Streamlined Thesis Analyses

Laura Vary

1/10/2022

Contents

Purpose of This Document:

The purpose of this document is to streamline files and associated analyses for the creation of generalized additive models that investigate spawning behavior and larval biogeography among fishes in the Bering Sea. This is mainly an automation document, with the goal of minimizing the back-and-forth between code files should data need to be modified or analyses re-ran. The document is organized by species, with species that have both egg and larval analyses appearing first.

Loading Data:

Alaska plaice: both egg and larval data are included for this species. Plaice spawn during April and May, live over 30 years, and transform to juveniles at standard lengths $\geq$ to 10.7 mm.

```
apsub<-read.csv(file='./Ichthyo Data/Cleaned_Cut_ApEggs.csv',header=TRUE,
                check.names=TRUE)
aplarv.ctd<-read.csv(file='./Ichthyo Data/Cleaned_Cut_ApLarv_wCTD.csv',
                     header=TRUE,check.names=TRUE)
aplarv.ctd<-subset(aplarv.ctd,doy>80&doy<180)

reg.sst<-read.csv('./Environmental Data/Mar_SST_RegionalIndex_NCEP_BS.csv',
                  header=TRUE,check.names=TRUE)

str_name<-"./Environmental Data/expanded_BS_bathy.tif"
bathy<-raster(str_name)
```

These data have been trimmed. The egg data are constrained to depths <151 meters; temporally, the egg data are constrained to above the 99th day of year and below the 182nd day of the year (temporally centered on the spawning period of plaice). The egg data are also joined to regional temperature indices for each year (the reg.sst dataset). The larval data are constrained to depths <151 meters, between 80 and 180 day of year, and are linked to CTD-derived, *in situ* temperature and salinity measurements.

The regional temperature index data are constrained to (-180, -151) degrees W and (50.5, 67.5) degrees N and reflect the average March temperature for each year across that region. March temperatures are chosen to estimate the conditions spawning plaice may have experienced, roughly two months before eggs appear in the water column.

##Descriptive Information: Alaska Plaice

Table 1: Descriptive Metrics for Alaska Plaice Egg Data

Lat Range	Lon Range	Day of Year Range	Bottom Depth Range
53.4-62.6	-178.4 to -158.2	100-181	0-150

Table 2: Descriptive Metrics for Alaska Plaice Larval Data

Lat Range	Lon Range	Day of Year Range	Bottom Depth Range
53.4-60.3	-176.8 to -158.2	99-175	23-150

The following three plots show *the day of year distribution for positive plaice egg catch*, *the year distribution for positive plaice egg catch.*

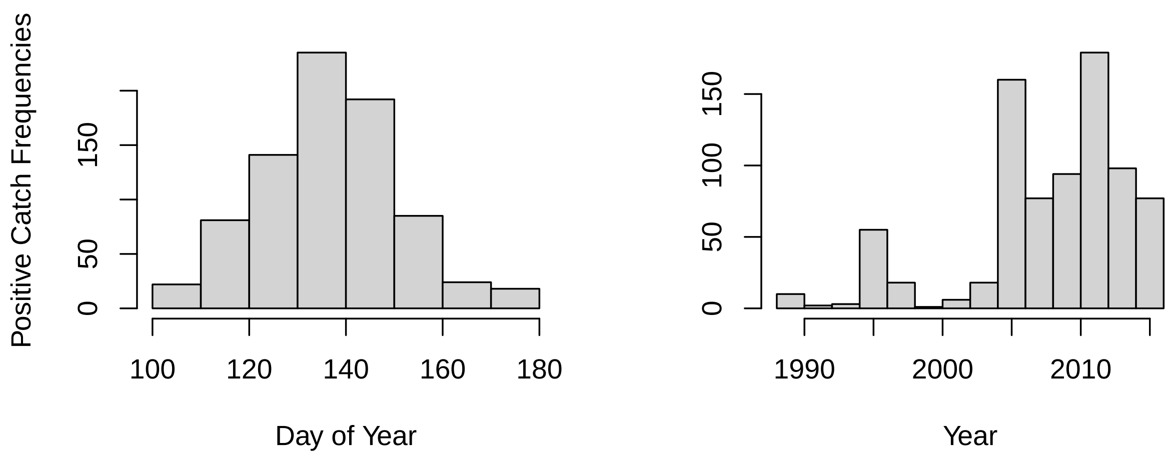
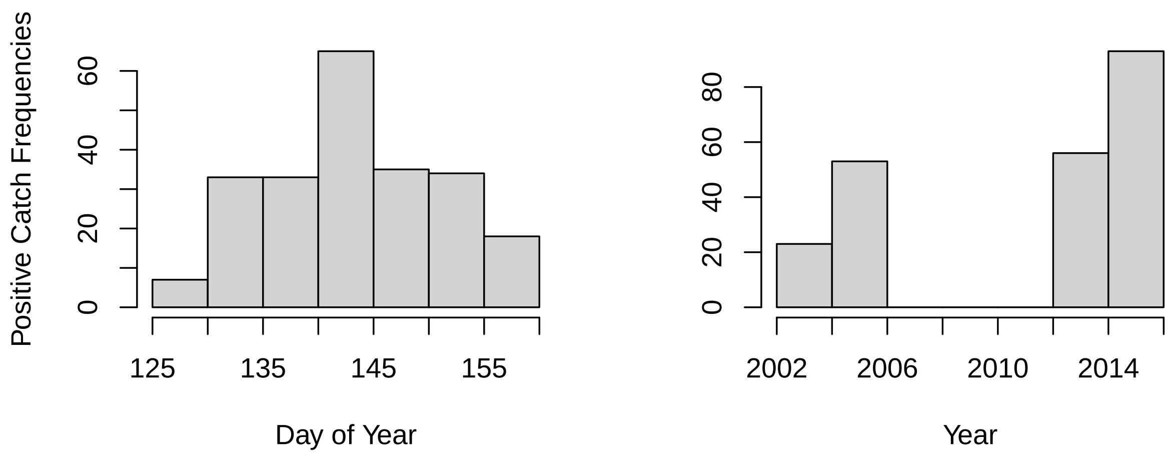


Figure 1: Alaska Plaice Eggs

The following three plots show *the day of year distribution for positive plaice larval catch*, *the year distribution for positive plaice larval catch.*



Now we'll move into the GAMs. The following code *is only necessary if the data were re-trimmed and new GAMs need to be run.* In this case, modify markdown document such that "{eval = TRUE}". The other model figures are marked as "eval = FALSE" if they, as of the last model run, do not produce the best model results. **Make sure to save the new models as RDS objects.**

Alaska plaice eggs were best explained by the threshold geography model, in which geographic distribution of eggs varied differently below and above 2.12 degrees Celsius.

