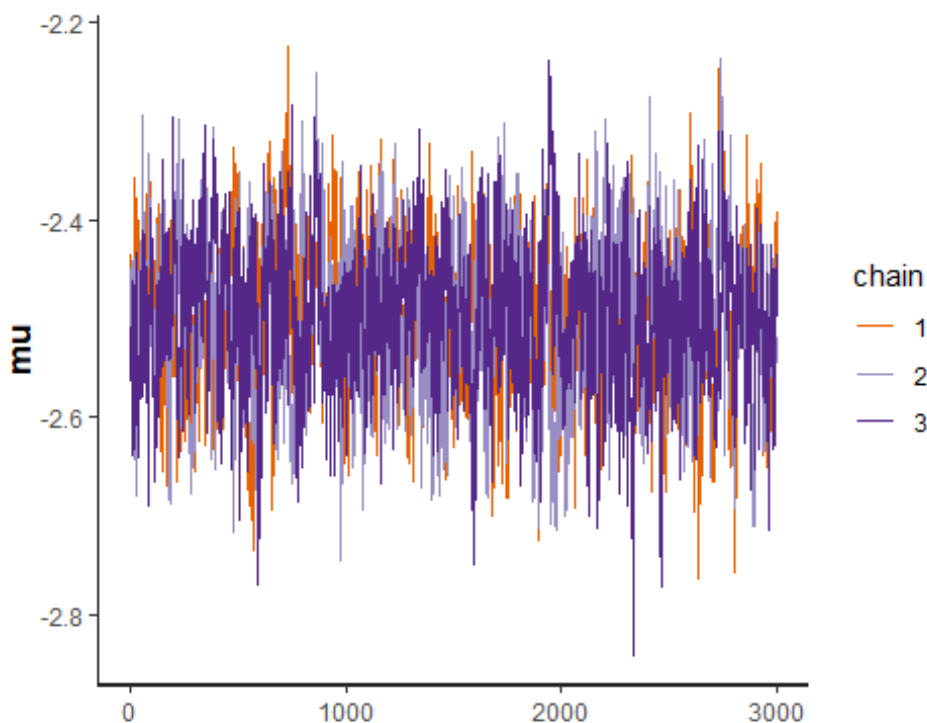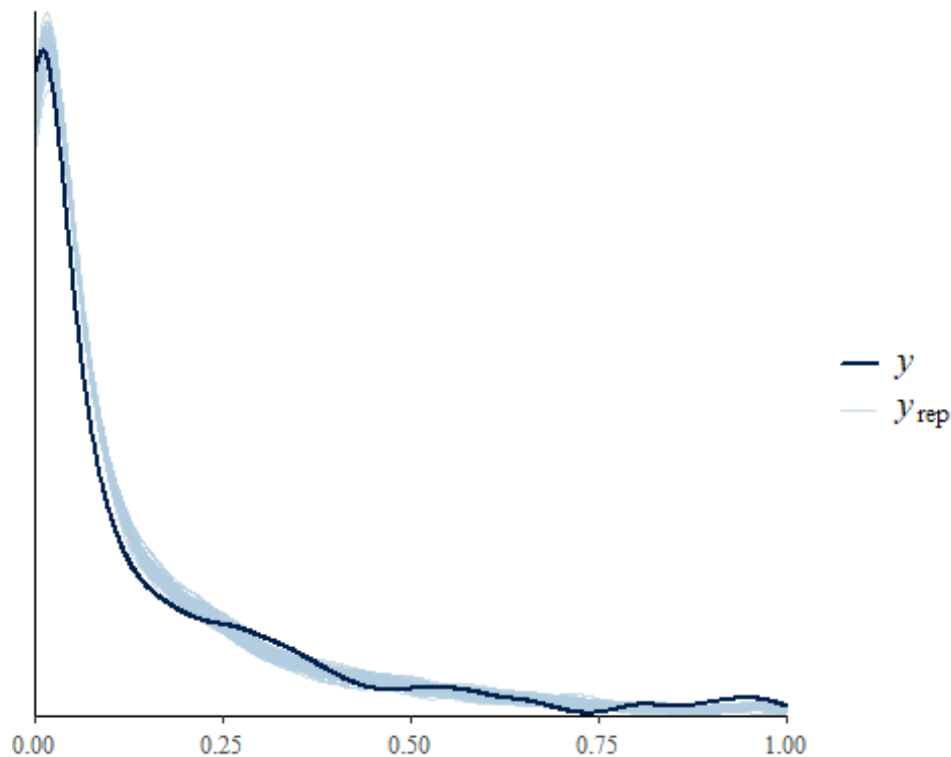


```
pp_check(
  stan.data$percent,
  rstan::extract(mod, par = 'y_rep')$y_rep[1:100, ],
```

