Streamlined Alaska Plaice Analyses & Figures

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.

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<182)

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 165 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	99-180	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.*

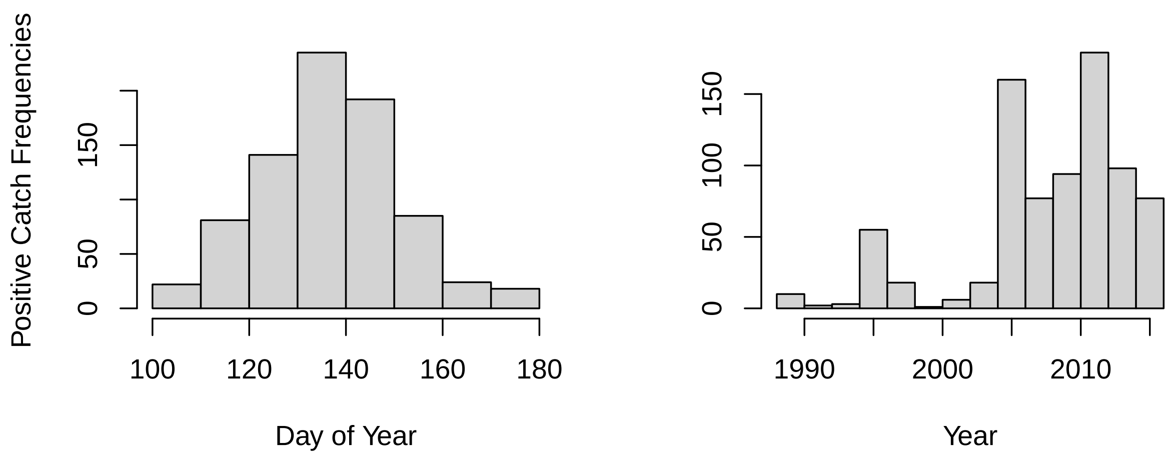
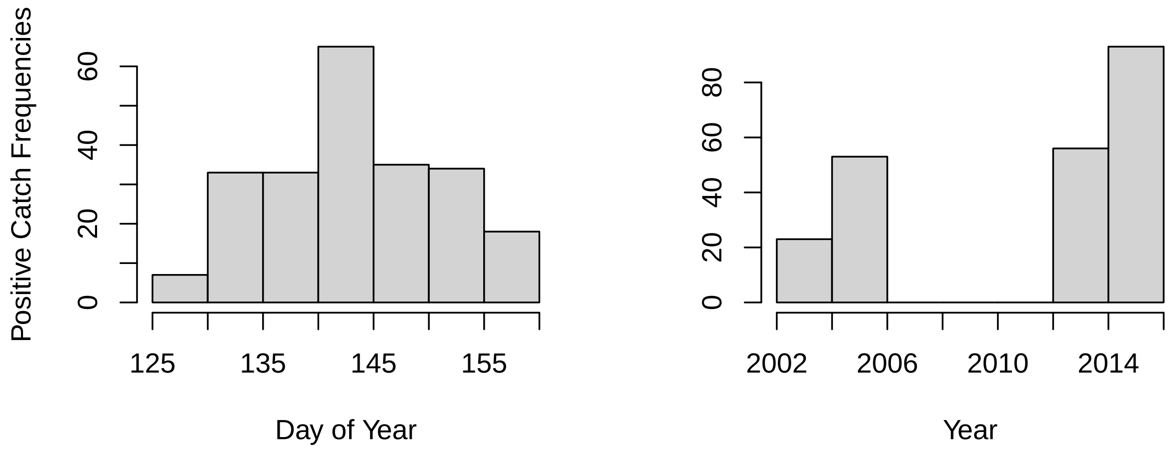
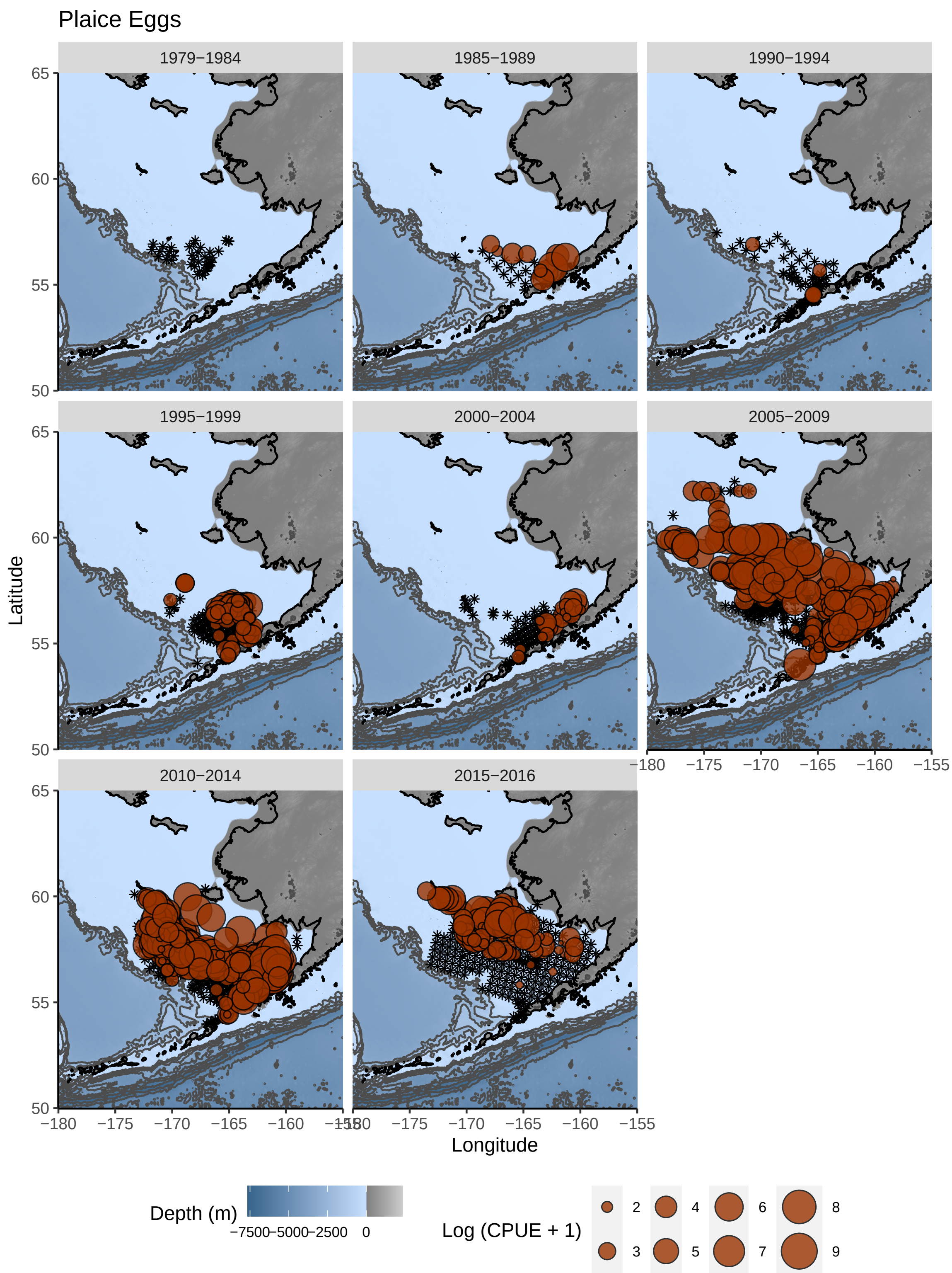