Generalized Additive Models: Alaska Plaice Eggs

The base model formulation:

```
eg.base<-gam((Cper10m2+1)~factor(year)+s(lon,lat)+s(doy)+s(bottom_depth,k=5),
             data=apsub,family=tw(link='log'),method='REML')

plot(eg.base,shade=FALSE,page=1,seWithMean=TRUE,scheme=2,scale=0)

saveRDS(eg.base,file="./GAM Models/ap_egg_base.rds")
```

```
eg.base<-readRDS("./GAM Models/ap_egg_base.rds")
summary(eg.base)
```

```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(lon, lat) + s(doy) + s(bottom_depth,
##     k = 5)
##
## Parametric coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.48430    0.19302   7.690 2.27e-14 ***
## factor(year)1988    2.51986    0.34999   7.200 8.42e-13 ***
```

```
## factor(year)1991  1.24741    0.34361   3.630 0.000290 ***
## factor(year)1993  0.27485    0.40839   0.673 0.501026
## factor(year)1994  0.53487    0.30597   1.748 0.080590 .
## factor(year)1995  1.24746    0.22058   5.655 1.77e-08 ***
## factor(year)1996 -0.55882    0.76879  -0.727 0.467378
## factor(year)1997  1.40638    0.29741   4.729 2.41e-06 ***
## factor(year)1998  1.55691    0.52719   2.953 0.003181 **
## factor(year)1999  0.99087    0.32183   3.079 0.002105 **
## factor(year)2000 -0.18017    0.33304  -0.541 0.588588
## factor(year)2002  0.06059    0.26695   0.227 0.820460
## factor(year)2003 -0.60516    0.27978  -2.163 0.030654 *
## factor(year)2005  0.98181    0.23801   4.125 3.85e-05 ***
## factor(year)2006  1.84056    0.22610   8.140 6.76e-16 ***
## factor(year)2007  0.34209    0.23567   1.452 0.146776
## factor(year)2008  2.97926    0.23767  12.535  < 2e-16 ***
## factor(year)2009  1.69714    0.23260   7.296 4.20e-13 ***
## factor(year)2010  2.16159    0.23092   9.361  < 2e-16 ***
## factor(year)2011  1.26047    0.27689   4.552 5.62e-06 ***
## factor(year)2012  3.00854    0.21989  13.682  < 2e-16 ***
## factor(year)2013  1.76616    0.34681   5.093 3.86e-07 ***
## factor(year)2014 -0.26866    0.21224  -1.266 0.205709
## factor(year)2015 -0.01786    0.33167  -0.054 0.957061
## factor(year)2016 -0.78827    0.21521  -3.663 0.000256 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                    edf Ref.df      F  p-value
## s(lon,lat)      27.653 28.877 61.494  < 2e-16 ***
## s(doy)           8.232  8.830  8.864  < 2e-16 ***
## s(bottom_depth)  3.840  3.981 10.356 6.49e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  -0.112   Deviance explained = 69.6%
## -REML = 7266.1  Scale est. = 1.6052    n = 2116
```

```
AIC(eg.base)
```

```
## [1] 14493.3
```

The variable-coefficient geography formulation (in which geographic egg distributions vary differently in relation to regional SST indices).

```
vc.geo<-gam((Cper10m2+1)~factor(year)+s(lon,lat)+s(doy)+s(bottom_depth,k=5)+
              s(lon,lat,by=reg.SST),data=apsub,family=tw(link='log'),
            method='REML')

par(mfrow=c(1,2))
plot(vc.geo,select=1,scheme=2,too.far=0.025,shade=FALSE,
     seWithMean=TRUE,xlab='Longitude',ylab='Latitude',
     main='V-C ap Egg Flex Geo, Avg. Variation')
map("world",fill=T,col="snow4",add=T)
```

```
plot(vc.geo,select=4,scheme=2,too.far=0.025,shade=FALSE,
     xlab='Longitude',ylab='Latitude',seWithMean=TRUE,
     main='V-C Flex Geo, Deviation from Avg. Variation')
map("world",fill=T,col="snow4",add=T)

saveRDS(vc.geo,file="./GAM Models/ap_egg_vc_geo.rds")
```

```
vc.geo<-readRDS("./GAM Models/ap_egg_vc_geo.rds")
summary(vc.geo)
```

```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(lon, lat) + s(doy) + s(bottom_depth,
##     k = 5) + s(lon, lat, by = reg.SST)
##
## Parametric coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.0000     0.0000     NaN      NaN
## factor(year)1988   2.9493     0.3138   9.397  < 2e-16 ***
## factor(year)1991   1.9263     0.3093   6.228 5.74e-10 ***
## factor(year)1993   1.0768     0.3803   2.832  0.00468 **
## factor(year)1994   1.1778     0.2773   4.247 2.26e-05 ***
## factor(year)1995   1.7886     0.1886   9.486  < 2e-16 ***
## factor(year)1996  -0.3831     0.7446  -0.515  0.60694
## factor(year)1997   1.8229     0.2748   6.633 4.19e-11 ***
## factor(year)1998   1.5508     0.4990   3.108  0.00191 **
## factor(year)1999   1.7748     0.2833   6.264 4.56e-10 ***
## factor(year)2000   0.1486     0.3000   0.495  0.62047
## factor(year)2002   0.5994     0.2462   2.434  0.01501 *
## factor(year)2003   0.1332     0.2916   0.457  0.64780
## factor(year)2005   1.1989     0.2481   4.832 1.45e-06 ***
## factor(year)2006   2.3334     0.2126  10.974  < 2e-16 ***
## factor(year)2007   0.8780     0.1994   4.402 1.13e-05 ***
## factor(year)2008   3.0602     0.1796  17.040  < 2e-16 ***
## factor(year)2009   2.2112     0.1934  11.431  < 2e-16 ***
## factor(year)2010   2.4785     0.1795  13.809  < 2e-16 ***
## factor(year)2011   1.9424     0.2590   7.500 9.50e-14 ***
## factor(year)2012   3.4133     0.1469  23.229  < 2e-16 ***
## factor(year)2013   2.1195     0.3070   6.905 6.72e-12 ***
## factor(year)2014   0.3203     0.1974   1.622  0.10488
## factor(year)2015   0.3791     0.3401   1.115  0.26514
## factor(year)2016  -0.4679     0.2187  -2.139  0.03255 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                     edf Ref.df      F  p-value
## s(lon,lat)       27.235 28.574 24.116  < 2e-16 ***
## s(doy)            7.978  8.713  8.892  < 2e-16 ***
## s(bottom_depth)   3.829  3.982  7.570 3.22e-06 ***
```

```
## s(lon,lat):reg.SST 28.013 29.528 18.398  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Rank: 96/97
## R-sq.(adj) =  0.107   Deviance explained = 74.5%
## -REML = 7141.1  Scale est. = 1.3887    n = 2116
```

```
AIC(vc.geo)
```

```
## [1] 14079
```

