Streamlined Rex Sole Analyses & Figures

Laura Vary

1/24/2022

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:

Rex sole: egg data are included for this species; larval data were too sparse to use effectively. Rex sole spawn from October to May, live roughly 24 years, and transform to juveniles at standard lengths $\geq$ 50 mm, which is huge and reflects the long period of time that rex sole larvae are traveling in the water column before settling.

These data have been trimmed. The data are only constrained to between April and July. The egg data are also joined to regional temperature indices for each year (the reg.sst dataset).

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 rex sole may have experienced, roughly two months before the peak amount of eggs in the water column occurs.

Descriptive Information:

Table 1: Descriptive Metrics for Rex Sole Egg Data

Lat Range	Lon Range	Day of Year Range	Bottom Depth Range
53.4-59.4	-173.5 to -161.6	91-212	0-3506

The following two plots show *the day of year distribution for positive Rex sole egg catch* (left) and *the year distribution for positive Rex sole egg catch* (right).

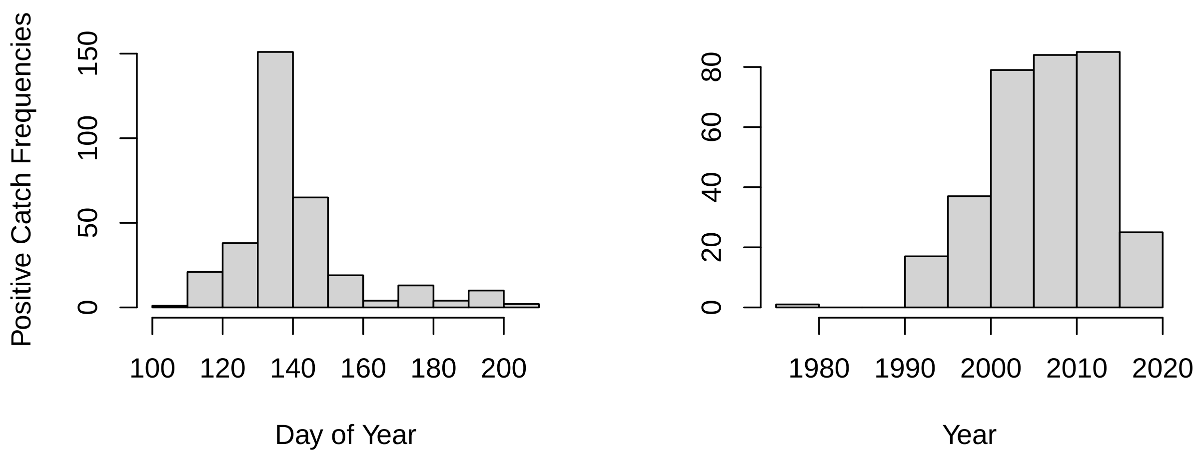
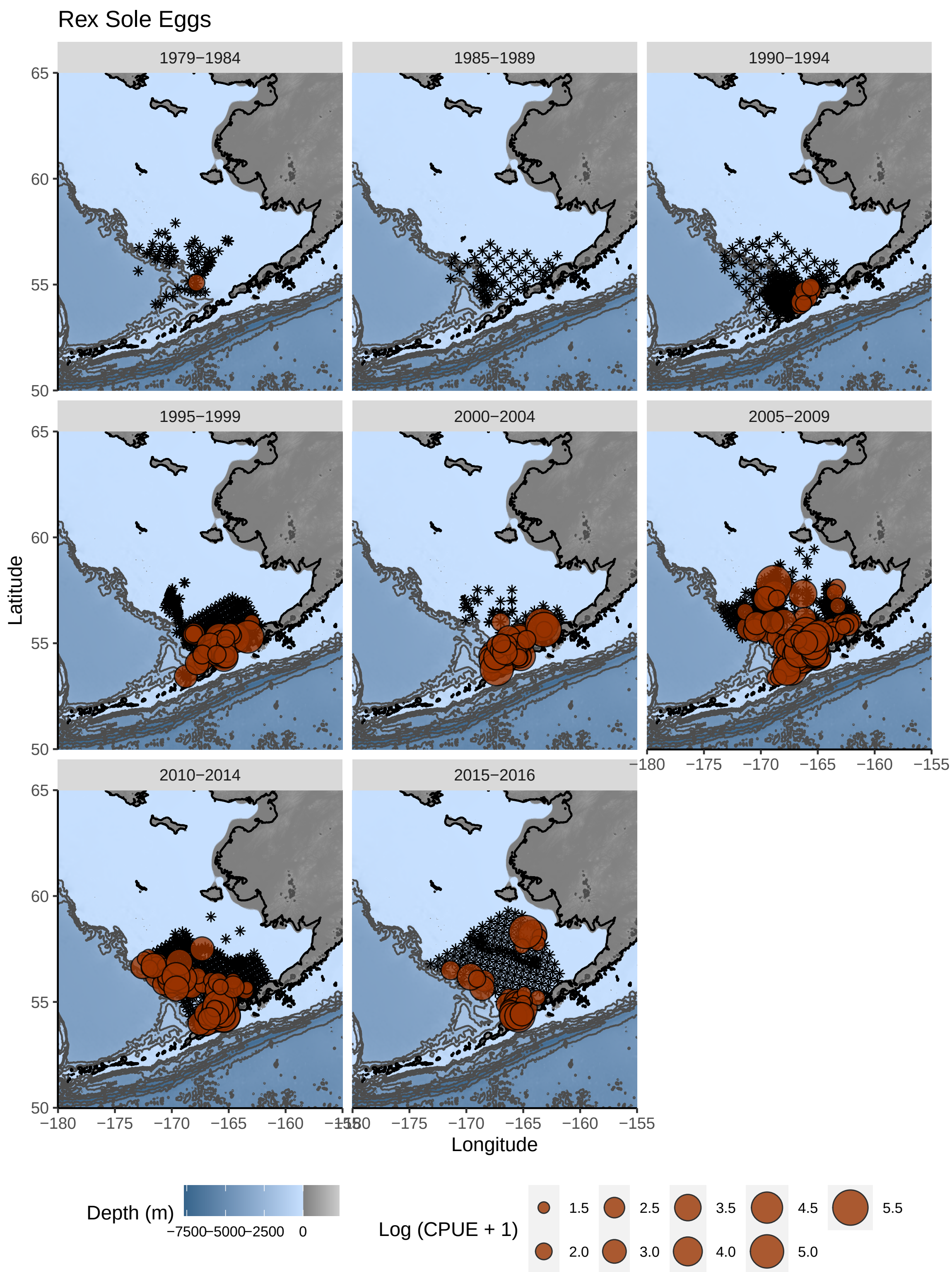