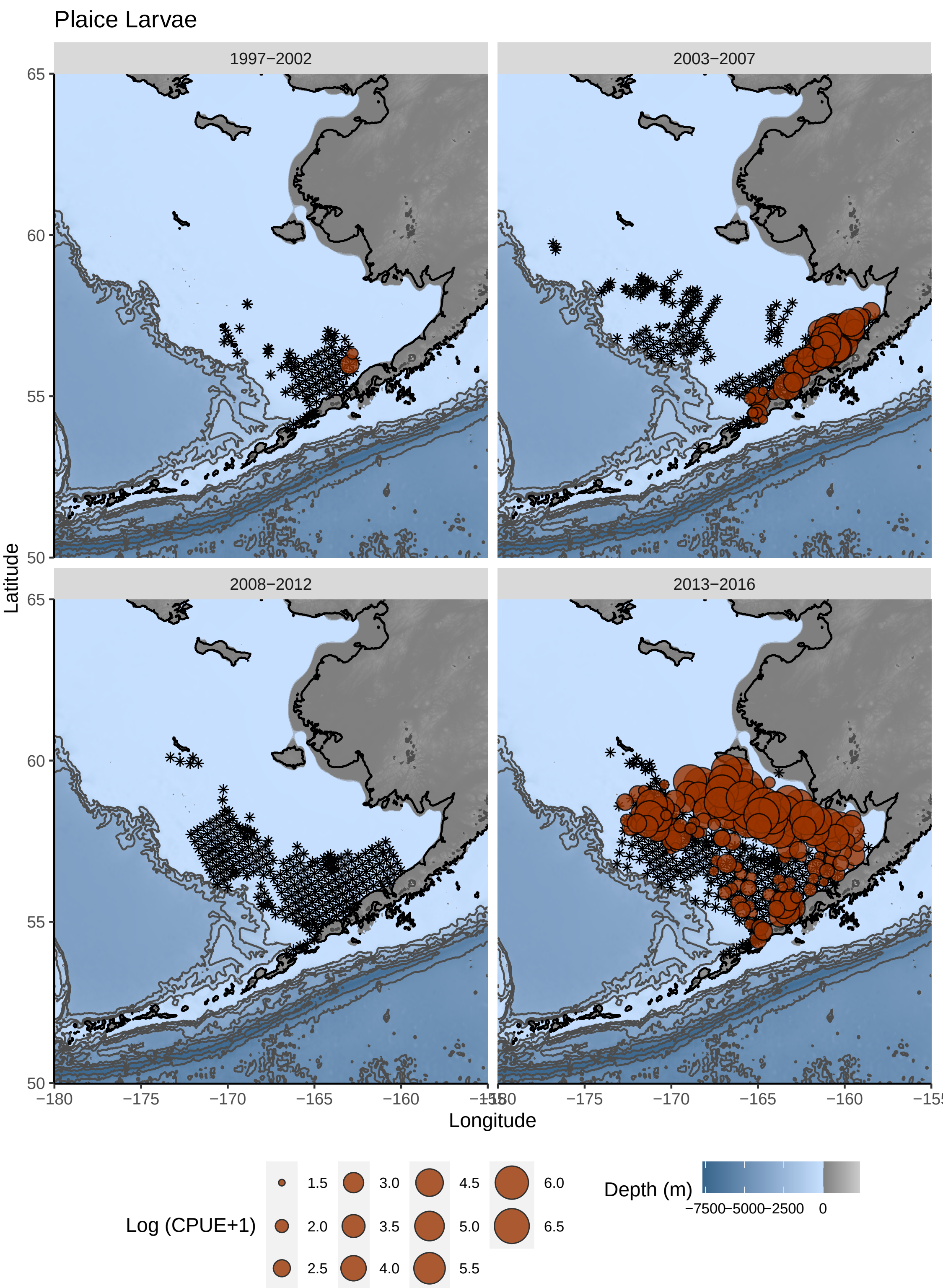
Figure 1: Alaska Plaice Eggs

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



The following plots show Plaice egg and larval catch distributions (Catch per unit effort, or per 10m^2) across five year increments from 1979 to 2016.


Plaice Eggs
1979–1984
1985–1989
1990–1994
1995–1999
2000–2004
2005–2009
2010–2014
2015–2016
Latitude
Longitude
65
60
55
50
−180
−175
−170
−165
−160
−155
Depth (m)
−7500−5000−2500 0
Log (CPUE + 1)
2
3
4
5
6
7
8
9



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")
```

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

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")
```

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

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
```

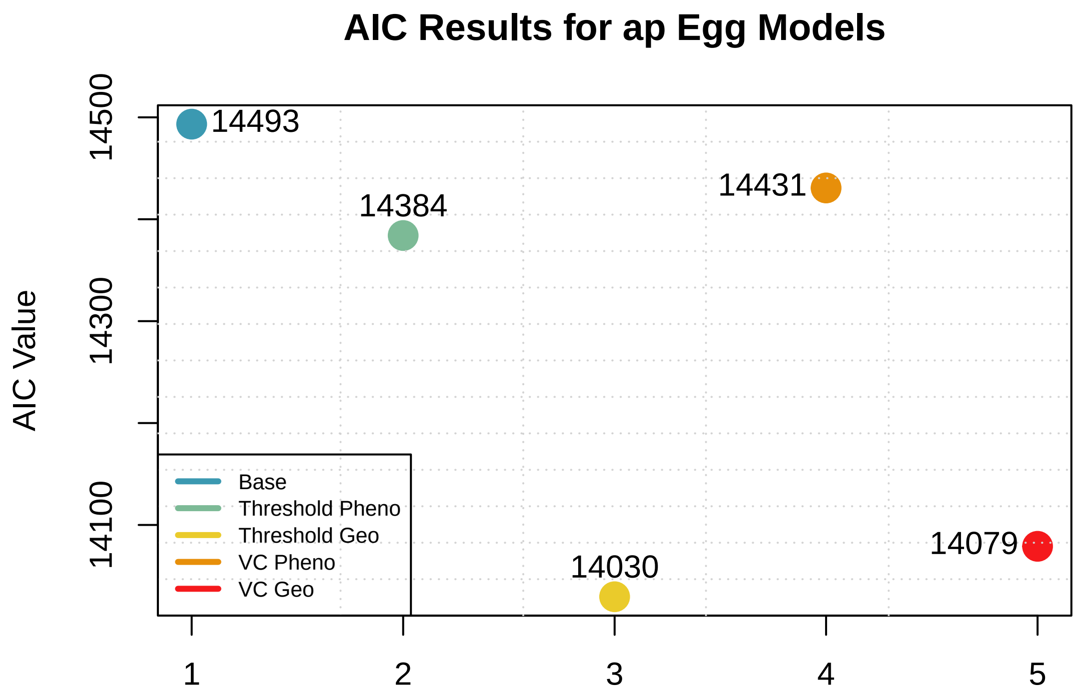
```
AIC(thr.geo)
```

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

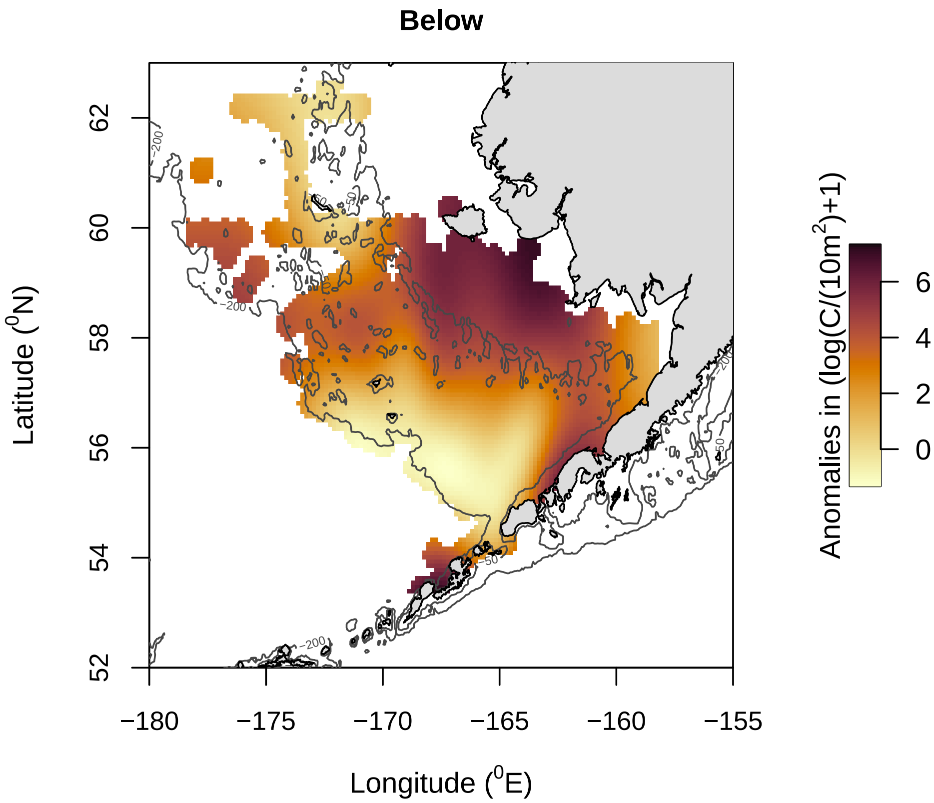
```
print(temps.in[[best.index.geo]])
```