The variable-coefficient phenology formulation (in which temporal (phenological) distribution of eggs vary in relation to regional SST indices).

```
vc.pheno<-gam((Cper10m2+1)~factor(year)+s(lon,lat)+s(doy)+s(bottom_depth,k=5)+
                s(doy,by=reg.SST),data=apsub,family=tw(link='log'),
              method='REML')

par(oma=c(1,1,1,0.5),mar=c(3,3,3,1.5))
plot(vc.pheno,select=2,main='Alaska Plaice VC Phenology, Eggs',seWithMean=TRUE,
     ylim=c(-25,11))
abline(h=0,col='mistyrose4',lty=2,lwd=1.3)
par(oma=c(1,1,1,0.5),mar=c(3,3,3,1.5),new=TRUE)
plot(vc.pheno,select=4,seWithMean=TRUE,shade=TRUE,shade.col=col,ylim=c(-25,11))
legend('topright',legend=c('Flexible Phenology Smooth','Deviation from Avg.Phenology'),
       col=c(NA,col),lwd=c(2,2),cex=0.8)
mtext(c("Day of Year","Anomalies in log(CPUE+1)"),side=c(1,2),line=2.5)

saveRDS(vc.pheno,file="./GAM Models/ap_egg_vc_pheno.rds")
```

```
vc.pheno<-readRDS("./GAM Models/ap_egg_vc_pheno.rds")
summary(vc.geo)
```

```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(lon, lat) + s(doy) + s(bottom_depth,
##     k = 5) + s(lon, lat, by = reg.SST)
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.0000     0.0000     NaN      NaN
## factor(year)1988    2.9493     0.3138   9.397  < 2e-16 ***
## factor(year)1991    1.9263     0.3093   6.228 5.74e-10 ***
## factor(year)1993    1.0768     0.3803   2.832  0.00468 **
## factor(year)1994    1.1778     0.2773   4.247 2.26e-05 ***
## factor(year)1995    1.7886     0.1886   9.486  < 2e-16 ***
## factor(year)1996   -0.3831     0.7446  -0.515  0.60694
## factor(year)1997    1.8229     0.2748   6.633 4.19e-11 ***
```

```
## factor(year)1998   1.5508     0.4990   3.108  0.00191 **
## factor(year)1999   1.7748     0.2833   6.264 4.56e-10 ***
## factor(year)2000   0.1486     0.3000   0.495  0.62047
## factor(year)2002   0.5994     0.2462   2.434  0.01501 *
## factor(year)2003   0.1332     0.2916   0.457  0.64780
## factor(year)2005   1.1989     0.2481   4.832 1.45e-06 ***
## factor(year)2006   2.3334     0.2126  10.974  < 2e-16 ***
## factor(year)2007   0.8780     0.1994   4.402 1.13e-05 ***
## factor(year)2008   3.0602     0.1796  17.040  < 2e-16 ***
## factor(year)2009   2.2112     0.1934  11.431  < 2e-16 ***
## factor(year)2010   2.4785     0.1795  13.809  < 2e-16 ***
## factor(year)2011   1.9424     0.2590   7.500 9.50e-14 ***
## factor(year)2012   3.4133     0.1469  23.229  < 2e-16 ***
## factor(year)2013   2.1195     0.3070   6.905 6.72e-12 ***
## factor(year)2014   0.3203     0.1974   1.622  0.10488
## factor(year)2015   0.3791     0.3401   1.115  0.26514
## factor(year)2016  -0.4679     0.2187  -2.139  0.03255 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                       edf Ref.df      F  p-value
## s(lon,lat)         27.235 28.574 24.116  < 2e-16 ***
## s(doy)              7.978  8.713  8.892  < 2e-16 ***
## s(bottom_depth)     3.829  3.982  7.570 3.22e-06 ***
## s(lon,lat):reg.SST 28.013 29.528 18.398  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Rank: 96/97
## R-sq.(adj) =  0.107   Deviance explained = 74.5%
## -REML = 7141.1  Scale est. = 1.3887    n = 2116
```

```
AIC(vc.geo)
```

```
## [1] 14079
```

The threshold phenology model formulation (in which the temporal (phenological) distribution of eggs vary differently above and below a threshold temperature:

```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(lon, lat) + s(bottom_depth,
##     k = 5) + s(doy, by = th)
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.57618    0.19640   8.025 1.69e-15 ***
## factor(year)1988   1.31835    0.35793   3.683 0.000236 ***
## factor(year)1991   0.42300    0.35765   1.183 0.237057
```

```
## factor(year)1993 -0.43793    0.41605  -1.053 0.292658
## factor(year)1994  0.07608    0.31659   0.240 0.810104
## factor(year)1995  0.95655    0.23566   4.059 5.11e-05 ***
## factor(year)1996 -0.74681    0.75435  -0.990 0.322292
## factor(year)1997  1.06124    0.30796   3.446 0.000580 ***
## factor(year)1998 -0.39828    0.45013  -0.885 0.376360
## factor(year)1999  2.75194    0.37170   7.404 1.92e-13 ***
## factor(year)2000 -0.24398    0.32485  -0.751 0.452705
## factor(year)2002 -0.08724    0.26301  -0.332 0.740154
## factor(year)2003 -0.48672    0.26607  -1.829 0.067505 .
## factor(year)2005  0.88985    0.23206   3.835 0.000130 ***
## factor(year)2006  1.59445    0.23084   6.907 6.58e-12 ***
## factor(year)2007  0.78550    0.24772   3.171 0.001542 **
## factor(year)2008  2.64464    0.24875  10.632  < 2e-16 ***
## factor(year)2009  1.29365    0.24627   5.253 1.65e-07 ***
## factor(year)2010  1.74450    0.24494   7.122 1.46e-12 ***
## factor(year)2011  1.65432    0.25707   6.435 1.53e-10 ***
## factor(year)2012  2.20870    0.23626   9.348  < 2e-16 ***
## factor(year)2013  1.14987    0.35788   3.213 0.001334 **
## factor(year)2014 -0.14051    0.19818  -0.709 0.478408
## factor(year)2015 -0.02122    0.33136  -0.064 0.948937
## factor(year)2016 -0.66457    0.20219  -3.287 0.001030 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                     edf Ref.df      F p-value
## s(lon,lat)       27.595 28.870 61.640  <2e-16 ***
## s(bottom_depth)   3.918  3.994 19.240  <2e-16 ***
## s(doy):thFALSE    1.000  1.001  2.887  0.0894 .
## s(doy):thTRUE     8.447  8.911 28.872  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  -0.00298   Deviance explained = 70.8%
## -REML = 7216.7  Scale est. = 1.5489    n = 2116
```

```
## [1] 14384.02
```

The threshold geography model formulation (in which the geographic distribution of eggs vary differently above and below a threshold temperature: *This is the best model to explain Alaska plaice egg variation across years, as of 1/10/2021.*

```
thr.geo<-readRDS("./GAM Models/ap_egg_thr_geo.rds")
best.index.geo<-readRDS("./GAM Models/ap_egg_best_index_geo.rds")
aic.geo<-readRDS("./GAM Models/ap_egg_aic_geo_list.rds")
summary(thr.geo)
```

