

eDNA climate analysis - models

Table of contents

Percentage of affected species	1
Removing species close to the current temperature limits (tropical)	17
Thermal limits: extracted X experimental	30

This document contain the preliminary models for the eDNA climate analysis.

At the end of the document there is a section called “**Thermal limits: extracted X experimental**” where I compare our thermal limits with those obtained through experiments for a subset of species.

Percentage of affected species

Question: considering a scenario of stability in species composition, how many species of each site will be affected by climate change, i.e. experience temperatures above their upper limits?

Hypothesis: the proportion of affected species will increase in all scenarios, being higher on the worst case scenarios. That trend should be apparent in all sites, but certain sites will be more affected than others. Both databases and eDNA will capture that trend.

```
library(arrow)
library(data.table)
library(lme4)

# Read species summaries
speciesthermsite <- read_parquet("../results/tsummaries_aggregated.parquet")

# Convert to DT
speciesthermsitedt <- data.table(speciesthermsite)
speciesthermsitedt$where <- ifelse(speciesthermsitedt$where == "Both" |
```

```

speciesthermsitedt$where == "OBIS/GBIF", "Databases",

# Produce summaries
sthighergeog <- speciesthermsitedt[,list(ctiavg=mean(q_0.5),
                                         sdcti=sd(q_0.5),
                                         str=mean(q_0.9-q_0.1),
                                         sstavg=mean(site_current),
                                         nspp=.N,
                                         pgt90now=sum(site_current>q_0.9)/.N,
                                         pgt90ssp1d50=sum(site_ssp126_dec50>q_0.9)/
                                         pgt90ssp1d100=sum(site_ssp126_dec100>q_0.9)/
                                         pgt90ssp2d50=sum(site_ssp245_dec50>q_0.9)/
                                         pgt90ssp2d100=sum(site_ssp245_dec100>q_0.9)/
                                         pgt90ssp3d50=sum(site_ssp370_dec50>q_0.9)/
                                         pgt90ssp3d100=sum(site_ssp370_dec100>q_0.9)/
                                         pgt90ssp4d50=sum(site_ssp460_dec50>q_0.9)/
                                         pgt90ssp4d100=sum(site_ssp460_dec100>q_0.9)/
                                         pgt90ssp5d50=sum(site_ssp585_dec50>q_0.9)/
                                         pgt90ssp5d100=sum(site_ssp585_dec100>q_0.9)/
                                         by=c("higherGeography","where"))

head(sthighergeog)

```

		higherGeography	where	ctiavg	sdcti	str	sstavg
1:		aldabra atoll	eDNA	24.77410	4.830172	5.838453	27.44886
2:		aldabra atoll	Databases	26.67343	2.855325	4.454912	27.44886
3:	archipielago de revillagigedo	Databases		23.36119	5.066075	6.748763	26.06002
4:	archipielago de revillagigedo	eDNA		22.10782	5.312405	9.174010	26.06002
5:	banc d arguin national park	eDNA		18.40543	5.638294	9.077047	22.08866
6:	banc d arguin national park	Databases		20.70475	5.581078	7.457039	22.08866
	nspp	pgt90now	pgt90ssp1d50	pgt90ssp1d100	pgt90ssp2d50	pgt90ssp2d100	
1:	399	0.2706767	0.3609023	0.3609023	0.4360902	0.7192982	
2:	1579	0.1209626	0.2058265	0.1988600	0.2596580	0.5104497	
3:	1528	0.2748691	0.3265707	0.3252618	0.3632199	0.4155759	
4:	181	0.2430939	0.2651934	0.2651934	0.3204420	0.4033149	
5:	131	0.3893130	0.4274809	0.4274809	0.4351145	0.4656489	
6:	501	0.2934132	0.3153693	0.3153693	0.3253493	0.3473054	
	pgt90ssp3d50	pgt90ssp3d100	pgt90ssp4d50	pgt90ssp4d100	pgt90ssp5d50		
1:	0.4812030	1.0000000	0.4812030	0.9147870	0.5112782		
2:	0.2932236	0.9987334	0.2913236	0.8777707	0.3299557		
3:	0.3723822	0.6492147	0.3625654	0.4587696	0.3946335		
4:	0.3425414	0.7071823	0.3204420	0.5027624	0.3812155		

5:	0.4351145	0.5190840	0.4351145	0.4885496	0.4580153
6:	0.3253493	0.3872255	0.3233533	0.3552894	0.3353293

pgt90ssp5d100

1:	1.0000000
2:	1.0000000
3:	0.8900524
4:	0.8895028
5:	0.5725191
6:	0.4111776

Once we prepare the data we can make the models. We start by doing a simple linear regression.

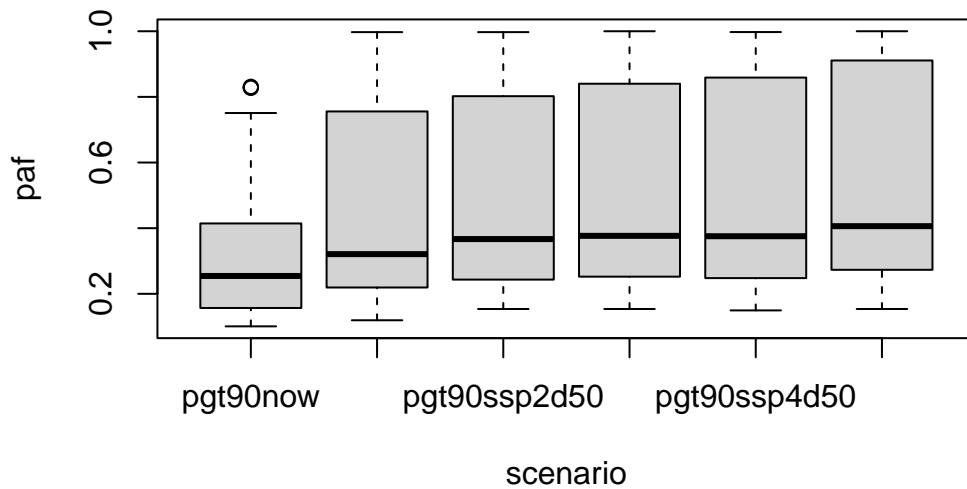
```
sthiergeog$where <- as.factor(sthiergeog$where)
sthiergeog$higherGeography <- as.factor(sthiergeog$higherGeography)

sthiergeog_l <- data.table(tidy::pivot_longer(sthiergeog,
                                              cols = 8:18,
                                              names_to = "scenario", values_to = "paf"))

sthiergeog_l50 <- sthiergeog_l[grepl("d50|now", scenario),,]
sthiergeog_l100 <- sthiergeog_l[grepl("d100|now", scenario),,]

sthiergeog_l50$scenario <- as.factor(sthiergeog_l50$scenario)
sthiergeog_l100$scenario <- as.factor(sthiergeog_l100$scenario)

boxplot(paf ~ scenario, data = sthiergeog_l50)
```



```
paf_lm1 <- glm(paf ~ scenario, data = sthighergeog_l50, family = binomial(), weights = nspp)
summary(paf_lm1)
```