```
## [1] 2.121005
```

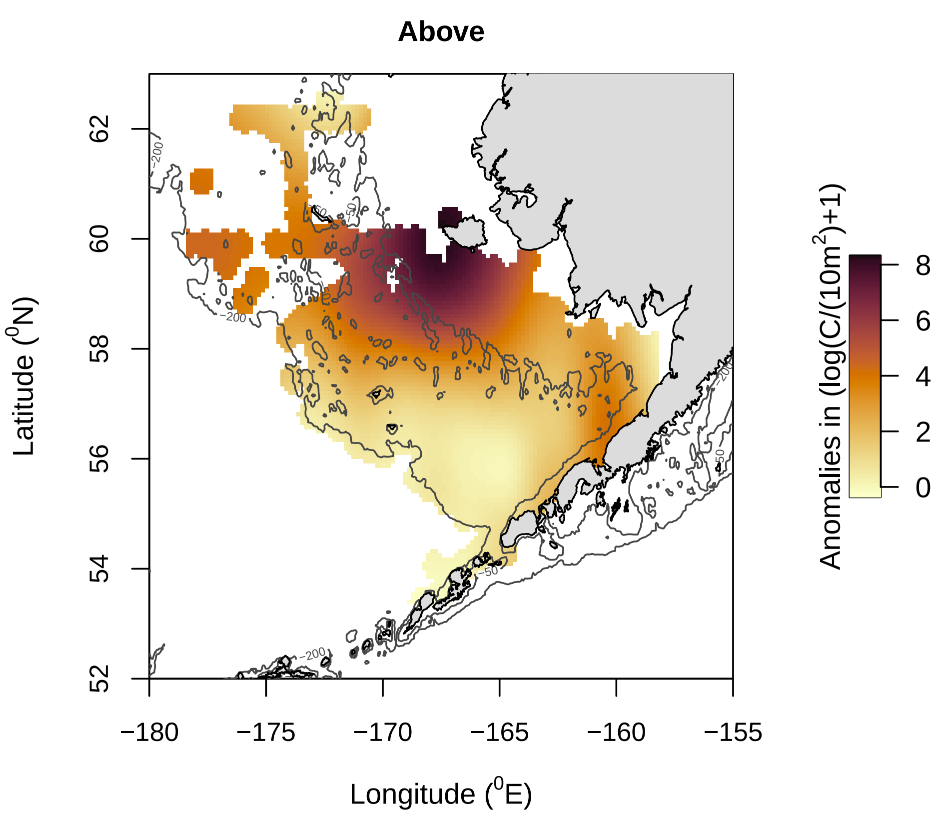
To confirm that this is indeed the best model, we can compare AIC values across all five tested models.



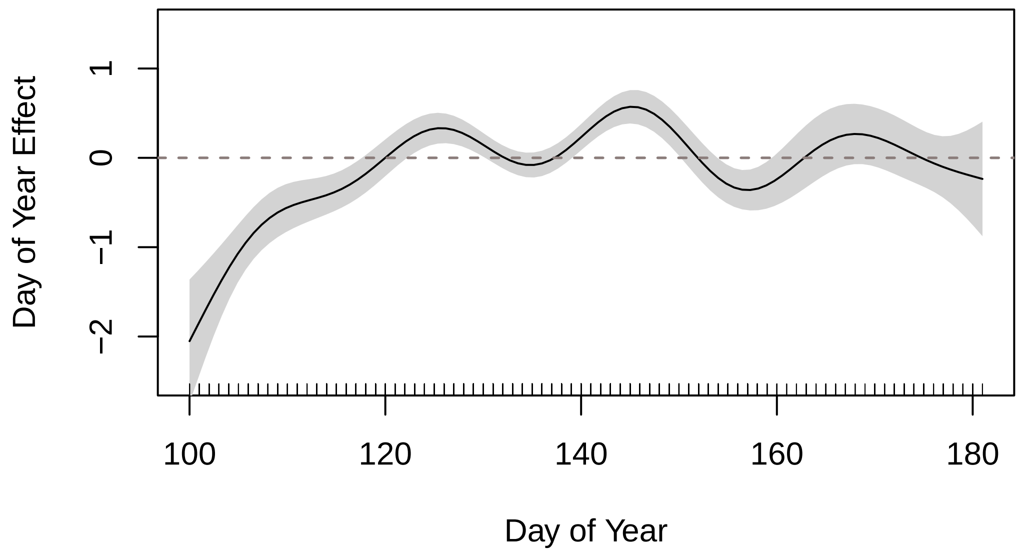
This is the below threshold (2.12 deg C) and above threshold predicted geographical distribution of Alaska plaice eggs (based on the threshold geography model).

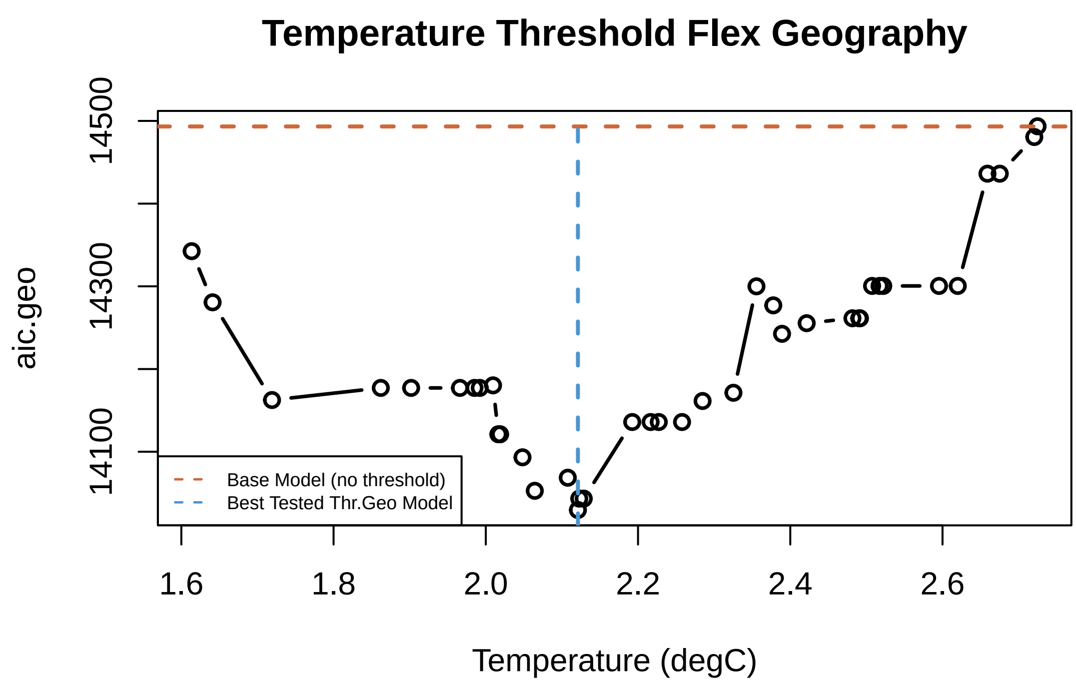


Below
Latitude (⁰N)
Longitude (⁰E)
Anomalies in (log(C/(10m²)+1)
−180
−175
−170
−165
−160
−155
52
54
56
58
60
62
0
2
4
6

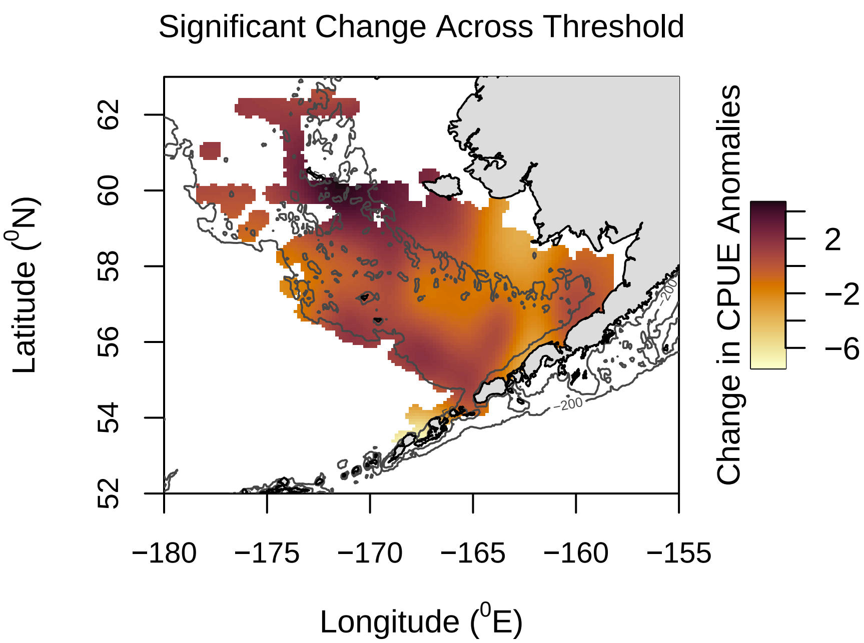


Above
Latitude (⁰N)
Longitude (⁰E)
Anomalies in (log(C/(10m²)+1)
−180
−175
−170
−165
−160
−155
52
54
56
58
60
62
0
2
4
6
8





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.



Reduction in MSE (%):