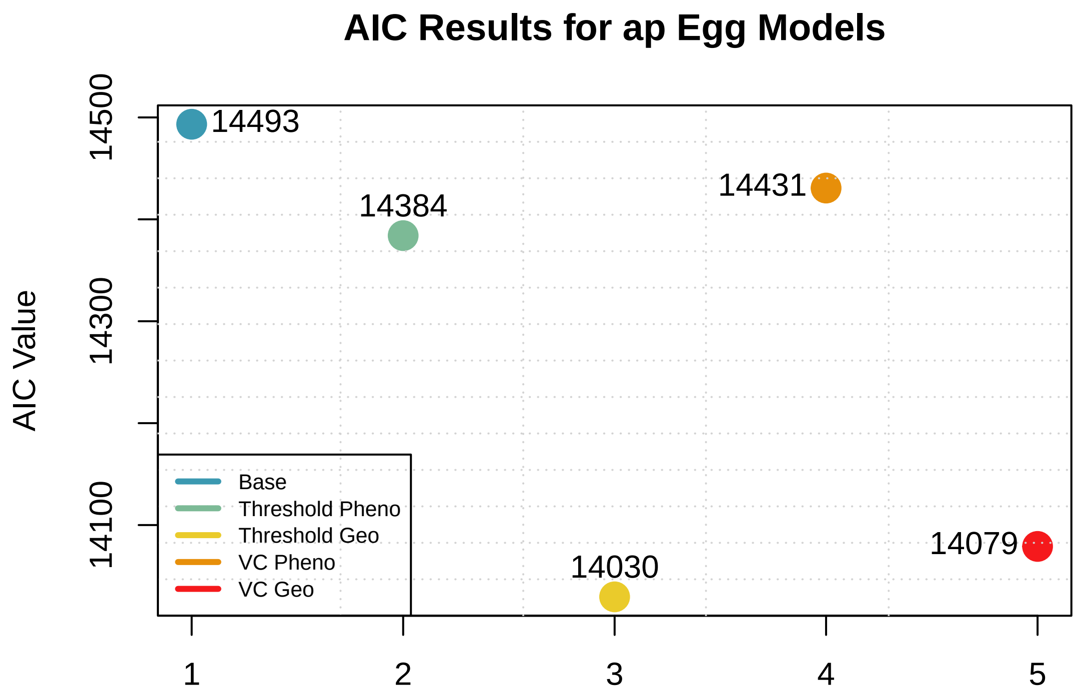
```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
```

```
## (Cper10m2 + 1) ~ factor(year) + s(doy) + s(bottom_depth, k = 5) +
##     s(lon, lat, by = th)
##
## Parametric coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.02544    0.18659   5.496 4.38e-08 ***
## factor(year)1988  2.61621    0.33468   7.817 8.63e-15 ***
## factor(year)1991  2.13895    0.32950   6.491 1.07e-10 ***
## factor(year)1993  0.98887    0.39266   2.518 0.011865 *
## factor(year)1994  0.96755    0.29909   3.235 0.001236 **
## factor(year)1995  1.82392    0.21488   8.488  < 2e-16 ***
## factor(year)1996 -0.28811    0.71833  -0.401 0.688403
## factor(year)1997  1.87579    0.27967   6.707 2.57e-11 ***
## factor(year)1998  1.97600    0.51276   3.854 0.000120 ***
## factor(year)1999  1.82942    0.30764   5.947 3.22e-09 ***
## factor(year)2000 -0.15258    0.31459  -0.485 0.627729
## factor(year)2002  0.68175    0.25963   2.626 0.008707 **
## factor(year)2003 -0.06705    0.27223  -0.246 0.805485
## factor(year)2005  1.30594    0.23060   5.663 1.70e-08 ***
## factor(year)2006  2.36688    0.21786  10.864  < 2e-16 ***
## factor(year)2007  0.79356    0.23138   3.430 0.000616 ***
## factor(year)2008  2.57392    0.22510  11.435  < 2e-16 ***
## factor(year)2009  1.96451    0.22090   8.893  < 2e-16 ***
## factor(year)2010  2.24061    0.22189  10.098  < 2e-16 ***
## factor(year)2011  1.81458    0.26431   6.865 8.80e-12 ***
## factor(year)2012  3.08484    0.21230  14.530  < 2e-16 ***
## factor(year)2013  2.03422    0.33288   6.111 1.18e-09 ***
## factor(year)2014  0.04052    0.20691   0.196 0.844762
## factor(year)2015  0.31682    0.31267   1.013 0.311050
## factor(year)2016 -0.50510    0.20715  -2.438 0.014843 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                       edf Ref.df     F p-value
## s(doy)              8.240  8.831 11.96  <2e-16 ***
## s(bottom_depth)     3.907  3.991 15.08  <2e-16 ***
## s(lon,lat):thFALSE 27.091 28.760 51.96  <2e-16 ***
## s(lon,lat):thTRUE  27.763 28.875 39.36  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.268   Deviance explained =   75%
## -REML = 7080.7  Scale est. = 1.3654    n = 2116
```