```
  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   7606.9  96.7
## p_loo      1990.7  25.1
## looic    -15213.8 193.3
## ------
## MCSE of elpd_loo is NA.
## 
## Pareto k diagnostic values:
##                        Count Pct.    Min. ESS
## (-Inf, 0.7]   (good)    1408  43.5%   107
##    (0.7, 1]   (bad)     1580  48.8%   <NA>
##    (1, Inf)   (very bad) 247   7.6%   <NA>
## See help('pareto-k-diagnostic') for details.
```

Test for effect of Region

Here I tested if extinction predictions change based on region - represented by continent and including marine habitats.

```r
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

#create model matrix for coefficients
betamat <- (model.matrix(~Region,data=data.use, contrasts.arg=list(Region="co
ntr.sum"))) #create model matrix 1-Africa, 2-Asia, 3-New Zealand/Australia, 4
-Europe, 5-marine, 6-N.America, 7-S.America, 8-world
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.fn<- function (chain_id) {
  list(beta = c(-2.5, rep(0,ncol(betamat)-1)))
}

# mod=stan(file="MetaRisk2 RSTAN beta mat.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("2geog5.rds") #previous version had divergences, this one run 18000 - 21
000
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] -2.4593080 0.002446096 0.07798388 -2.6135710 -2.30947271 1016.3953
## beta[2] -0.1912490 0.007496676 0.12131568 -0.4281321  0.04397541  261.8764
## beta[3] -0.3751275 0.005617334 0.11277391 -0.5969375 -0.14716383  403.0480
## beta[4]  0.7717284 0.005306964 0.11610710  0.5399118  0.99446831  478.6579
## beta[5] -0.2421157 0.003910200 0.13257187 -0.5028510  0.01476373 1149.4888
## beta[6] -0.2700705 0.007028646 0.27631256 -0.8090136  0.27329845 1545.4604
## beta[7] -0.2407106 0.004193293 0.13719256 -0.5107548  0.02505386 1070.4113
## beta[8]  0.5442511 0.004941583 0.16817678  0.2085484  0.86642240 1158.2434
##             Rhat
## beta[1] 1.004860
## beta[2] 1.003693
## beta[3] 1.009072
```
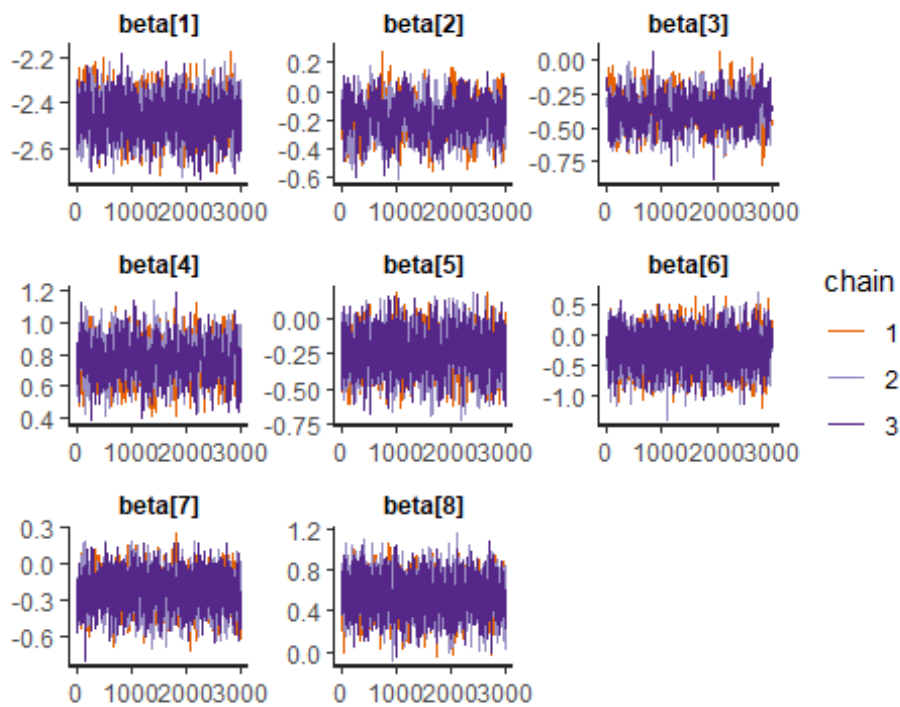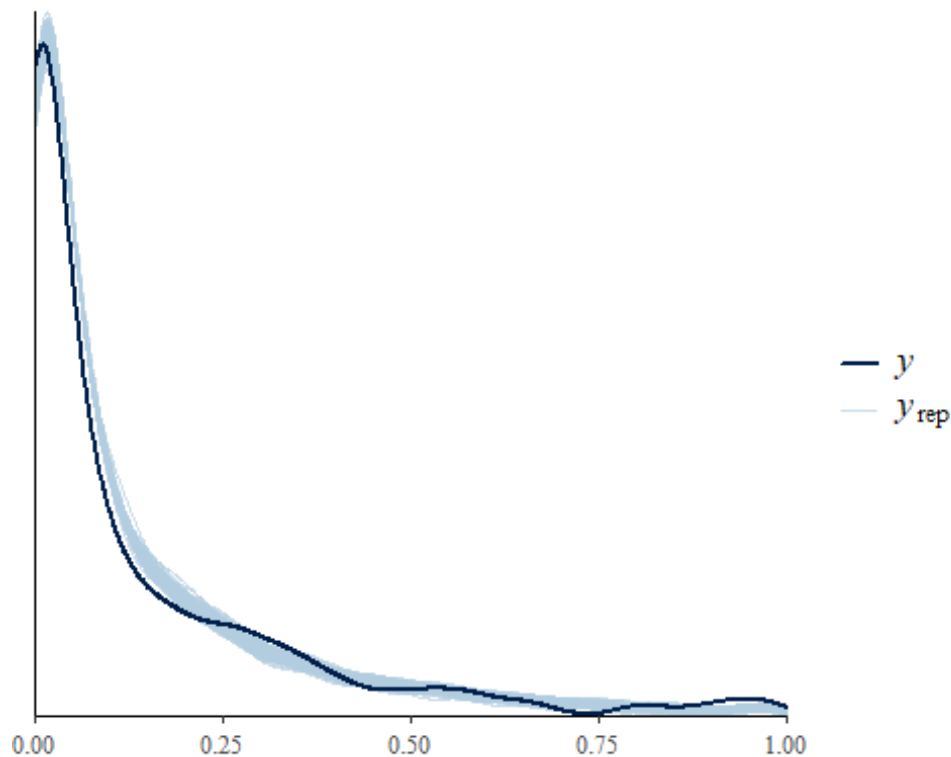
```
## beta[4] 1.014267
## beta[5] 1.002382
## beta[6] 1.002713
## beta[7] 1.001691
## beta[8] 1.003744