Call:

```
glm(formula = paf ~ scenario, family = binomial(), data = sthighergeog_l50,
     weights = nspp)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.09544	0.01064	-102.94	<2e-16 ***
scenariopgt90ssp1d50	0.55442	0.01431	38.75	<2e-16 ***
scenariopgt90ssp2d50	0.69130	0.01421	48.66	<2e-16 ***
scenariopgt90ssp3d50	0.74629	0.01418	52.65	<2e-16 ***
scenariopgt90ssp4d50	0.73866	0.01418	52.09	<2e-16 ***
scenariopgt90ssp5d50	0.86154	0.01412	61.00	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

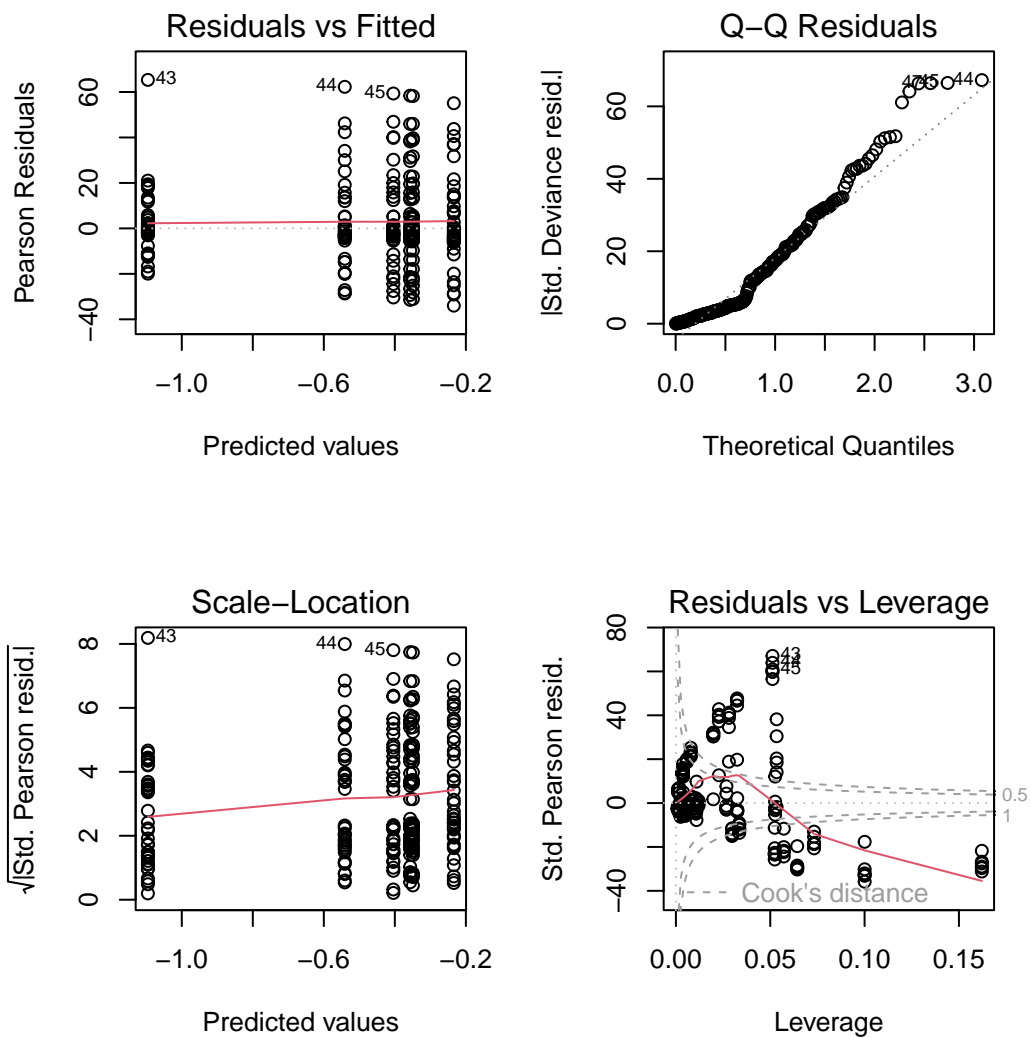
Null deviance: 104532 on 239 degrees of freedom

Residual deviance: 99705 on 234 degrees of freedom
AIC: 101116

Number of Fisher Scoring iterations: 4

Check the residuals:

```
par(mfrow = c(2,2))  
plot(paf_lm1)
```



QQ plot is not good, but the rest is ok.

Model considering the source of the data:

```
paf_lm2 <- glm(paf ~ scenario + where, data = sthighergeog_150, family = binomial(), weights = nspp)

summary(paf_lm2)
```

Call:

```
glm(formula = paf ~ scenario + where, family = binomial(), data = sthighergeog_150,
     weights = nspp)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.14991	0.01076	-106.83	<2e-16 ***
scenariopgt90ssp1d50	0.55839	0.01436	38.89	<2e-16 ***
scenariopgt90ssp2d50	0.69628	0.01426	48.83	<2e-16 ***
scenariopgt90ssp3d50	0.75167	0.01423	52.83	<2e-16 ***
scenariopgt90ssp4d50	0.74398	0.01423	52.28	<2e-16 ***
scenariopgt90ssp5d50	0.86775	0.01418	61.21	<2e-16 ***
whereeDNA	0.64526	0.01457	44.30	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

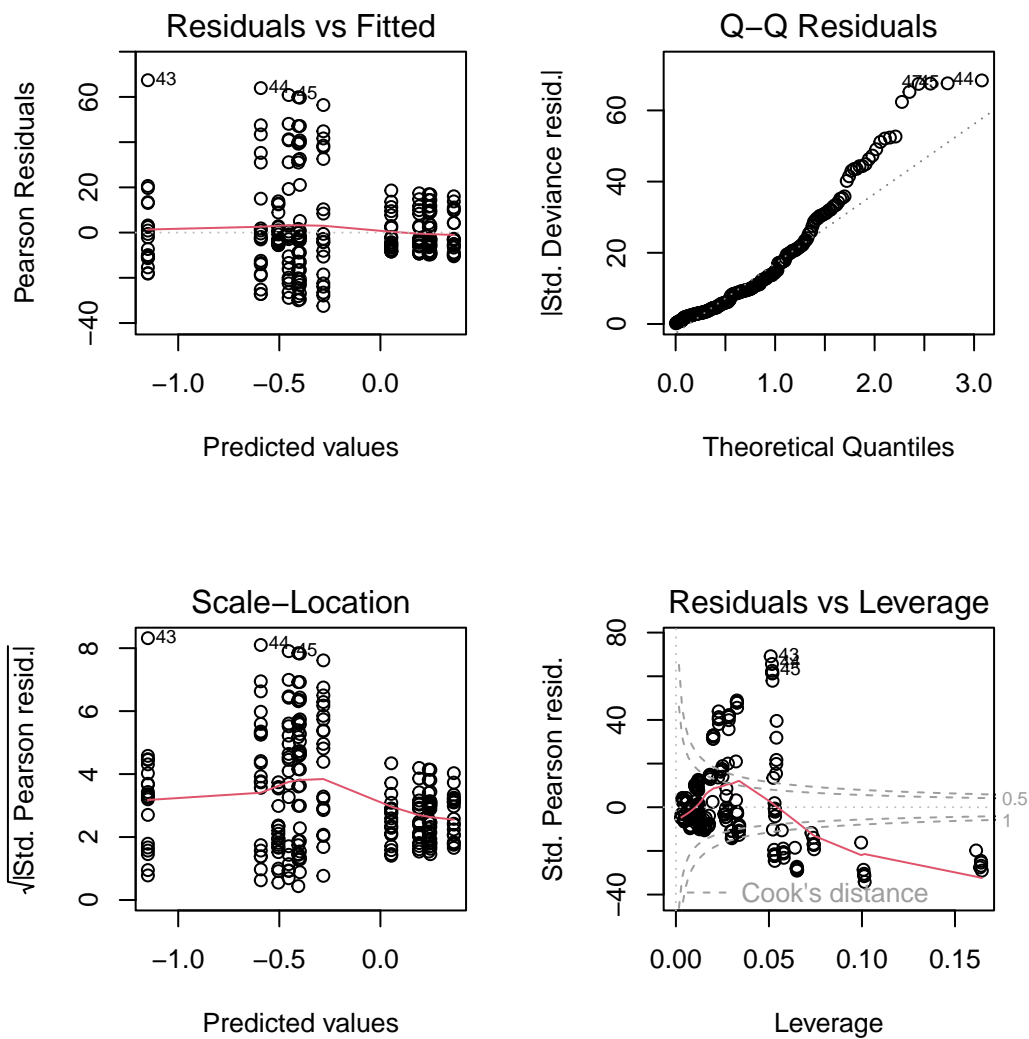
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 104532 on 239 degrees of freedom
Residual deviance: 97747 on 233 degrees of freedom
AIC: 99160

Number of Fisher Scoring iterations: 4

Check the residuals:

```
par(mfrow = c(2,2))
plot(paf_lm2)
```



Try to include the interaction between “where” and “scenario”:

```
paf_lm3 <- glm(paf ~ scenario*where, data = sthigergeog_150, family = binomial(), weights = w)
summary(paf_lm3)
```