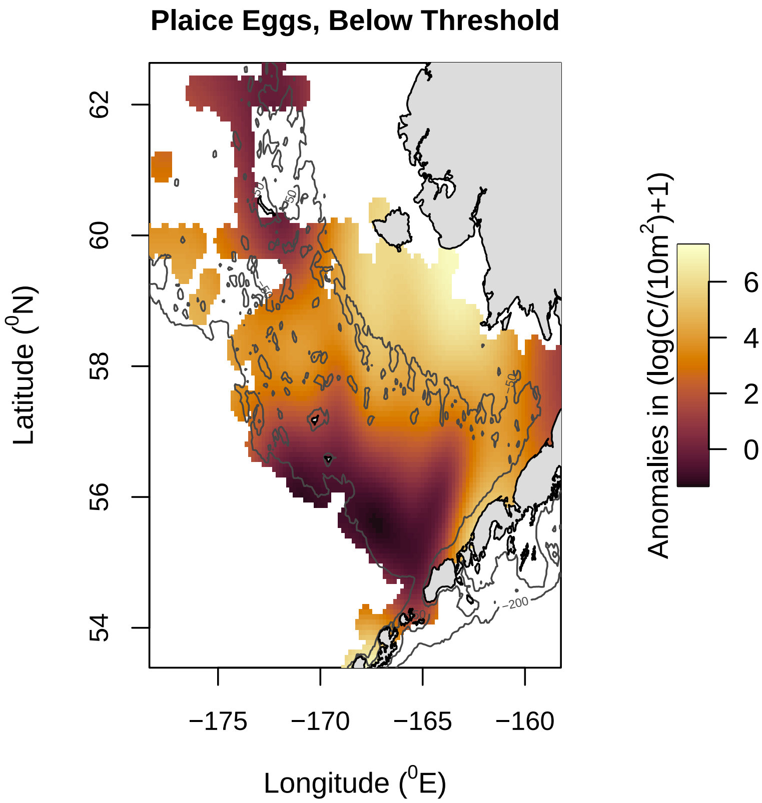
```
AIC(thr.geo)
```

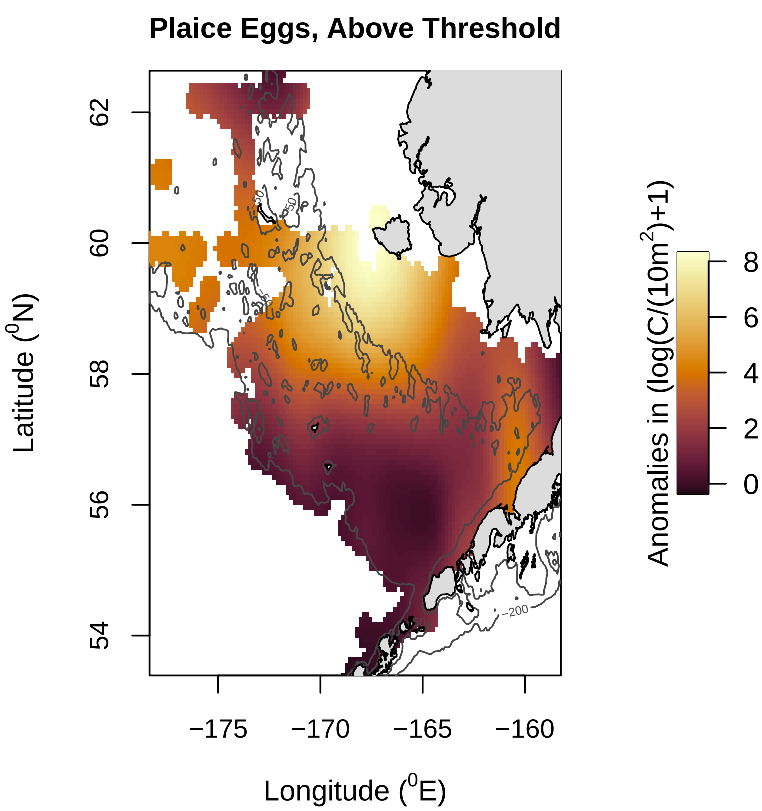
```
## [1] 14029.54
```

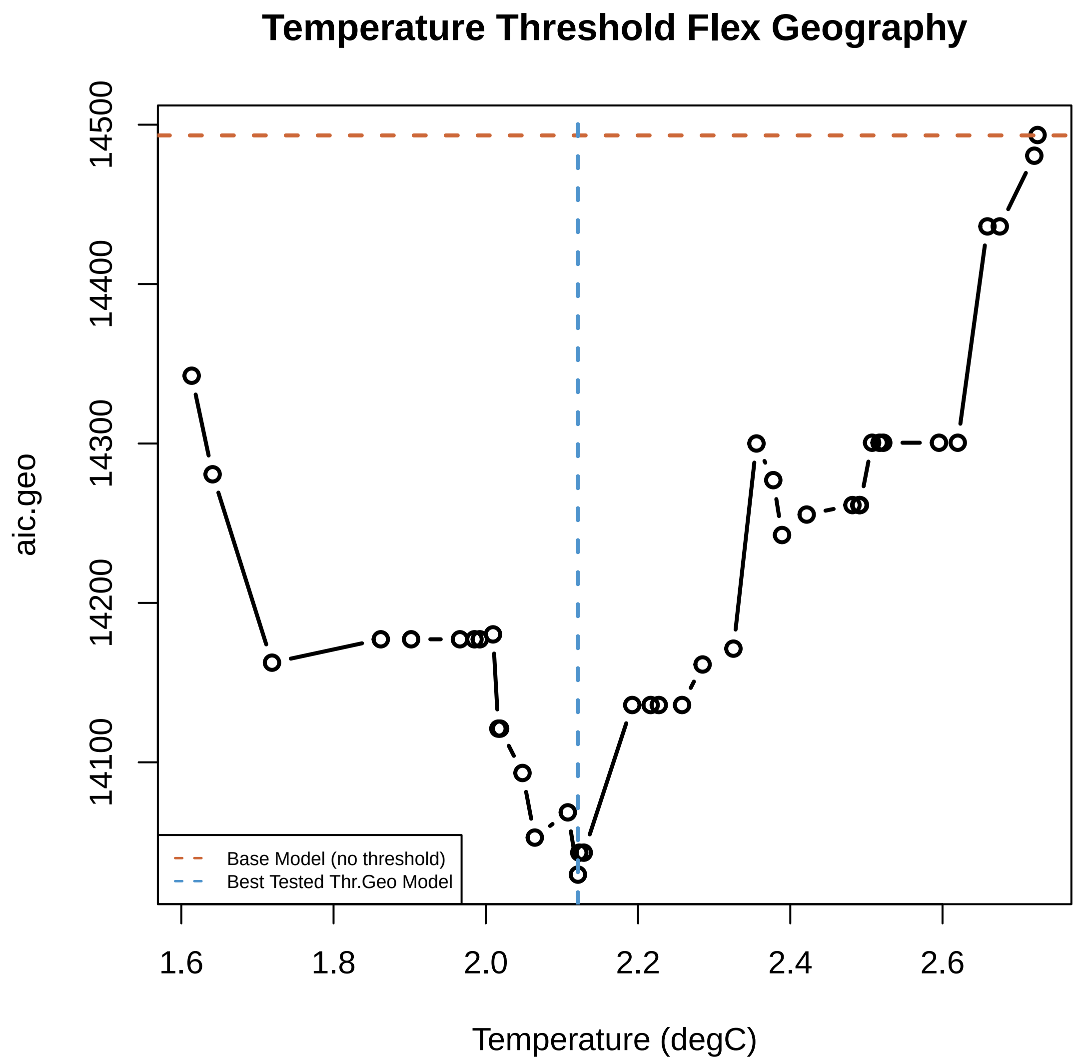
To confirm that this is indeed the best model, we can compare AIC values across all five tested models.



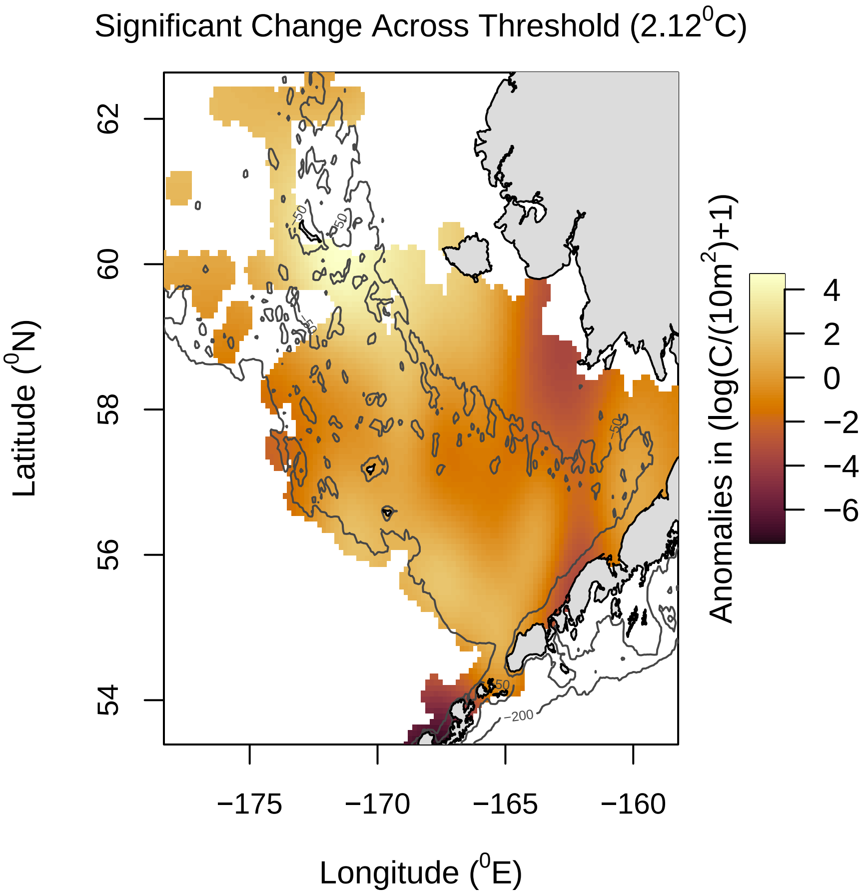
This is the below threshold (2.12 deg C) and above threshold predicted geographical distribution of Alaska plaice eggs (based on the threshold geography model).