#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    7609.1  96.9
## p_loo       1985.5  25.5
## looic     -15218.2 193.8
## ------
## MCSE of elpd_loo is NA.
##
## Pareto k diagnostic values:
##                          Count Pct.    Min. ESS
## (-Inf, 0.7]   (good)     1407  43.5%   117
##    (0.7, 1]   (bad)      1598  49.4%   <NA>
##    (1, Inf)   (very bad)  230   7.1%   <NA>
## See help('pareto-k-diagnostic') for details.

#create data frame of looics from two models
load("2geog5.rds")
loo.mod2=loo.mod # rename loo.mod so can load n
mod2 = mod
```

```
#load("beta interc only.rds")
load("2all_interc.rds") #use full data (dataP) intercept only model for compa
rison

table.data<-data.frame(
  Model = c("Intercept-only model","Model including geography"),
  LOOic = c(loo.mod$estimates[3],loo.mod2$estimates[3]),
  SE = c(loo.mod$estimates[6],loo.mod2$estimates[6])
)
loo.diff = loo.mod2$estimates[3] - loo.mod$estimates[3]
print(loo.diff)

## [1] -4.47479

knitr::kable(table.data, caption = "Table x: Comparisons of LOOic between bas
eline and Model with geography", format = "markdown")
```

Table 1: Comparisons of LOOic between baseline and Model with geography

Model	LOOic	SE
Intercept-only model	-15213.77	193.3243
Model including geography	-15218.24	193.7686

```
load("2geog5.rds")

modx = mod
cats <- c("Africa", "Asia", "New Zealand/Australia", "Europe", "Oceans", "Nor
th America", "South America", "Global")
#Calculate estimates; note original is 1 in matrix
posterior=as.data.frame(modx);

#absolute risks
gm<-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<-8*gm -(beta.1+beta.2+beta.3+beta.4+beta.5+beta.6+beta.7) #last level
is missing in sum contrasts so needs to be calculated from rest

#use common median threat
load("2all_interc.rds")
posterior2=as.data.frame(mod)
grand.mean = posterior2[["mu"]]
grand.mean.inv =invlogit(grand.mean)
```

```r
grand.mean.pred = quantile(grand.mean.inv,probs = c(0.025, 0.5, 0.975))
grand.mean.pred
```

```
##       2.5%        50%      97.5%
## 0.06568248 0.07596676 0.08662260
```