Call:

```
glm(formula = paf ~ scenario * where, family = binomial(), data = sthigergeog_150,
```

```
weights = nspp)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.16205	0.01125	-103.293	< 2e-16	***
scenariopgt90ssp1d50	0.57152	0.01506	37.958	< 2e-16	***
scenariopgt90ssp2d50	0.71042	0.01494	47.550	< 2e-16	***
scenariopgt90ssp3d50	0.76500	0.01490	51.329	< 2e-16	***
scenariopgt90ssp4d50	0.75802	0.01491	50.846	< 2e-16	***
scenariopgt90ssp5d50	0.88302	0.01484	59.499	< 2e-16	***
whereeDNA	0.77327	0.03628	21.311	< 2e-16	***
scenariopgt90ssp1d50:whereeDNA	-0.14032	0.05063	-2.772	0.00558	**
scenariopgt90ssp2d50:whereeDNA	-0.15391	0.05067	-3.038	0.00239	**
scenariopgt90ssp3d50:whereeDNA	-0.14366	0.05073	-2.832	0.00463	**
scenariopgt90ssp4d50:whereeDNA	-0.15293	0.05071	-3.016	0.00257	**
scenariopgt90ssp5d50:whereeDNA	-0.17052	0.05086	-3.353	0.00080	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

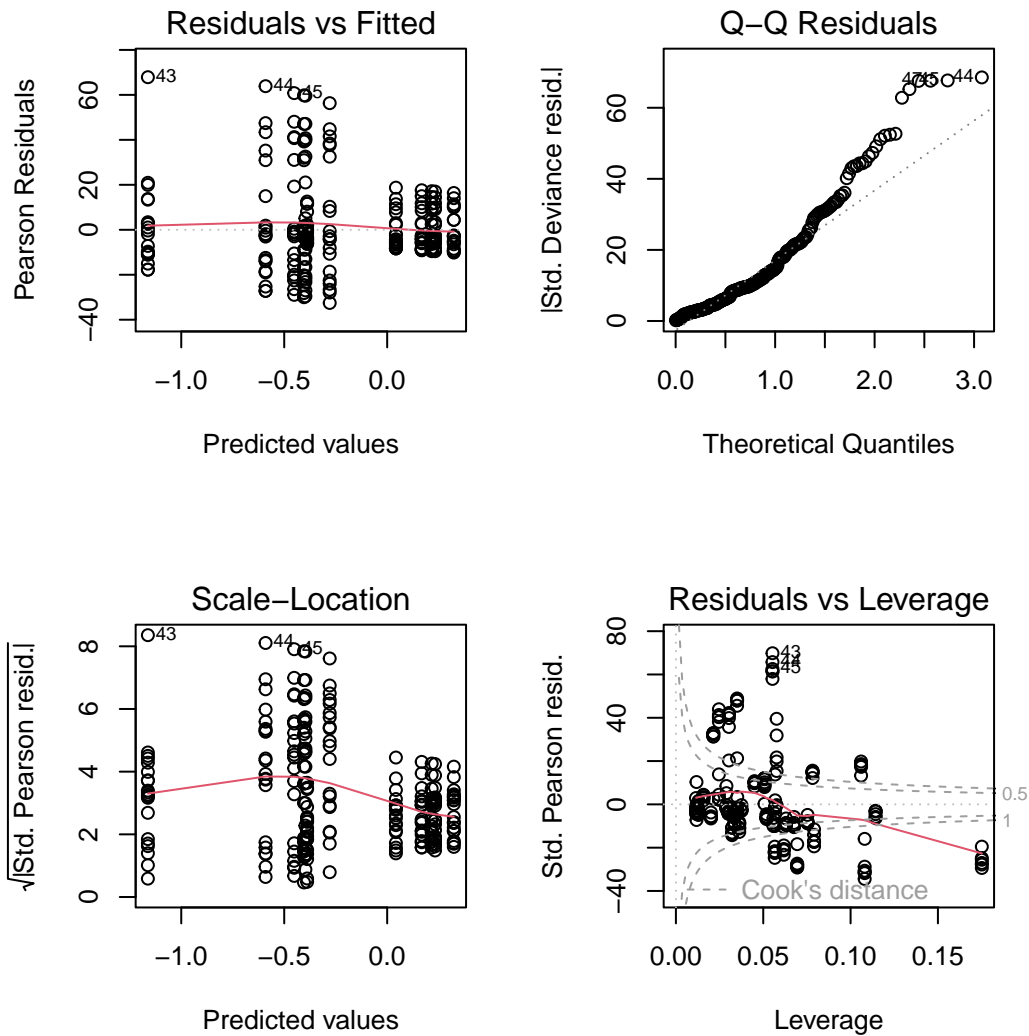
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 104532 on 239 degrees of freedom
Residual deviance: 97732 on 228 degrees of freedom
AIC: 99155

Number of Fisher Scoring iterations: 4

Check residuals:

```
par(mfrow = c(2,2))  
plot(paf_lm3)
```

Try a Linear Mixed-Effects model. In this case, we will focus in modelling $\text{PAF} \sim \text{scenario}$, with “where” as a random effect.

```
paf_lme1 <- glmer(paf ~ scenario + (1|where), data = sthighergeog_l50, family = binomial(),
                  weights = nspp)

summary(paf_lme1)
```

Generalized linear mixed model fit by maximum likelihood (Laplace

```

Approximation) [glmerMod]
Family: binomial ( logit )
Formula: paf ~ scenario + (1 | where)
Data: sthighergeog_l50
Weights: nspp

```

```

      AIC      BIC   logLik deviance df.resid
99177.4  99201.7 -49581.7  99163.4      233

```

Scaled residuals:

```

      Min       1Q   Median       3Q      Max
-32.438  -8.815  -3.060   9.148  67.390

```

Random effects:

```

Groups Name          Variance Std.Dev.
where (Intercept) 0.104    0.3225
Number of obs: 240, groups:  where, 2

```

Fixed effects:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.82755    0.22831  -3.625 0.000289 ***
scenariopgt90ssp1d50  0.55839    0.01436  38.888 < 2e-16 ***
scenariopgt90ssp2d50  0.69628    0.01426  48.832 < 2e-16 ***
scenariopgt90ssp3d50  0.75167    0.01423  52.834 < 2e-16 ***
scenariopgt90ssp4d50  0.74398    0.01423  52.278 < 2e-16 ***
scenariopgt90ssp5d50  0.86775    0.01418  61.212 < 2e-16 ***
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Correlation of Fixed Effects:

```

      (Intr) s90150 s90250 s90350 s90450
scnrpg90150 -0.034
scnrpg90250 -0.035  0.557
scnrpg90350 -0.035  0.558  0.562
scnrpg90450 -0.035  0.558  0.562  0.563
scnrpg90550 -0.035  0.560  0.564  0.566  0.566

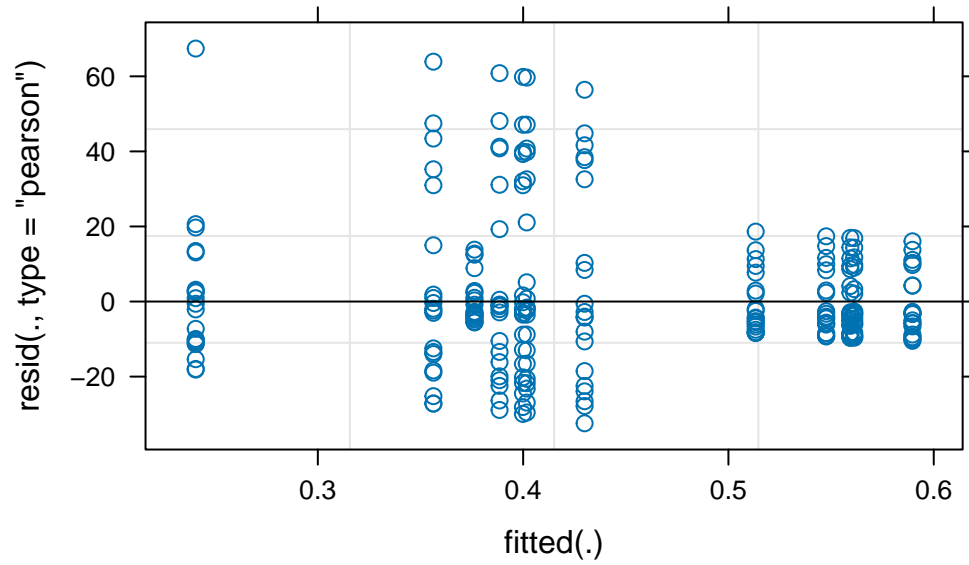
```

The source does explain some of the variation, but not a great amount compared to the scenarios.

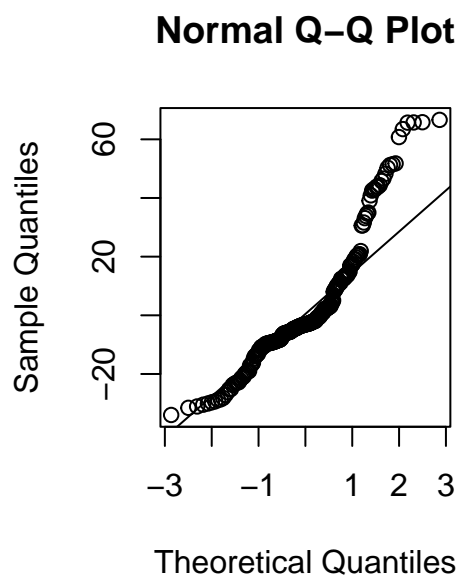
```

par(mfrow = c(1,2))
plot(paf_lme1)

```



```
qqnorm(resid(paf_lme1))
qqline(resid(paf_lme1))
```



Try now a model with higherGeography as a random effect:

```
paf_lme2 <- glmer(paf ~ scenario + (1|higherGeography), data = sthighergeog_150, family =  
summary(paf_lme2)
```

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]

Family: binomial (logit)

Formula: paf ~ scenario + (1 | higherGeography)

Data: sthighergeog_150

Weights: nspp

AIC	BIC	logLik	deviance	df.resid
10576.5	10600.9	-5281.3	10562.5	233

Scaled residuals:

Min	1Q	Median	3Q	Max
-34.627	-1.972	0.862	4.198	20.998

Random effects:

Groups	Name	Variance	Std.Dev.
higherGeography	(Intercept)	2.306	1.519

Number of obs: 240, groups: higherGeography, 20

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.04305	0.33983	-3.069	0.00215 **
scenariopgt90ssp1d50	0.85993	0.01802	47.731	< 2e-16 ***
scenariopgt90ssp2d50	1.05616	0.01784	59.203	< 2e-16 ***
scenariopgt90ssp3d50	1.13329	0.01778	63.736	< 2e-16 ***
scenariopgt90ssp4d50	1.12264	0.01779	63.110	< 2e-16 ***
scenariopgt90ssp5d50	1.29197	0.01768	73.077	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

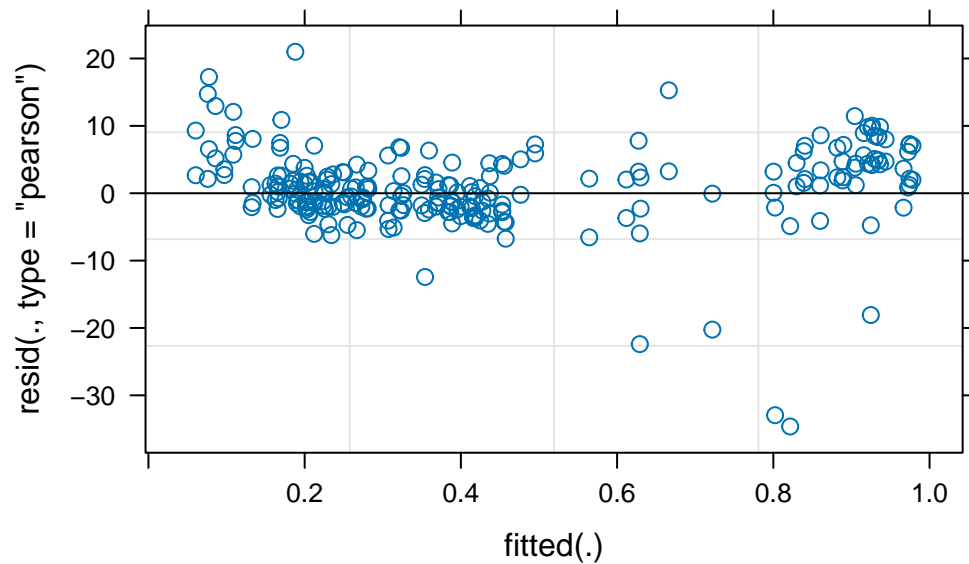
(Intr)	s90150	s90250	s90350	s90450
scnrpg90150	-0.030			
scnrpg90250	-0.031	0.598		
scnrpg90350	-0.031	0.601	0.610	
scnrpg90450	-0.031	0.601	0.610	0.613

```
scnrpg90550 -0.031  0.606  0.616  0.619  0.619
```

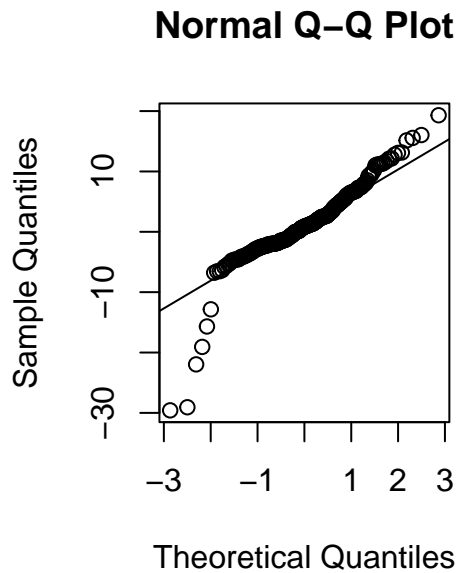
In this case, the higher Geography have more influence on the values of PAF. This seems to be an interesting model to proceed with.

Check residuals:

```
par(mfrow = c(1,2))  
plot(paf_lme2)
```



```
qqnorm(resid(paf_lme2))  
qqline(resid(paf_lme2))
```



Both models show the trend of increase in affected species as we go to the worst case scenarios.

Try to include the temperature trend on the model:

```
paf_lm_sst1 <- glm(paf ~ scenario + where + sstavg, data = sthigergeog_150, family = binomial)
summary(paf_lm_sst1)
```

Call:

```
glm(formula = paf ~ scenario + where + sstavg, family = binomial(),
     data = sthigergeog_150, weights = nspp)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.1932606	0.0233542	-136.73	<2e-16 ***
scenariopgt90ssp1d50	0.5777583	0.0146050	39.56	<2e-16 ***
scenariopgt90ssp2d50	0.7216029	0.0145155	49.71	<2e-16 ***
scenariopgt90ssp3d50	0.7795244	0.0144884	53.80	<2e-16 ***
scenariopgt90ssp4d50	0.7714791	0.0144919	53.23	<2e-16 ***
scenariopgt90ssp5d50	0.9011562	0.0144477	62.37	<2e-16 ***
whereeDNA	0.5216678	0.0148397	35.15	<2e-16 ***

```
sstavg          0.0847700  0.0008321  101.88   <2e-16 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 104532  on 239  degrees of freedom
Residual deviance:  85073  on 232  degrees of freedom
AIC: 86488
```

Number of Fisher Scoring iterations: 5

```
sthigergeog_l50_sc <- sthigergeog_l50
sthigergeog_l50_sc$sstavg <- scale(sthigergeog_l50_sc$sstavg)

paf_lme_sst1 <- glmer(paf ~ scenario + sstavg + (1|where), data = sthigergeog_l50_sc, fam
summary(paf_lme_sst1)
```

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial (logit)
Formula: paf ~ scenario + sstavg + (1 | where)
Data: sthigergeog_l50_sc
Weights: nspp

AIC	BIC	logLik	deviance	df.resid
86504.3	86532.2	-43244.2	86488.3	232

Scaled residuals:

Min	1Q	Median	3Q	Max
-37.228	-7.345	-0.848	9.814	57.331

Random effects:

Groups	Name	Variance	Std.Dev.
where	(Intercept)	0.06792	0.2606

Number of obs: 240, groups: where, 2

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.940267	0.184703	-5.091	3.57e-07 ***

scenariopgt90ssp1d50	0.577752	0.014605	39.559	< 2e-16 ***
scenariopgt90ssp2d50	0.721597	0.014515	49.712	< 2e-16 ***
scenariopgt90ssp3d50	0.779517	0.014488	53.803	< 2e-16 ***
scenariopgt90ssp4d50	0.771472	0.014492	53.235	< 2e-16 ***
scenariopgt90ssp5d50	0.901148	0.014448	62.373	< 2e-16 ***
sstavg	0.520552	0.005109	101.882	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	s90150	s90250	s90350	s90450	s90550
scnrpg90150	-0.043					
scnrpg90250	-0.043	0.554				
scnrpg90350	-0.043	0.555	0.559			
scnrpg90450	-0.043	0.555	0.558	0.560		
scnrpg90550	-0.044	0.557	0.560	0.562	0.561	
sstavg	-0.008	0.020	0.026	0.029	0.029	0.036

```
paf_gam_sst1 <- mgcv::gam(paf ~ scenario + where + s(sstavg, k = 4), data = sthighergeog_1)
summary(paf_gam_sst1)
```