Plaice Eggs, Below Threshold
Latitude (⁰N)
62
60
58
56
54
−175
−170
−165
−160
Longitude (⁰E)
Anomalies in (log(C/(10m²)+1)
0
2
4
6
−200


Plaice Eggs, Above Threshold
Latitude (⁰N)
62
60
58
56
54
−175
−170
−165
−160
Longitude (⁰E)
Anomalies in (log(C/(10m²)+1)
0
2
4
6
8
−200



With the threshold geography model, we can predict places where there is a significant *difference* in predictions when comparing the above threshold spatial prediction to the below threshold spatial prediction. This plot is calculated by subtracting the spatial distribution *below* the threshold from the prediction *above* the threshold.



Larval Generalized Additive Models:

The following code *is only necessary if the data were re-trimmed and new GAMs need to be run.* In this case, modify markdown document such that "{eval = TRUE}". The other model figures are marked as "eval = FALSE" if they, as of the last model run, do not produce the best model results. These models are produced using conductivity-temperature-depth derived temperature and salinity measurements.

We begin with the base larval model:

```
lv.base<-gam((Cper10m2+1)~factor(year)+s(doy,k=7)+s(lon,lat)+
               s(bottom_depth,k=5),
             data=aplarv.ctd,family=tw(link='log'),method='REML')

saveRDS(lv.base,file="./GAM Models/ap_larval_base.rds")
```

```
lv.base<-readRDS("./GAM Models/ap_larval_base.rds")
summary(lv.base)
```

```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
```

```
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(doy, k = 7) + s(lon, lat) +
##     s(bottom_depth, k = 5)
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.63411    0.23912   2.652  0.00810 **
## factor(year)1998  1.16038    0.50728   2.287  0.02233 *
## factor(year)1999  1.21284    0.37542   3.231  0.00127 **
## factor(year)2000 -0.05472    0.33479  -0.163  0.87018
## factor(year)2002  0.09467    0.27540   0.344  0.73108
## factor(year)2003  0.93324    0.28221   3.307  0.00097 ***
## factor(year)2005  1.41838    0.26149   5.424 6.95e-08 ***
## factor(year)2006  0.35275    0.25632   1.376  0.16899
## factor(year)2007  0.53981    0.28540   1.891  0.05879 .
## factor(year)2008  0.18211    0.36928   0.493  0.62199
## factor(year)2009 -0.19599    0.26305  -0.745  0.45636
## factor(year)2010  0.02927    0.26021   0.112  0.91046
## factor(year)2011 -0.46593    0.33389  -1.395  0.16313
## factor(year)2012 -0.39148    0.25059  -1.562  0.11848
## factor(year)2013 -0.01145    0.30574  -0.037  0.97012
## factor(year)2014  0.45777    0.25278   1.811  0.07038 .
## factor(year)2015  0.19521    0.30067   0.649  0.51630
## factor(year)2016  0.61153    0.25228   2.424  0.01549 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                     edf Ref.df     F p-value
## s(doy)            3.323  4.108 18.56  <2e-16 ***
## s(lon,lat)       27.760 28.890 32.43  <2e-16 ***
## s(bottom_depth)  3.880  3.988 11.70  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   -9.8   Deviance explained = 78.8%
## -REML = 2731.6  Scale est. = 0.81516   n = 1338
```

Then we add in additive salinity:

```
lv.add.sal<-gam((Cper10m2+1)~factor(year)+s(doy,k=7)+s(lon,lat)+
                  s(bottom_depth,k=5)+
                  s(salinity),data=aplarv.ctd,family=tw(link='log'),
                method='REML')

saveRDS(lv.add.sal,file="./GAM Models/ap_larval_addsal.rds")
```

```
lv.add.sal<-readRDS("./GAM Models/ap_larval_addsal.rds")
summary(lv.add.sal)
```

```
##
## Family: Tweedie(p=1.99)
## Link function: log
```

```
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(doy, k = 7) + s(lon, lat) +
##     s(bottom_depth, k = 5) + s(salinity)
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.64775    0.24189   2.678 0.007505 **
## factor(year)1998  1.16866    0.50452   2.316 0.020695 *
## factor(year)1999  1.16481    0.37864   3.076 0.002141 **
## factor(year)2000 -0.13485    0.33837  -0.399 0.690299
## factor(year)2002  0.03407    0.28176   0.121 0.903775
## factor(year)2003  0.95278    0.28309   3.366 0.000786 ***
## factor(year)2005  1.40458    0.26604   5.280 1.52e-07 ***
## factor(year)2006  0.34278    0.26053   1.316 0.188513
## factor(year)2007  0.53085    0.28834   1.841 0.065843 .
## factor(year)2008  0.19844    0.37106   0.535 0.592888
## factor(year)2009 -0.19231    0.26608  -0.723 0.469964
## factor(year)2010  0.05271    0.26128   0.202 0.840158
## factor(year)2011 -0.46456    0.33362  -1.392 0.164019
## factor(year)2012 -0.39331    0.25120  -1.566 0.117663
## factor(year)2013  0.03260    0.30771   0.106 0.915638
## factor(year)2014  0.43796    0.25627   1.709 0.087698 .
## factor(year)2015  0.19262    0.30253   0.637 0.524429
## factor(year)2016  0.56619    0.25722   2.201 0.027903 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                     edf Ref.df      F  p-value
## s(doy)            3.223  3.994 19.173  < 2e-16 ***
## s(lon,lat)       27.761 28.887 32.520  < 2e-16 ***
## s(bottom_depth)   3.871  3.986 10.926 5.83e-07 ***
## s(salinity)       3.159  4.063  1.538    0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =    -10   Deviance explained = 78.9%
## -REML = 2732.9  Scale est. = 0.81295   n = 1338
```

Then additive temperature:

```
lv.add.temp<-gam((Cper10m2+1)~factor(year)+s(doy,k=7)+s(lon,lat)+
                   s(bottom_depth,k=5)+
                   s(temperature),data=aplarv.ctd,family=tw(link='log'),
                 method='REML')

saveRDS(lv.add.temp,file="./GAM Models/ap_larval_addtemp.rds")
```

```
lv.add.temp<-readRDS("./GAM Models/ap_larval_addtemp.rds")
summary(lv.add.temp)
```

```
##
```

```
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(doy, k = 7) + s(lon, lat) +
##     s(bottom_depth, k = 5) + s(temperature)
##
## Parametric coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.59194    0.23169   2.555  0.01074 *
## factor(year)1998  0.64599    0.39297   1.644  0.10045
## factor(year)1999  0.91984    0.30948   2.972  0.00301 **
## factor(year)2000 -0.24128    0.31035  -0.777  0.43703
## factor(year)2002 -0.39314    0.27942  -1.407  0.15967
## factor(year)2003  0.66580    0.28516   2.335  0.01971 *
## factor(year)2005  1.07189    0.27233   3.936 8.73e-05 ***
## factor(year)2006  0.48049    0.24263   1.980  0.04788 *
## factor(year)2007  0.57692    0.27565   2.093  0.03655 *
## factor(year)2008  0.62522    0.37181   1.682  0.09290 .
## factor(year)2009 -0.13711    0.25494  -0.538  0.59079
## factor(year)2010  0.19428    0.25501   0.762  0.44630
## factor(year)2011 -0.05887    0.31918  -0.184  0.85369
## factor(year)2012 -0.10688    0.25165  -0.425  0.67111
## factor(year)2013  0.45213    0.30787   1.469  0.14219
## factor(year)2014  0.17342    0.25873   0.670  0.50281
## factor(year)2015 -0.12128    0.29652  -0.409  0.68259
## factor(year)2016  0.84714    0.28642   2.958  0.00316 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                    edf Ref.df      F  p-value
## s(doy)           1.471  1.819 16.370 1.23e-06 ***
## s(lon,lat)      27.472 28.835 26.441  < 2e-16 ***
## s(bottom_depth)  3.841  3.980  9.853 1.38e-06 ***
## s(temperature)   6.688  7.861 21.001  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  -0.645   Deviance explained = 80.6%
## -REML = 2673.4  Scale est. = 0.75576   n = 1338
```