```r
beta.cat<-cbind(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$cats = with (pred.cat.df, reorder(cats, mean))
pred.cat.df <- pred.cat.df[pred.cat.df$cats != "Oceans",] #remove Oceans, whi
ch is covered by habitat grouping now

#relative risks
rbeta.1<-posterior[["beta[2]"]]
rbeta.2<-posterior[["beta[3]"]]
rbeta.3<-posterior[["beta[4]"]]
rbeta.4<-posterior[["beta[5]"]]
rbeta.5<-posterior[["beta[6]"]]
rbeta.6<-posterior[["beta[7]"]]
rbeta.7<-posterior[["beta[8]"]]
rbeta.8<-beta.8 - grand.mean

rbeta.cat = invlogit(beta.cat)-invlogit(grand.mean)
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"
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$cats = with (rpred.cat.df, reorder(cats, mean))
rpred.cat.df <- rpred.cat.df[rpred.cat.df$cats != "Oceans",]

#calculate global mean
g.mean = grand.mean.pred[2]
```

```r
pred.cat.df.r <- pred.cat.df[,1:4]
pred.cat.df.r[,2:4] <- round((pred.cat.df.r[,2:4]), digits = 4)
knitr::kable(pred.cat.df.r, caption = "Table x: Extinction risk by continent
and difference from mean, with 95% credible intervals", format = "markdown",
font = 5)
```

Table 2: Extinction risk by continent and difference from mean, with 95% credible intervals

	x	mean	low	hi
beta.1	Africa	0.0660	0.0512	0.0846
beta.2	Asia	0.0553	0.0434	0.0718
beta.3	New Zealand/Australia	0.1565	0.1221	0.1942
beta.4	Europe	0.0629	0.0487	0.0800
beta.6	North America	0.0630	0.0479	0.0826
beta.7	South America	0.1283	0.0927	0.1736
beta.8	Global	0.0793	0.0609	0.0997

```r
#Figures
Fig1a<-ggplot(data = pred.cat.df)+
  geom_vline(xintercept=g.mean) +
  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,.26)) +
  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
```
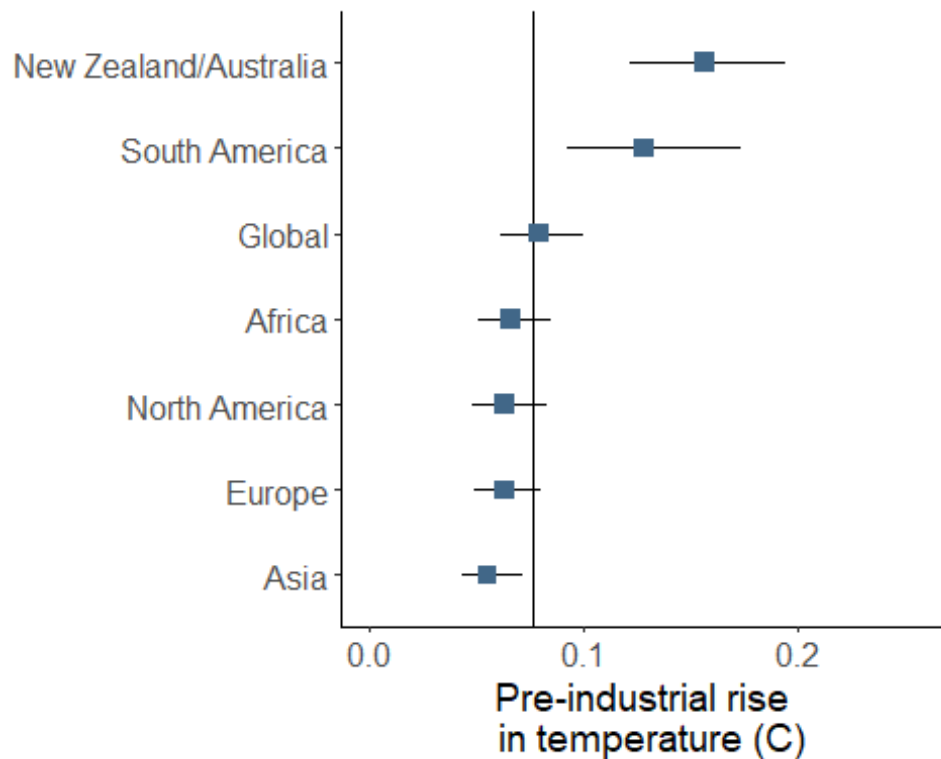
Fig. 1. Predicted extinction risk based on continent.

```r
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,.2)) +
  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
```
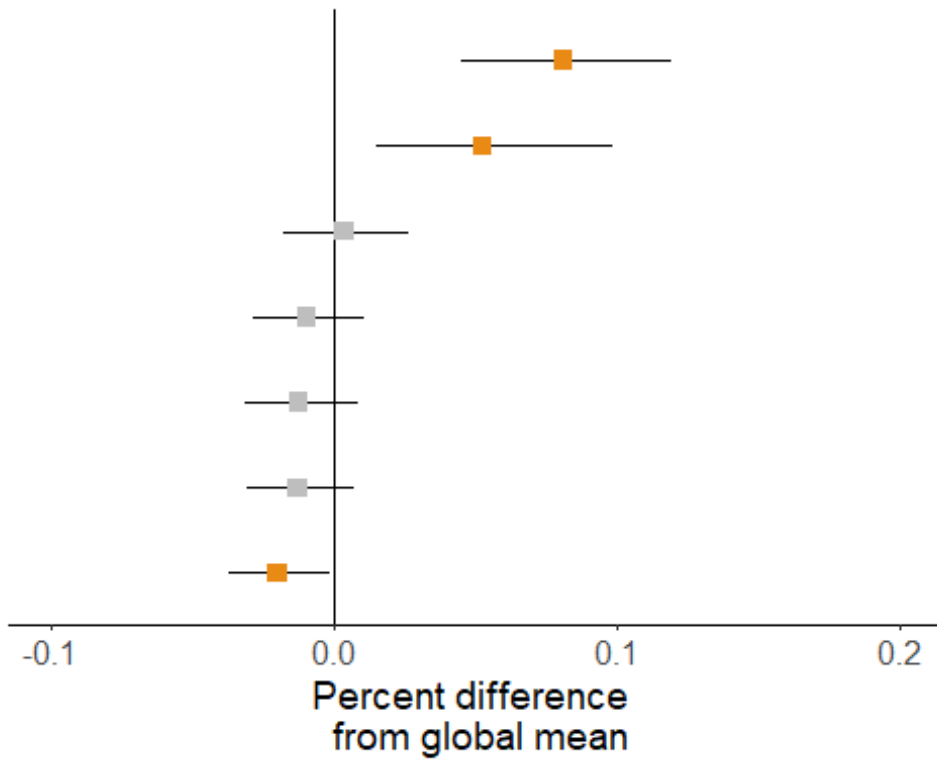
Fig. 2. Predicted extinction risk based on continent.