```
## [1] 14.94038
```

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")
```

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")
```

Then additive temperature:

```r
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")
```

Then additive temperature and salinity, in individual additive terms:

```r
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")
```

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

```r
lv.2d<-gam((Cper10m2+1)~factor(year)+s(lon,lat)+s(doy,k=7)+s(bottom_depth,k=5)+
             te(salinity,temperature),data=aplarv.ctd,family=tw(link='log'),
           method='REML')

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

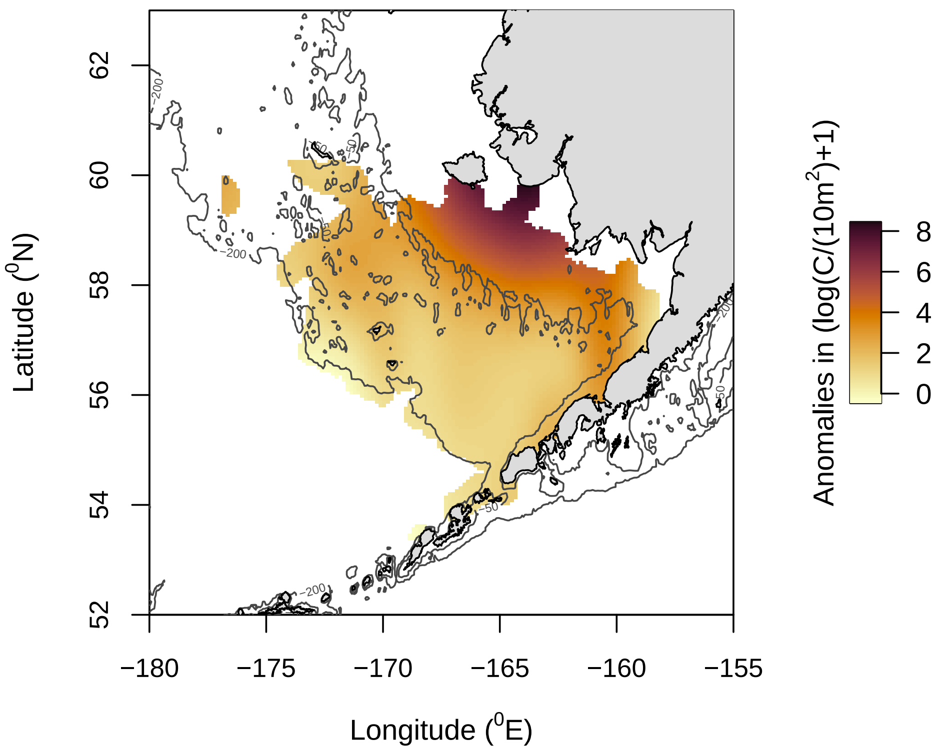
```
##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(lon, lat) + s(doy, k = 7) +
##     s(bottom_depth, k = 5) + te(salinity, temperature)
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.614210   0.234478   2.619 0.008911 **
## factor(year)1998  0.686424   0.363255   1.890 0.059032 .
## factor(year)1999  1.021116   0.294075   3.472 0.000533 ***
## factor(year)2000 -0.099012   0.310653  -0.319 0.749989
## factor(year)2002 -0.056671   0.276383  -0.205 0.837570
## factor(year)2003  0.437833   0.288013   1.520 0.128713
## factor(year)2005  1.264438   0.271872   4.651 3.65e-06 ***
## factor(year)2006  0.506688   0.252880   2.004 0.045317 *
## factor(year)2007  0.493034   0.284493   1.733 0.083333 .
## factor(year)2008  0.697257   0.382321   1.824 0.068425 .
## factor(year)2009 -0.063489   0.260341  -0.244 0.807372
## factor(year)2010  0.236880   0.255236   0.928 0.353540
## factor(year)2011 -0.315271   0.319531  -0.987 0.323992
## factor(year)2012 -0.129643   0.255120  -0.508 0.611425
## factor(year)2013  0.391386   0.314177   1.246 0.213086
## factor(year)2014 -0.008963   0.260672  -0.034 0.972577
## factor(year)2015 -0.291407   0.296308  -0.983 0.325569
## factor(year)2016  0.742205   0.284329   2.610 0.009150 **
```