Then additive temperature and salinity, in individual additive terms:

```
lv.temp.sal<-gam((Cper10m2+1)~factor(year)+s(doy,k=7)+s(lon,lat)+
                   s(bottom_depth,k=5)+
                   s(temperature)+s(salinity),data=aplarv.ctd,
                 family=tw(link='log'),method='REML')

saveRDS(lv.temp.sal,file="./GAM Models/ap_larval_addtempsal.rds")
```

```
lv.temp.sal<-readRDS("./GAM Models/ap_larval_addtempsal.rds")
summary(lv.temp.sal)
```

```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(doy, k = 7) + s(lon, lat) +
##     s(bottom_depth, k = 5) + s(temperature) + s(salinity)
##
## Parametric coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.43294    0.23419   1.849 0.064744 .
## factor(year)1998  0.59330    0.36160   1.641 0.101091
## factor(year)1999  1.10130    0.28911   3.809 0.000146 ***
## factor(year)2000 -0.09667    0.31141  -0.310 0.756280
## factor(year)2002 -0.16055    0.28475  -0.564 0.572977
## factor(year)2003  0.75631    0.28541   2.650 0.008150 **
## factor(year)2005  1.28050    0.27567   4.645 3.75e-06 ***
## factor(year)2006  0.70559    0.24686   2.858 0.004328 **
## factor(year)2007  0.79860    0.27840   2.868 0.004192 **
## factor(year)2008  0.89967    0.37626   2.391 0.016943 *
## factor(year)2009  0.04515    0.25874   0.174 0.861510
## factor(year)2010  0.26848    0.25527   1.052 0.293113
## factor(year)2011  0.05109    0.31708   0.161 0.872018
## factor(year)2012  0.05492    0.25196   0.218 0.827487
## factor(year)2013  0.59457    0.30990   1.919 0.055258 .
## factor(year)2014  0.26661    0.26119   1.021 0.307579
## factor(year)2015 -0.05541    0.29721  -0.186 0.852124
## factor(year)2016  1.02048    0.28889   3.532 0.000426 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                     edf Ref.df      F  p-value
## s(doy)            1.000  1.000 17.740 2.76e-05 ***
## s(lon,lat)       27.319 28.800 25.201  < 2e-16 ***
## s(bottom_depth)   3.839  3.979  9.219 3.39e-06 ***
## s(temperature)    6.824  7.974 23.938  < 2e-16 ***
## s(salinity)       4.604  5.768  4.204 0.000471 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  -0.196   Deviance explained = 80.9%
## -REML = 2668.3  Scale est. = 0.74691   n = 1338
```

And finally, the best performing model: the bivariate salinity-temperature additive term:

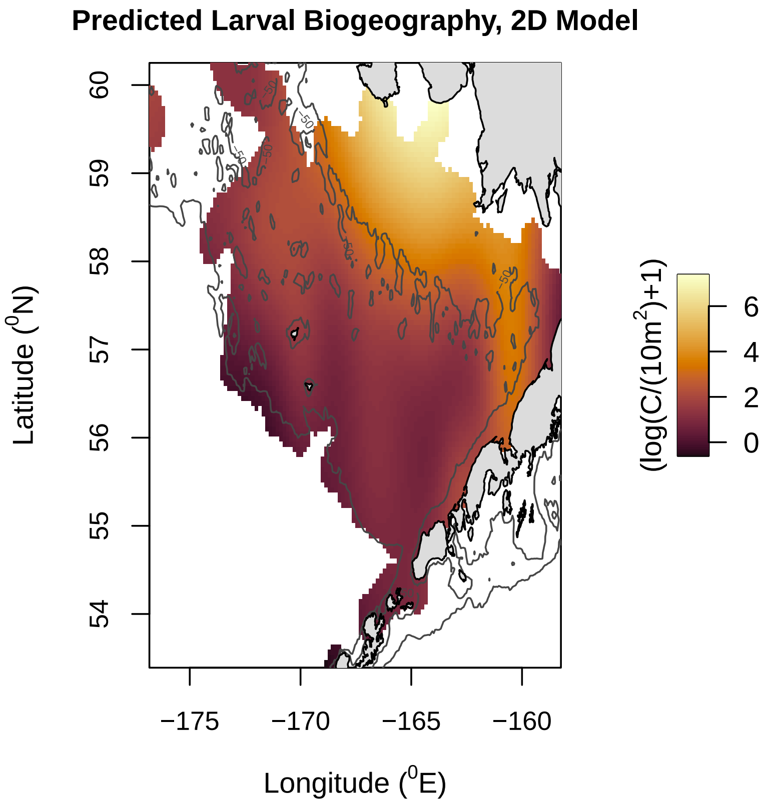
```
lv.2d<-gam((Cper10m2+1)~factor(year)+s(lon,lat)+s(doy,k=7)+s(bottom_depth)+
             s(salinity,temperature),data=aplarv.ctd,family=tw(link='log'),
           method='REML')

saveRDS(lv.2d,file="./GAM Models/ap_larval_2d.rds")
```