Family: binomial

Link function: logit

Formula:

paf ~ scenario + where + s(sstavg, k = 4)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.18525	0.01201	-98.65	<2e-16 ***
scenariopgt90ssp1d50	0.69078	0.01599	43.19	<2e-16 ***
scenariopgt90ssp2d50	0.86188	0.01590	54.20	<2e-16 ***
scenariopgt90ssp3d50	0.93053	0.01587	58.62	<2e-16 ***
scenariopgt90ssp4d50	0.92100	0.01588	58.00	<2e-16 ***
scenariopgt90ssp5d50	1.07411	0.01583	67.83	<2e-16 ***
whereeDNA	0.33379	0.01695	19.70	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:


```

          edf Ref.df Chi.sq p-value
s(sstavg) 2.998      3 44035 <2e-16 ***
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

R-sq.(adj) = 0.663  Deviance explained = 60%
UBRE = 173.09  Scale est. = 1          n = 240

```

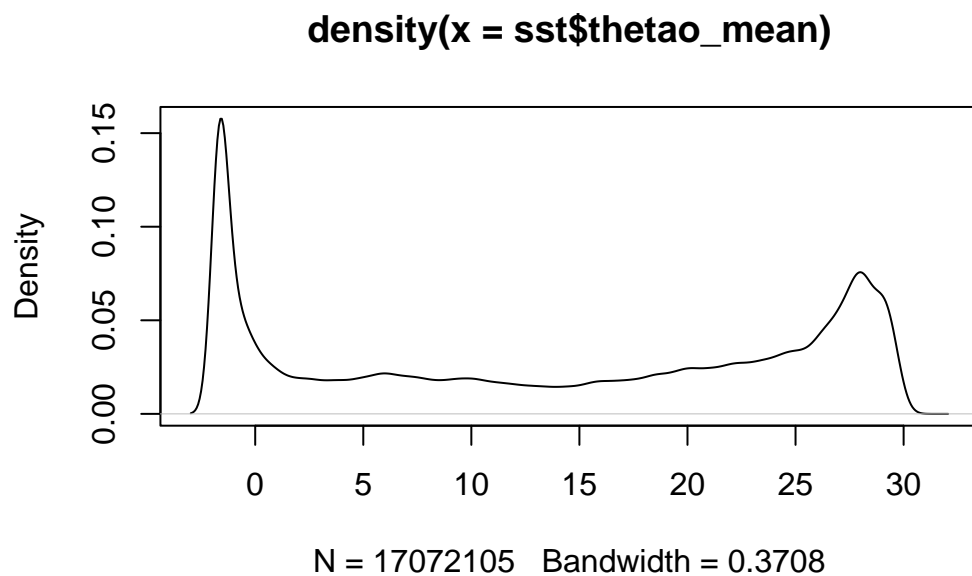
Removing species close to the current temperature limits (tropical)

```

sst <- terra::rast("~/Research/mpa_europe/mpaeu_sdm/data/env/current/thetao_baseline_depth")
sst <- as.data.frame(sst)

plot(density(sst$thetao_mean))

```



```

sst_q <- quantile(sst$thetao_mean, 0.9) # We can try other values

speciesthermsitedt <- speciesthermsitedt[q_1 <= sst_q,,]

sthigergeog <- speciesthermsitedt[,list(ctiavg=mean(q_0.5),

```

```

sdcti=sd(q_0.5),
str=mean(q_0.9-q_0.1),
sstavg=mean(site_current),
nspp=.N,
pgt90now=sum(site_current>q_0.9)/.N,
pgt90ssp1d50=sum(site_ssp126_dec50>q_0.9)/
pgt90ssp1d100=sum(site_ssp126_dec100>q_0.9)/
pgt90ssp2d50=sum(site_ssp245_dec50>q_0.9)/
pgt90ssp2d100=sum(site_ssp245_dec100>q_0.9)/
pgt90ssp3d50=sum(site_ssp370_dec50>q_0.9)/
pgt90ssp3d100=sum(site_ssp370_dec100>q_0.9)/
pgt90ssp4d50=sum(site_ssp460_dec50>q_0.9)/
pgt90ssp4d100=sum(site_ssp460_dec100>q_0.9)/
pgt90ssp5d50=sum(site_ssp585_dec50>q_0.9)/
pgt90ssp5d100=sum(site_ssp585_dec100>q_0.9)/
by=c("higherGeography", "where")]

sthigergeog$where <- as.factor(sthigergeog$where)
sthigergeog$higherGeography <- as.factor(sthigergeog$higherGeography)

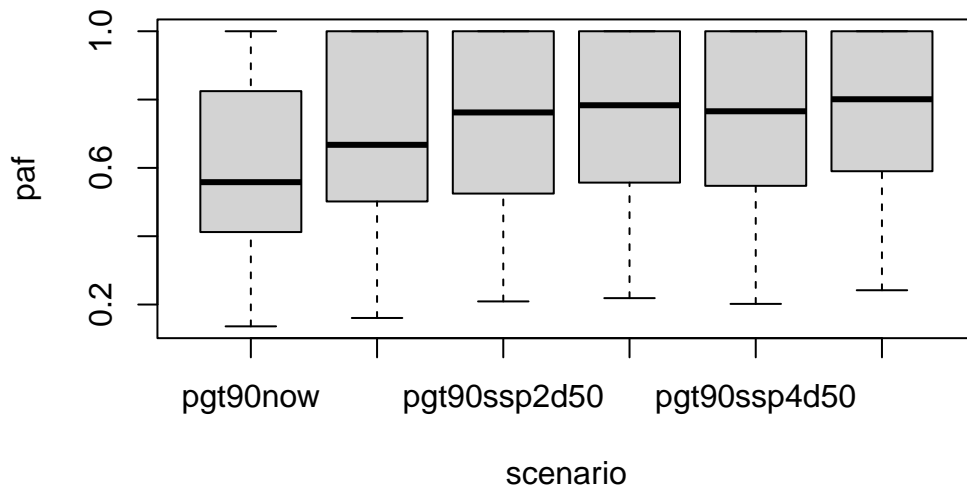
sthigergeog_l <- data.table(tidyr::pivot_longer(sthigergeog,
                                                cols = 8:18,
                                                names_to = "scenario", values_to = "paf

sthigergeog_l50 <- sthigergeog_l[grepl("d50|now", scenario),,]
sthigergeog_l100 <- sthigergeog_l[grepl("d100|now", scenario),,]

sthigergeog_l50$scenario <- as.factor(sthigergeog_l50$scenario)
sthigergeog_l100$scenario <- as.factor(sthigergeog_l100$scenario)

boxplot(paf ~ scenario, data = sthigergeog_l50)

```



Try again the same models:

```
paf_lm1_rs <- glm(paf ~ scenario, data = sthigergeog_150, family = binomial(), weights =
summary(paf_lm1_rs)
```