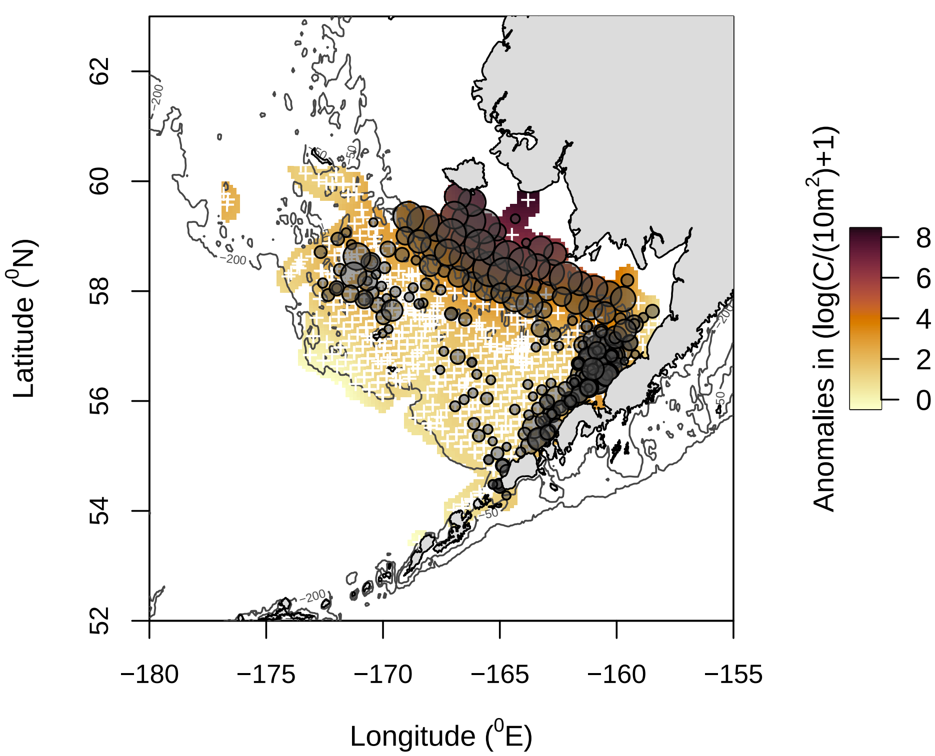
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                              edf Ref.df      F  p-value
## s(lon,lat)                26.996 28.714 21.970  < 2e-16 ***
## s(doy)                     1.000  1.000 28.337 3.61e-07 ***
## s(bottom_depth)            3.808  3.971  7.495 5.13e-05 ***
## te(salinity,temperature) 15.723 18.351 14.055  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0321   Deviance explained = 81.2%
## -REML =   2657  Scale est. = 0.73566   n = 1341
```

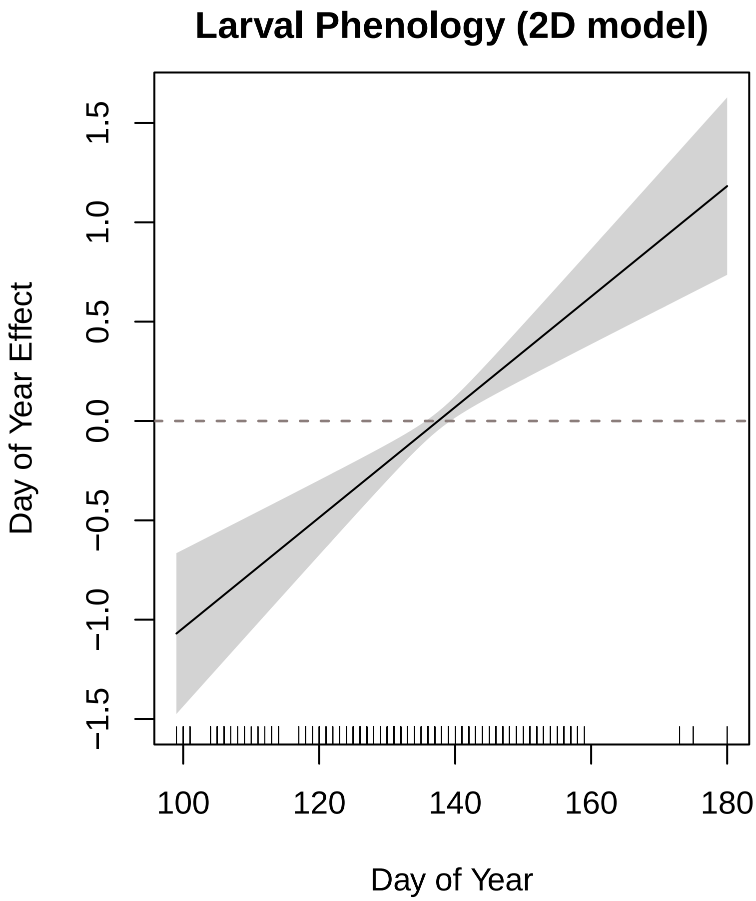
To confirm that this is indeed the best model, we can compare AIC values across all five tested models.

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.

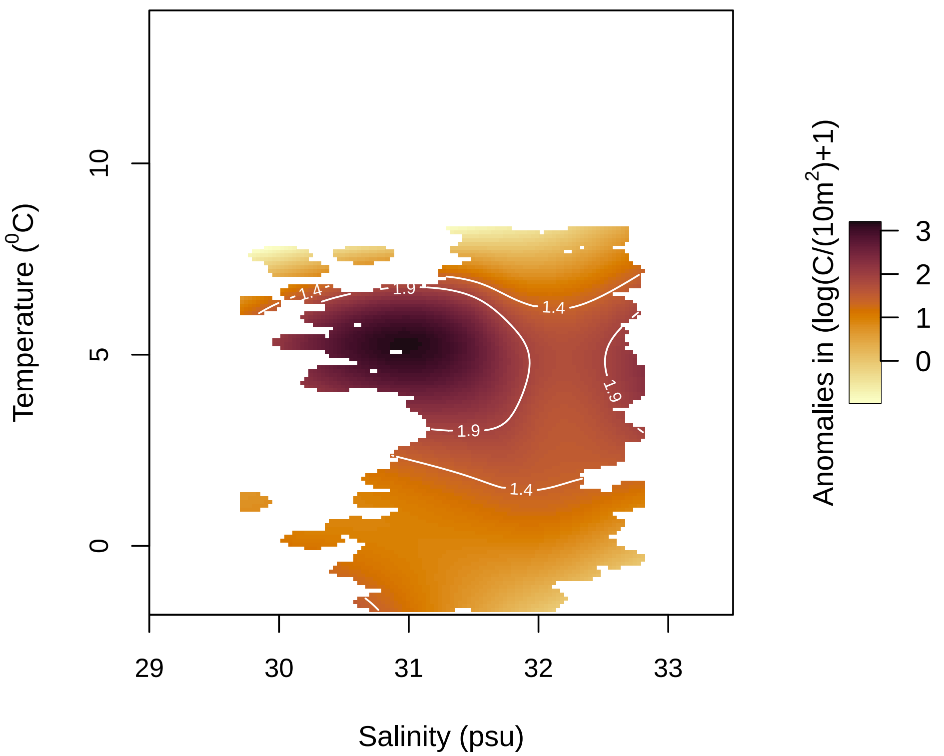
Predicted Larval Biogeography, 2D Model



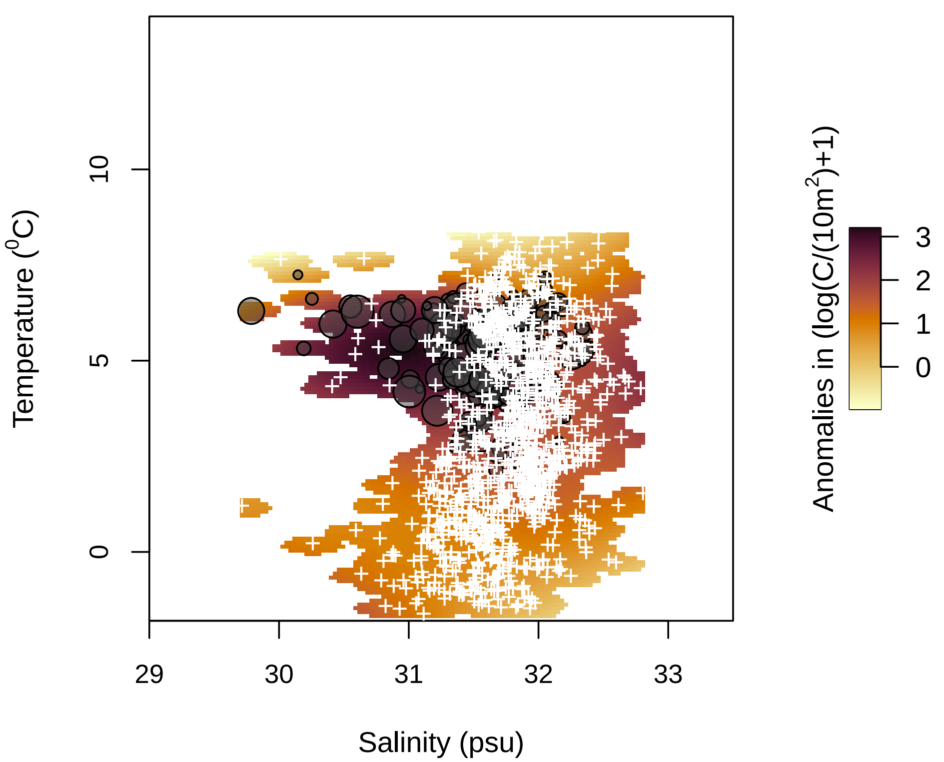




With this bivariate model, we can also calculate the predicted anomalous larval catch (more or less than expected) on a salinity-temperature plot. This figure shows that prediction, with observed larval catch (log(n=1)) overlaid.



Larval Biogeography By Temperature and Salinity



Now I'll calculate a specific range of temperature, salinity, and both temperature and salinity to evaluate breadth of environmental tolerances.

For the univariate predictions, make a grid that holds either temperature or salinity constant depending on the variable of interest. The temperature and salinity values held constant were chosen based on the bivariate temperature,salinity plots; I attempted to capture a representative snapshot of variability in the univariate dimension.

First, I'll determine the sum of larval log(CPUE+1) predictions and 60% of that sum, to know how many rows to cut from the larger extent.