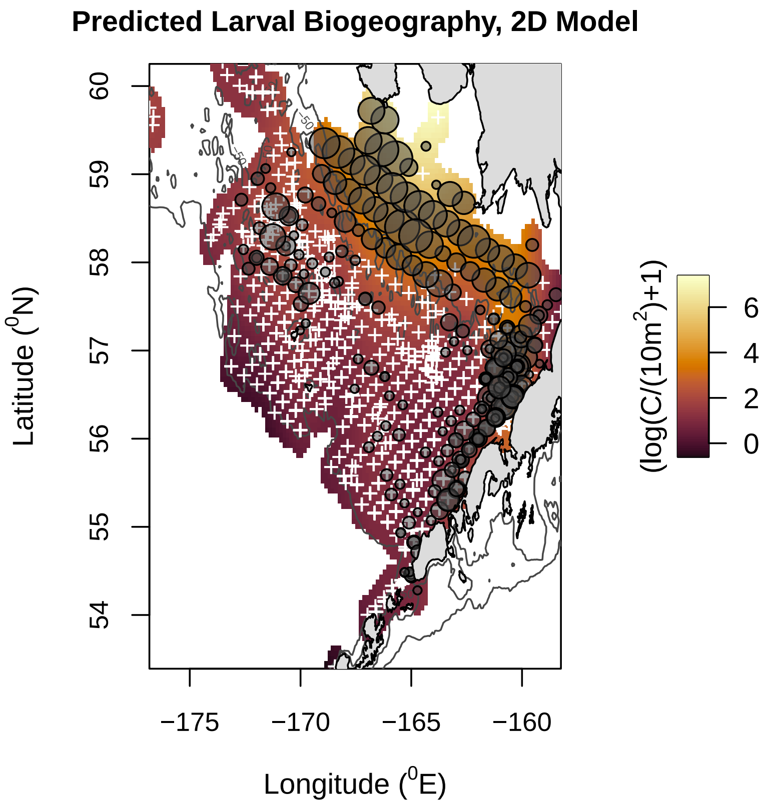
```r
lv.2d<-readRDS("./GAM Models/ap_larval_2d.rds")
summary(lv.2d)
```

```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(lon, lat) + s(doy, k = 7) +
##     s(bottom_depth) + s(salinity, temperature)
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.7316     0.2324   3.148 0.001685 **
## factor(year)1998    0.7555     0.3983   1.897 0.058112 .
## factor(year)1999    0.8160     0.3176   2.569 0.010307 *
## factor(year)2000   -0.3541     0.3107  -1.140 0.254508
## factor(year)2002   -0.6627     0.2913  -2.275 0.023086 *
## factor(year)2003    0.2463     0.2890   0.852 0.394182
## factor(year)2005    0.9614     0.2748   3.499 0.000484 ***
## factor(year)2006    0.4274     0.2471   1.730 0.083871 .
## factor(year)2007    0.3203     0.2808   1.141 0.254250
## factor(year)2008    0.5872     0.3747   1.567 0.117339
## factor(year)2009   -0.1899     0.2559  -0.742 0.458355
## factor(year)2010    0.1983     0.2486   0.798 0.425249
## factor(year)2011   -0.3025     0.3127  -0.967 0.333639
## factor(year)2012   -0.2040     0.2520  -0.810 0.418328
## factor(year)2013    0.3103     0.3070   1.011 0.312447
## factor(year)2014   -0.1790     0.2596  -0.690 0.490593
## factor(year)2015   -0.3525     0.2906  -1.213 0.225391
## factor(year)2016    0.6020     0.2856   2.108 0.035238 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                              edf Ref.df      F  p-value
## s(lon,lat)                26.690 28.615 20.489  < 2e-16 ***
## s(doy)                     1.615  2.031 12.330 4.76e-06 ***
## s(bottom_depth)            8.086  8.765 12.668  < 2e-16 ***
## s(salinity,temperature) 23.212 27.085  9.794  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  -0.209   Deviance explained = 82.7%
## -REML = 2621.1  Scale est. = 0.68704   n = 1338
```

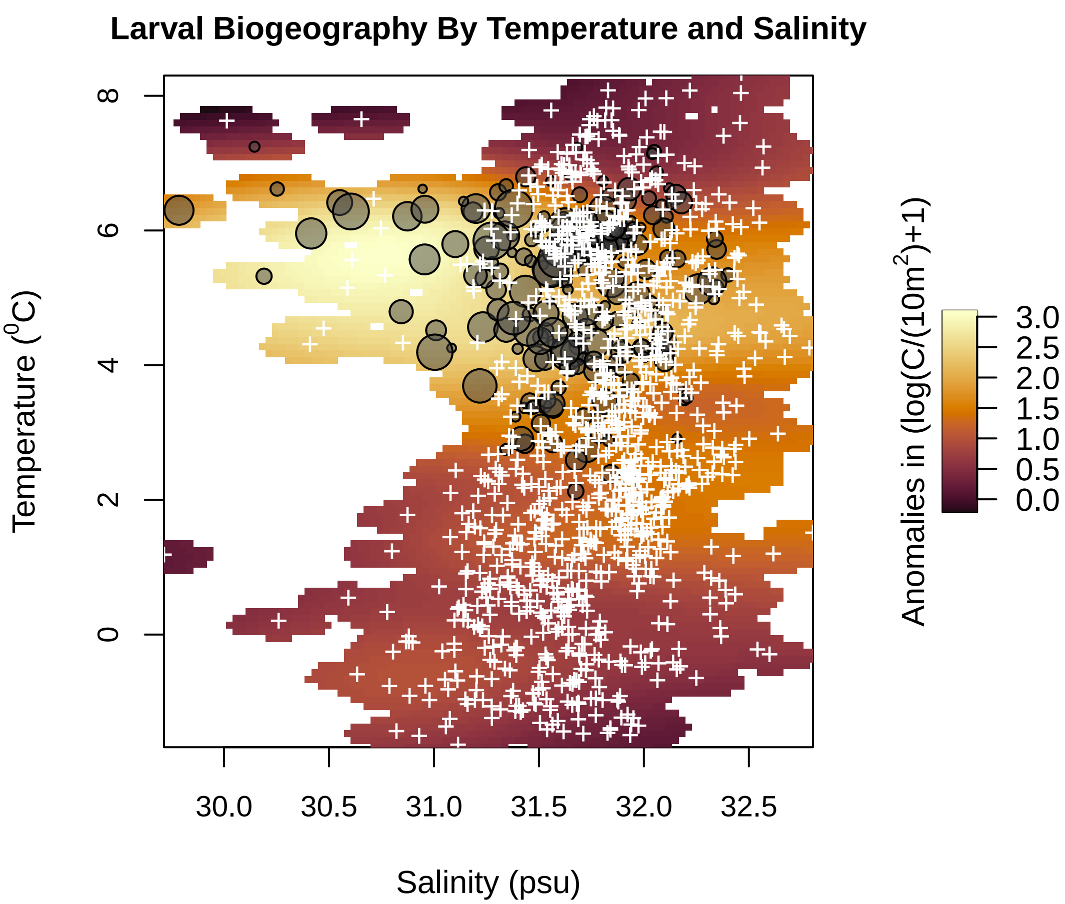
To confirm that this is indeed the best model, we can compare AIC values across all five tested models.

The following plot is the predicted Alaska Plaice larval biogeography based on the best performing model, the bivariate salinity-temperature GAM, next to the same plot with log(CPUE+1) observations shown.


Predicted Larval Biogeography, 2D Model
Latitude (0N)
60
59
58
57
56
55
54
−175
−170
−165
−160
Longitude (0E)
(log(C/(10m2)+1)
6
4
2
0


Predicted Larval Biogeography, 2D Model
Latitude (0N)
60
59
58
57
56
55
54
−175
−170
−165
−160
Longitude (0E)
(log(C/(10m2)+1)
6
4
2
0

With this bivariate model, we can also calculate the predicted anomalous larval catch (more or less than expected) on a salinity-temperature plot. This figure shows that prediction, with observed larval catch (log(n=1)) overlaid.



To again share the improvements of the best performing models from the base models, we can look at the AIC division produces.

Table 3: Model Power through Comparisons, Alaska plaice

	Best Divided By Base	Best Divided By Second Best
Eggs	0.9680024	0.9964871
Larvae	0.9522328	0.9772080