Call:

```
glm(formula = paf ~ scenario, family = binomial(), data = sthigergeog_150,
weights = nspp)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.18866	0.01592	-11.85	<2e-16 ***
scenariopgt90ssp1d50	0.52257	0.02263	23.10	<2e-16 ***
scenariopgt90ssp2d50	0.71503	0.02286	31.28	<2e-16 ***
scenariopgt90ssp3d50	0.80354	0.02301	34.92	<2e-16 ***
scenariopgt90ssp4d50	0.75099	0.02292	32.77	<2e-16 ***
scenariopgt90ssp5d50	0.90641	0.02321	39.05	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

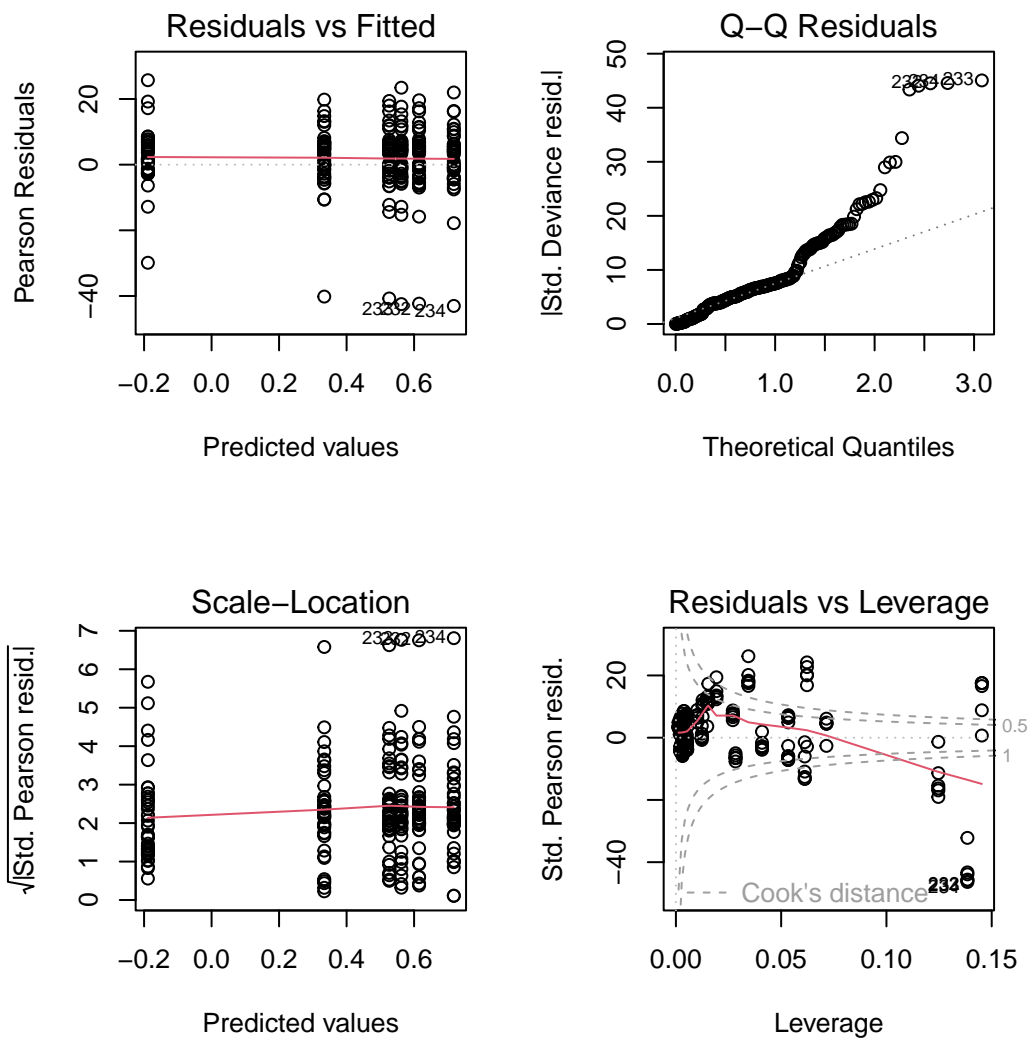
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 30953 on 239 degrees of freedom
Residual deviance: 28879 on 234 degrees of freedom
AIC: 29752

Number of Fisher Scoring iterations: 4

Check the residuals:

```
par(mfrow = c(2,2))  
plot(paf_lm1_rs)
```



QQ plot is not good, but the rest is ok.

Model considering the source of the data:

```
paf_lm2_rs <- glm(paf ~ scenario + where, data = sthigergeog_150, family = binomial(), weights = w)
summary(paf_lm2_rs)
```

Call:

```
glm(formula = paf ~ scenario + where, family = binomial(), data = sthighergeog_150,
     weights = nspp)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.21480	0.01606	-13.38	<2e-16 ***
scenariopgt90ssp1d50	0.52370	0.02265	23.12	<2e-16 ***
scenariopgt90ssp2d50	0.71654	0.02289	31.31	<2e-16 ***
scenariopgt90ssp3d50	0.80522	0.02303	34.96	<2e-16 ***
scenariopgt90ssp4d50	0.75257	0.02294	32.80	<2e-16 ***
scenariopgt90ssp5d50	0.90827	0.02323	39.09	<2e-16 ***
whereeDNA	0.37128	0.02744	13.53	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

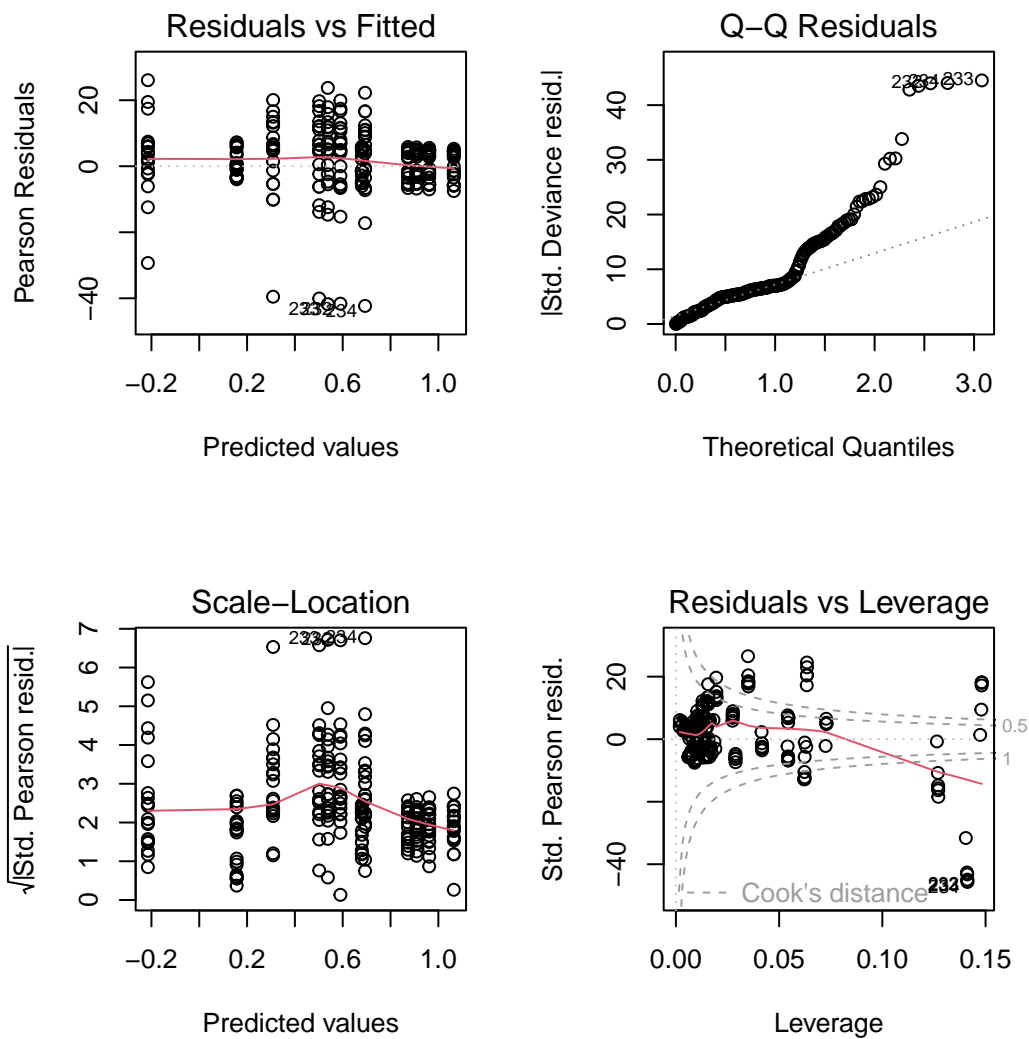
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 30953 on 239 degrees of freedom
 Residual deviance: 28689 on 233 degrees of freedom
 AIC: 29564

Number of Fisher Scoring iterations: 4

Check the residuals:

```
par(mfrow = c(2,2))
plot(paf_lm2_rs)
```



Try to include the interaction between “where” and “scenario”:

```
paf_lm3_rs <- glm(paf ~ scenario*where, data = sthigergeog_150, family = binomial(), weights = w)
summary(paf_lm3_rs)
```