```
## [1] 9440.009
```

```
## [1] 5664.005
```

```
##    salinity temperature       dist year       lon   lat doy bottom_depth
## 1  30.96310    5.277909 0.19659444 2005 -165.4712 56.59 138           78
## 2  30.99432    5.277909 0.21131119 2005 -165.4712 56.59 138           78
## 3  30.93188    5.277909 0.16550113 2005 -165.4712 56.59 138           78
## 4  30.96310    5.177170 0.22825435 2005 -165.4712 56.59 138           78
## 5  31.02554    5.277909 0.18142766 2005 -165.4712 56.59 138           78
## 6  30.99432    5.177170 0.20149253 2005 -165.4712 56.59 138           78
## 7  30.93188    5.177170 0.20209231 2005 -165.4712 56.59 138           78
## 8  30.90065    5.277909 0.13446770 2005 -165.4712 56.59 138           78
## 9  30.96310    5.378647 0.18014016 2005 -165.4712 56.59 138           78
## 10 31.02554    5.177170 0.17172849 2005 -165.4712 56.59 138           78
## 11 30.99432    5.378647 0.15233113 2005 -165.4712 56.59 138           78
## 12 30.93188    5.378647 0.18503058 2005 -165.4712 56.59 138           78
## 13 31.05677    5.277909 0.15208308 2005 -165.4712 56.59 138           78
## 14 30.90065    5.177170 0.17757320 2005 -165.4712 56.59 138           78
## 15 31.02554    5.378647 0.12612992 2005 -165.4712 56.59 138           78
## 16 30.90065    5.378647 0.15788368 2005 -165.4712 56.59 138           78
## 17 31.05677    5.177170 0.14258961 2005 -165.4712 56.59 138           78
## 18 30.86943    5.277909 0.10354801 2005 -165.4712 56.59 138           78
## 19 30.96310    5.076432 0.22709564 2005 -165.4712 56.59 138           78
## 20 30.99432    5.076432 0.19590750 2005 -165.4712 56.59 138           78
## 21 31.05677    5.378647 0.10277369 2005 -165.4712 56.59 138           78
## 22 30.86943    5.177170 0.15547626 2005 -165.4712 56.59 138           78
## 23 31.08799    5.277909 0.12366171 2005 -165.4712 56.59 138           78
## 24 31.02554    5.076432 0.16473237 2005 -165.4712 56.59 138           78
## 25 30.86943    5.378647 0.13254389 2005 -165.4712 56.59 138           78
## 26 31.08799    5.177170 0.11455395 2005 -165.4712 56.59 138           78
## 27 30.83821    5.277909 0.07288696 2005 -165.4712 56.59 138           78
## 28 30.96310    5.479385 0.09148024 2005 -165.4712 56.59 138           78
## 29 30.99432    5.479385 0.09899439 2005 -165.4712 56.59 138           78
## 30 31.05677    5.076432 0.13357938 2005 -165.4712 56.59 138           78
## 31 30.93188    5.479385 0.09427125 2005 -165.4712 56.59 138           78
## 32 31.08799    5.378647 0.08465101 2005 -165.4712 56.59 138           78
## 33 30.83821    5.177170 0.13697878 2005 -165.4712 56.59 138           78
## 34 31.02554    5.479385 0.10438878 2005 -165.4712 56.59 138           78
## 35 30.83821    5.378647 0.11026423 2005 -165.4712 56.59 138           78
## 36 30.90065    5.479385 0.10656074 2005 -165.4712 56.59 138           78
## 37 31.11921    5.277909 0.09697866 2005 -165.4712 56.59 138           78
## 38 31.11921    5.177170 0.08867417 2005 -165.4712 56.59 138           78
```