```
ggarrange(Fig1a, NULL, Fig1b, ncol=3, widths = c(4,.4, 2))
```
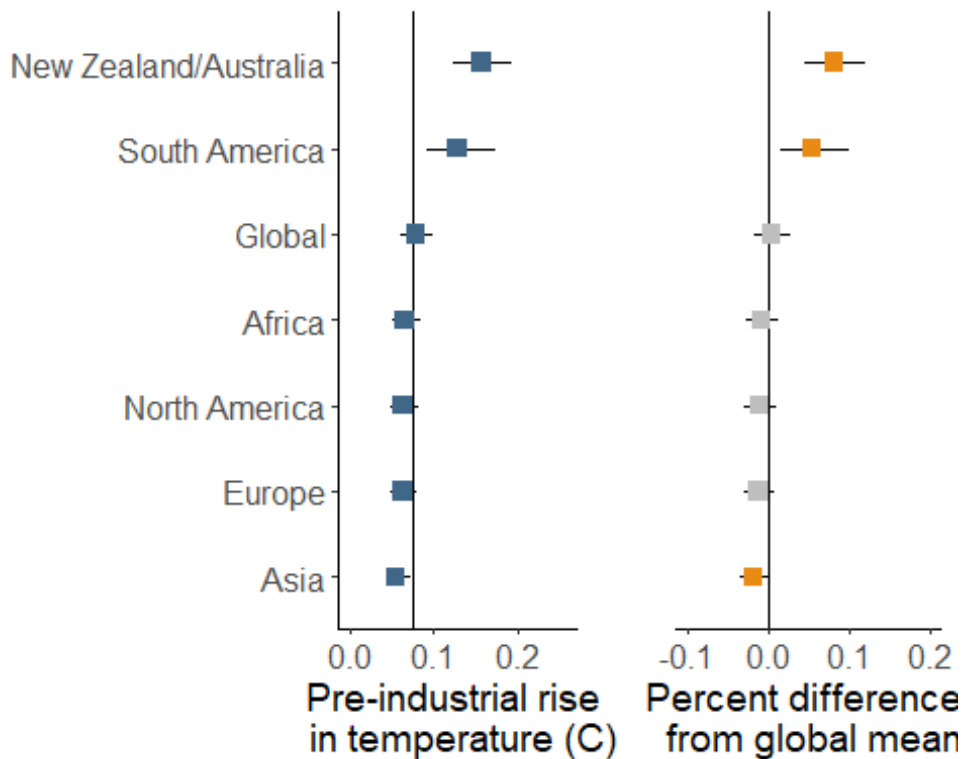
Fig. 3. Predicted extinction risk based on continent.

```r
ggsave("Metarisk2 geog.png",width=6,height=5,unit="in",dpi="print")

(sum(dataP$Region == "Namerica") + sum(dataP$Region == "Europe"))/nrow(dataP)

## [1] 0.4046368

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

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

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

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

N.6.st<-length(unique(dataP$Study[dataP$Region == "Namerica"]))
N.6.mod<-length((dataP$Study[dataP$Region == "Namerica"]))

N.7.st<-length(unique(dataP$Study[dataP$Region == "Africa"]))
```

```
N.7.mod<-length((dataP$Study[dataP$Region == "Africa"]))

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

table.data<-data.frame(
  Factor = c("Asia","South America","Global","Europe","North America","Africa
","Australia/NewZealand"),
  Studies = c(N.1.st,N.3.st,N.4.st,N.5.st,N.6.st,N.7.st,N.8.st),
  Models = c(N.1.mod,N.3.mod,N.4.mod,N.5.mod,N.6.mod,N.7.mod,N.8.mod)
)
knitr::kable(table.data, caption = "Table 2: Number of studies and models for
each factor", format = "markdown")
```

Table 2: Number of studies and models for each factor

Factor	Studies	Models
Asia	67	381
South America	63	307
Global	57	528
Europe	109	711
North America	88	598
Africa	47	289
Australia/NewZealand	45	305

Conclusion

Geographic areas vary in risk, with the highest risks in South America, Australia, and New Zealand. The least risks are in Asia, but this is likely due to a bias toward studies in temperate Asia and focus on trees.