Call:

```
glm(formula = paf ~ scenario * where, family = binomial(), data = sthigergeog_150,
```

```
weights = nspp)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.23671	0.01655	-14.300	< 2e-16 ***
scenariopgt90ssp1d50	0.54361	0.02347	23.167	< 2e-16 ***
scenariopgt90ssp2d50	0.74451	0.02371	31.406	< 2e-16 ***
scenariopgt90ssp3d50	0.83338	0.02385	34.938	< 2e-16 ***
scenariopgt90ssp4d50	0.77922	0.02376	32.795	< 2e-16 ***
scenariopgt90ssp5d50	0.93889	0.02406	39.023	< 2e-16 ***
whereeDNA	0.68807	0.06367	10.807	< 2e-16 ***
scenariopgt90ssp1d50:whereeDNA	-0.28560	0.09163	-3.117	0.00183 **
scenariopgt90ssp2d50:whereeDNA	-0.41237	0.09228	-4.469	7.87e-06 ***
scenariopgt90ssp3d50:whereeDNA	-0.41638	0.09308	-4.473	7.70e-06 ***
scenariopgt90ssp4d50:whereeDNA	-0.39225	0.09278	-4.228	2.36e-05 ***
scenariopgt90ssp5d50:whereeDNA	-0.45628	0.09378	-4.866	1.14e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

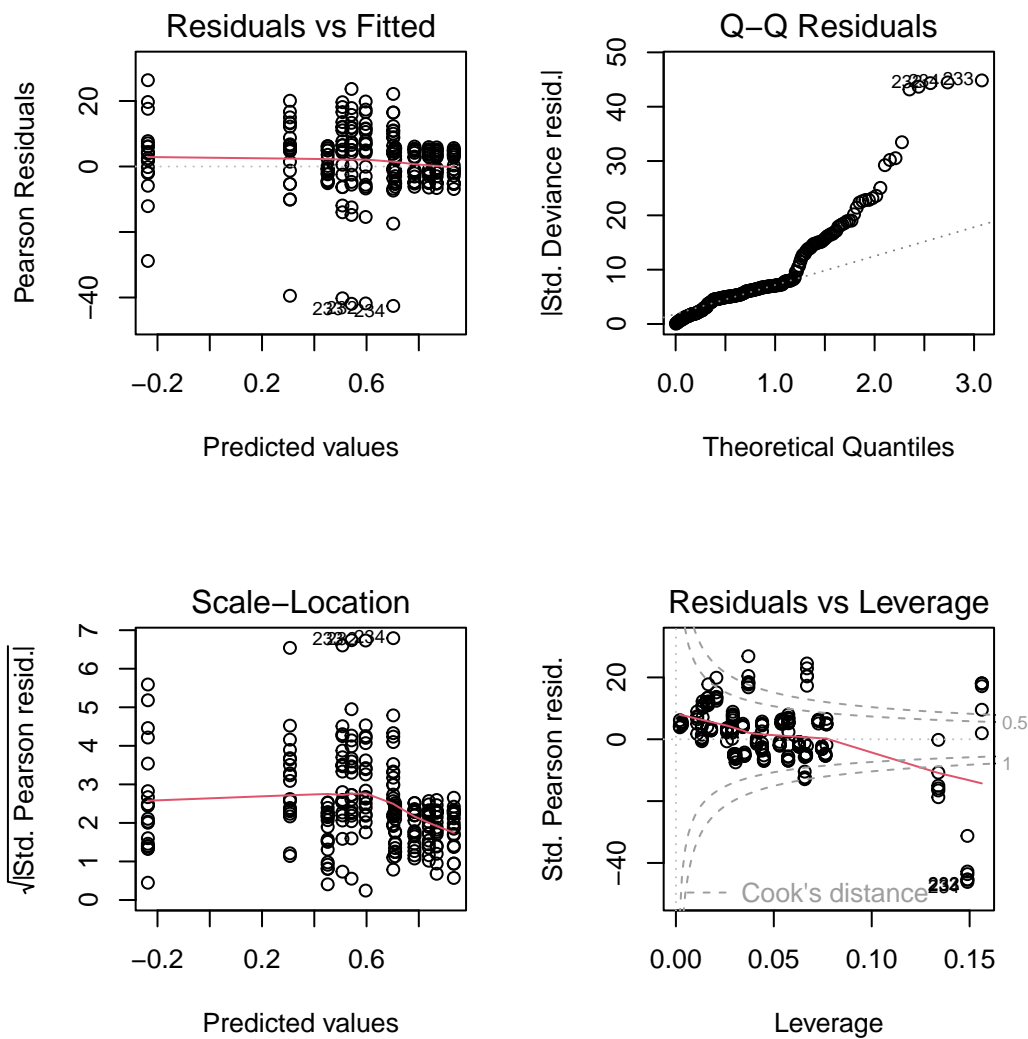
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 30953 on 239 degrees of freedom
Residual deviance: 28655 on 228 degrees of freedom
AIC: 29539

Number of Fisher Scoring iterations: 4

Check residuals:

```
par(mfrow = c(2,2))  
plot(paf_lm3_rs)
```

Try a Linear Mixed-Effects model. In this case, we will focus in modelling $\text{PAF} \sim \text{scenario}$, with “where” as a random effect.

```
paf_lme1_rs <- glmer(paf ~ scenario + (1|where), data = sthighergeog_150, family = binomial, weights = nspp)

summary(paf_lme1_rs)
```

Generalized linear mixed model fit by maximum likelihood (Laplace

```

Approximation) [glmerMod]
Family: binomial ( logit )
Formula: paf ~ scenario + (1 | where)
Data: sthighergeog_l50
Weights: nspp

      AIC      BIC   logLik deviance df.resid
29576.7  29601.0 -14781.3  29562.7      233

Scaled residuals:
      Min       1Q   Median       3Q      Max
-42.325  -3.183   2.870   5.858  26.060

Random effects:
Groups Name          Variance Std.Dev.
where (Intercept) 0.0341   0.1847
Number of obs: 240, groups:  where, 2

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -0.03086    0.13208  -0.234    0.815
scenariopgt90ssp1d50  0.52369    0.02265  23.121 <2e-16 ***
scenariopgt90ssp2d50  0.71652    0.02289  31.308 <2e-16 ***
scenariopgt90ssp3d50  0.80520    0.02303  34.960 <2e-16 ***
scenariopgt90ssp4d50  0.75255    0.02294  32.801 <2e-16 ***
scenariopgt90ssp5d50  0.90825    0.02323  39.094 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) s90150 s90250 s90350 s90450
scnrpg90150 -0.084
scnrpg90250 -0.083  0.490
scnrpg90350 -0.083  0.487  0.482
scnrpg90450 -0.083  0.489  0.484  0.481
scnrpg90550 -0.082  0.483  0.478  0.475  0.477

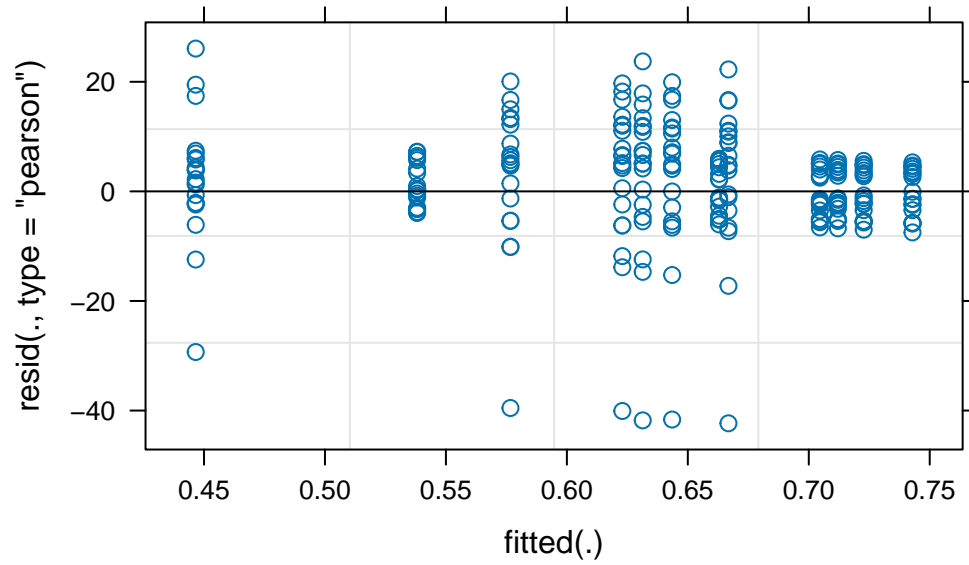
```

The source does explain some of the variation, but not a great amount compared to the scenarios.