```
## 39 31.08799    5.076432 0.10246870 2005 -165.4712 56.59 138            78
## 40 30.80699    5.277909 0.04304085 2005 -165.4712 56.59 138            78
## 41 31.05677    5.479385 0.07450296 2005 -165.4712 56.59 138            78
## 42 30.86943    5.479385 0.12559061 2005 -165.4712 56.59 138            78
## 43 31.11921    5.378647 0.07562304 2005 -165.4712 56.59 138            78
## 44 30.80699    5.177170 0.12370607 2005 -165.4712 56.59 138            78
## 45 30.99432    4.975694 0.21571950 2005 -165.4712 56.59 138            78
##        pred    con.sum
## 1  3.178714   3.178714
## 2  3.178001   6.356715
## 3  3.176552   9.533267
## 4  3.174696  12.707963
## 5  3.174339  15.882302
## 6  3.174190  19.056492
## 7  3.172386  22.228878
## 8  3.171586  25.400465
## 9  3.171450  28.571915
## 10 3.170796  31.742711
## 11 3.170601  34.913312
## 12 3.169376  38.082688
## 13 3.167658  41.250346
## 14 3.167331  44.417677
## 15 3.166754  47.584432
## 16 3.164450  50.748881
## 17 3.164446  53.913328
## 18 3.163888  57.077216
## 19 3.160175  60.237390
## 20 3.159941  63.397332
## 21 3.159837  66.557169
## 22 3.159600  69.716769
## 23 3.157886  72.874655
## 24 3.156893  76.031548
## 25 3.156746  79.188294
## 26 3.155069  82.343363
## 27 3.153530  85.496893
## 28 3.152125  88.649018
## 29 3.151217  91.800236
## 30 3.150961  94.951196
## 31 3.150074  98.101270
## 32 3.149777 101.251047
## 33 3.149265 104.400311
## 34 3.147275 107.547586
## 35 3.146338 110.693924
## 36 3.145135 113.839059
## 37 3.144950 116.984009
## 38 3.142594 120.126603
## 39 3.142076 123.268679
## 40 3.140584 126.409263
## 41 3.140224 129.549487
## 42 3.137385 132.686872
## 43 3.136500 135.823372
## 44 3.136394 138.959766
## 45 3.136029 142.095795
```

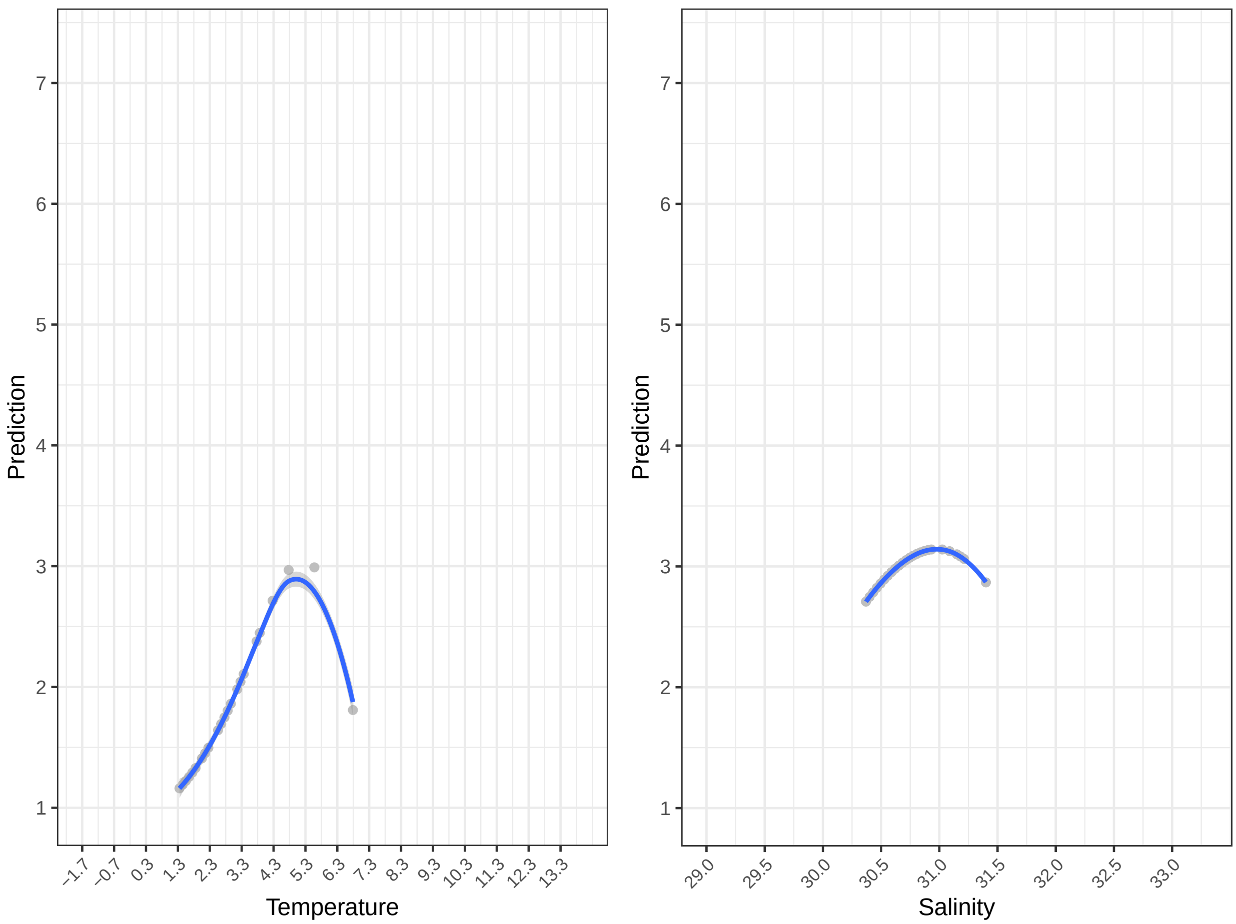
```
## [1] 38.8046
```

So this percentage (percbi, 38.8%) is the percentage of the total grid extent (in terms of nrow) within which 60% of the predicted log(CPUE+1) lies. I.e., 38.8% percent of the whole grid extent encompasses 60% of the predicted observations.

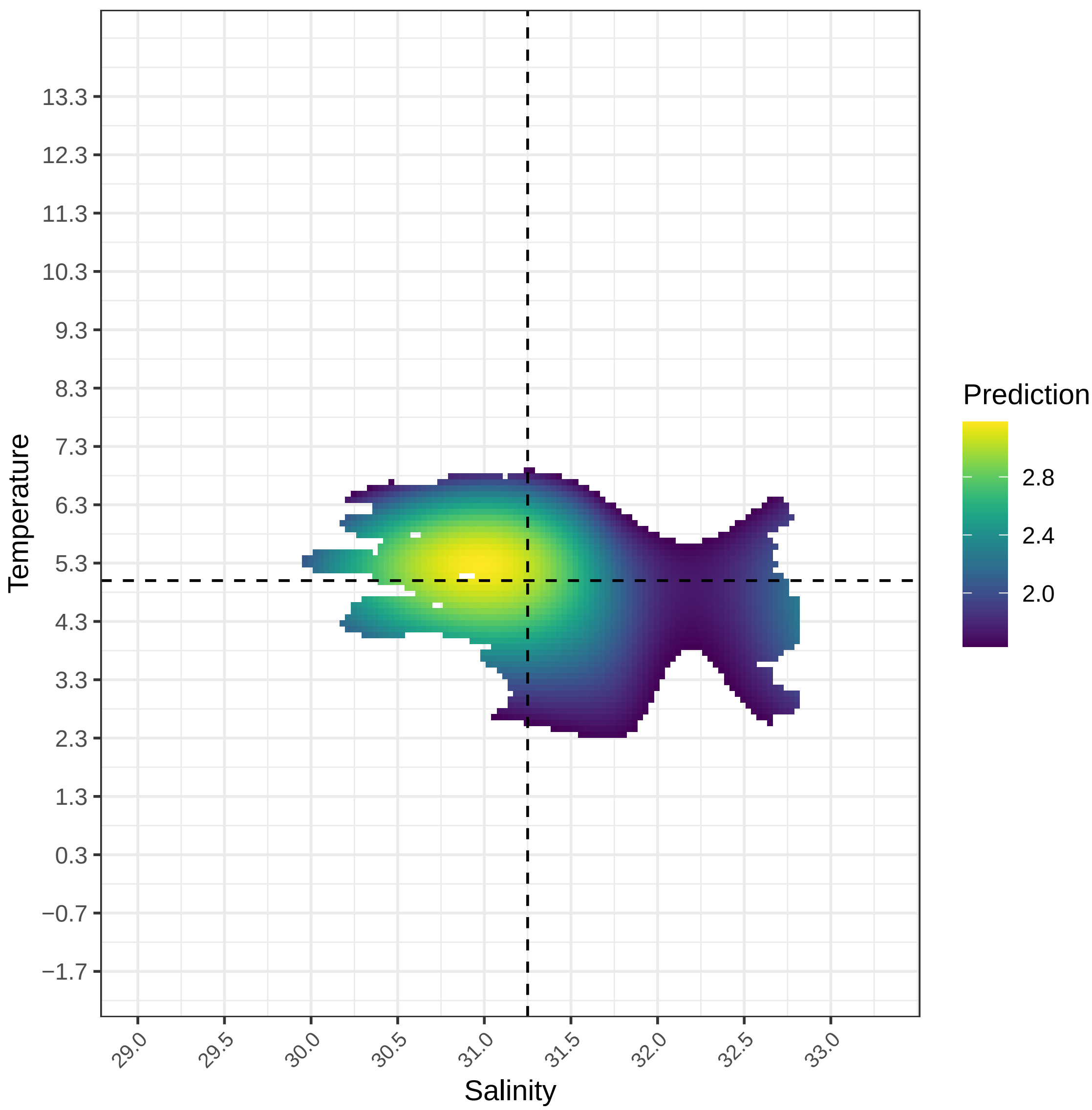
```
## [1] 100
```

```
## [1] 52.08333
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



For the bivariate analysis:



Determine area in T-S units over which 60% of observations are contained.

```
## [1] 41.78145
```

41.78% of the salinity-temperature area contains 60% of the total predicted log(CPUE+1).

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 AIC Comparisons, Alaska plaice

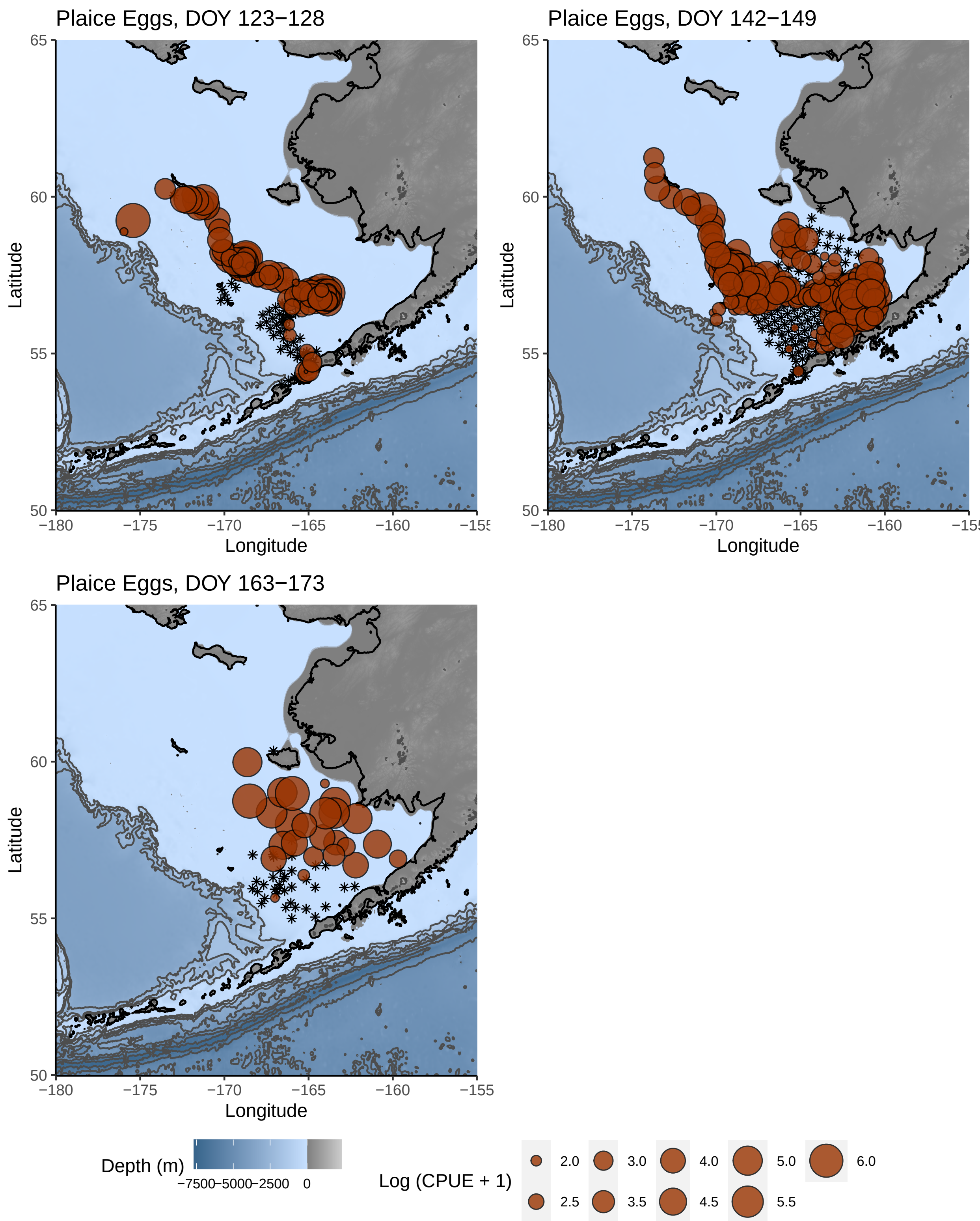
	Best Divided By Base	Best Divided By Second Best
Eggs	0.9680024	0.9964871
Larvae	0.9709863	0.9961160

Reduction in MSE (%):

```
## [1] 9.623696
```

Appendices:

The following plots investigate the phenological peaks in egg distribution to evaluate any spatial differences in egg distribution among these three peaks.


Plaice Eggs, DOY 123−128
Plaice Eggs, DOY 142−149
Plaice Eggs, DOY 163−173
Latitude
Longitude
65
60
55
50
−180
−175
−170
−165
−160
−155
Depth (m)
−7500
−5000
−2500
0
Log (CPUE + 1)
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0