```

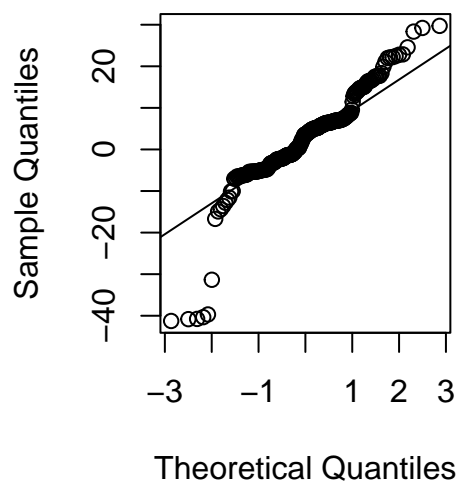
par(mfrow = c(1,2))
plot(paf_lme1_rs)

```



```
qqnorm(resid(paf_lme1_rs))
qqline(resid(paf_lme1_rs))
```

Normal Q-Q Plot



Try now a model with higherGeography as a random effect:

```
paf_lme2_rs <- glmer(paf ~ scenario + (1|higherGeography), data = sthighergeog_150, family = binomial)
summary(paf_lme2_rs)
```

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]

Family: binomial (logit)

Formula: paf ~ scenario + (1 | higherGeography)

Data: sthighergeog_150

Weights: nspp

AIC	BIC	logLik	deviance	df.resid
3710.4	3734.7	-1848.2	3696.4	233

Scaled residuals:

Min	1Q	Median	3Q	Max
-22.880	-1.439	0.158	1.230	16.582

Random effects:

Groups	Name	Variance	Std.Dev.
higherGeography	(Intercept)	12.02	3.466

Number of obs: 240, groups: higherGeography, 20

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.75402	0.78958	2.221	0.0263 *
scenariopgt90ssp1d50	0.70357	0.02634	26.711	<2e-16 ***
scenariopgt90ssp2d50	0.95545	0.02657	35.955	<2e-16 ***
scenariopgt90ssp3d50	1.07013	0.02673	40.039	<2e-16 ***
scenariopgt90ssp4d50	1.00213	0.02663	37.628	<2e-16 ***
scenariopgt90ssp5d50	1.20249	0.02694	44.632	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

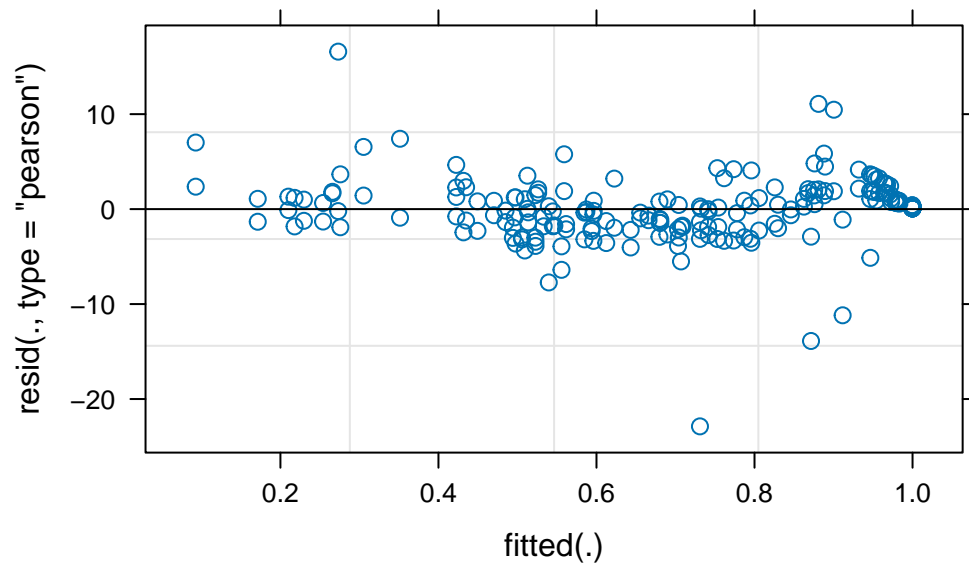
	(Intr)	s90150	s90250	s90350	s90450
scnrpg90150	-0.016				
scnrpg90250	-0.015	0.507			
scnrpg90350	-0.015	0.505	0.503		
scnrpg90450	-0.015	0.506	0.504	0.502	

```
scnrpg90550 -0.015  0.502  0.500  0.499  0.500
```

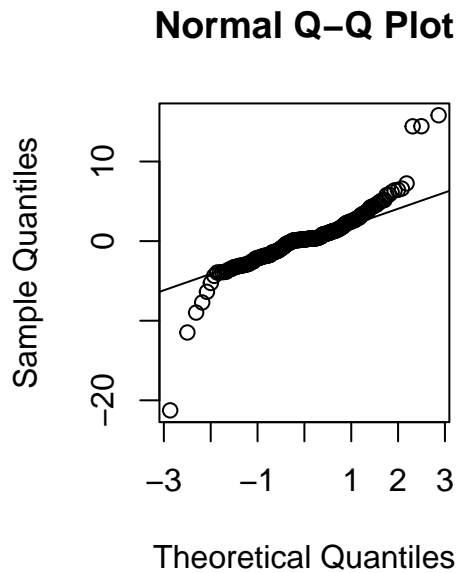
The higher Geography have even more influence on the values of PAF.

Check residuals:

```
par(mfrow = c(1,2))  
plot(paf_lme2_rs)
```



```
qqnorm(resid(paf_lme2_rs))  
qqline(resid(paf_lme2_rs))
```



Thermal limits: extracted X experimental

I obtained the thermal limits from the [GlobTherm dataset](#).

```
suppressPackageStartupMessages(library(dplyr))

# Get only unique species - because we will use only the thermal limits, we can get from t
species_filt <- speciethermsitedt[!duplicated(speciethermsitedt$species),]

# Load the GlobTherm dataset
globtherm <- suppressMessages(suppressWarnings(readxl::read_xlsx("~/Downloads/GlobalTherm_
globtherm_sp <- paste(globtherm$Genus, globtherm$Species)

# See which species are available
gt_sel <- which(globtherm_sp %in% species_filt$species)
gt_sel <- globtherm[gt_sel,]
gt_sel <- gt_sel %>%
  mutate(species = globtherm_sp[globtherm_sp %in% species_filt$species]) %>%
  select(species, Tmax, max_metric)

# See which metrics are available
```

```
table(gt_sel$max_metric)
```

ctmax	LT0	LT100	LT50	UTNZ
32	93	5	36	2

```
# Select only those that are on our list
species_filt_sel <- species_filt[species_filt$species %in% gt_sel$species,]
species_filt_sel <- species_filt_sel %>%
  select(species, q_0.9)

# Join both tables
tlimits <- species_filt_sel %>%
  left_join(gt_sel) %>%
  filter(!is.na(Tmax))
```

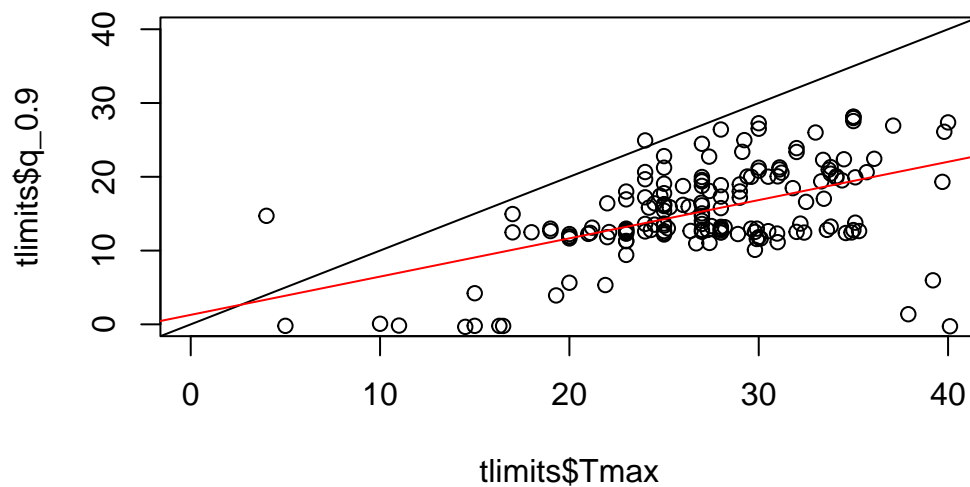
Joining with `by = join_by(species)`

```
nrow(tlimits)
```

```
[1] 168
```

```
# Plot
plot(tlimits$q_0.9 ~ tlimits$Tmax, xlim = c(0, 40), ylim = c(0,40))
abline(a = 0, b = 1)

m1 <- lm(q_0.9~Tmax, data = tlimits)
abline(m1, col = "red")
```



We see that for a great portion of the species, the species thermal maxima is actually much higher than the one we estimated through the occurrence data.

It is also possible to use only “CTmax”:

```
gt_sel_ctmax <- gt_sel %>%
  filter(max_metric == "ctmax")

# Join both tables
tlimits_ctmax <- species_filt_sel %>%
  left_join(gt_sel_ctmax) %>%
  filter(!is.na(Tmax))
```

Joining with `by = join_by(species)`

```
nrow(tlimits_ctmax)
```

```
[1] 32
```



```
# Plot
plot(tlimits_ctmax$q_0.9 ~ tlimits_ctmax$Tmax, xlim = c(0, 40), ylim = c(0,40))
abline(a = 0, b = 1)

m2 <- lm(q_0.9~Tmax, data = tlimits_ctmax)
abline(m1, col = "red")
```