Figure 1: Rex Sole Eggs

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



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.**

Rex sole eggs were best explained by the threshold geography model, in which the spatial distribution of eggs varied differently below and above 2.29 degrees Celsius.

Generalized Additive Models: Rex Eggs

The base model formulation:

```
eg.base<-readRDS("./GAM Models/rx_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)       -0.1253     0.1281  -0.979 0.327918
## factor(year)1988   0.9494     0.2065   4.597 4.49e-06 ***
## factor(year)1991   0.8892     0.1981   4.489 7.46e-06 ***
## factor(year)1992   0.4631     0.2328   1.989 0.046813 *
## factor(year)1993   0.2276     0.1754   1.298 0.194540
## factor(year)1994   0.1328     0.1697   0.782 0.434099
## factor(year)1995   0.7562     0.1481   5.106 3.54e-07 ***
## factor(year)1996   1.4614     0.2990   4.888 1.08e-06 ***
## factor(year)1997   0.5736     0.1604   3.577 0.000354 ***
## factor(year)1998   1.8004     0.2981   6.039 1.78e-09 ***
## factor(year)1999   1.5587     0.1776   8.776  < 2e-16 ***
## factor(year)2000   2.4171     0.2066  11.698  < 2e-16 ***
## factor(year)2001   0.7979     0.3643   2.190 0.028583 *
## factor(year)2002   1.2869     0.1702   7.562 5.50e-14 ***
## factor(year)2003   1.1838     0.1882   6.291 3.70e-10 ***
## factor(year)2004   1.2004     0.4531   2.649 0.008117 **
## factor(year)2005   1.9244     0.1694  11.360  < 2e-16 ***
## factor(year)2006   1.3586     0.1518   8.949  < 2e-16 ***
## factor(year)2007   1.7022     0.1548  10.994  < 2e-16 ***
## factor(year)2008   0.7096     0.1625   4.367 1.31e-05 ***
## factor(year)2009   0.5949     0.1616   3.682 0.000237 ***
## factor(year)2010   0.9516     0.1554   6.124 1.05e-09 ***
## factor(year)2011   1.2195     0.2055   5.936 3.33e-09 ***
## factor(year)2012   0.6406     0.1435   4.463 8.44e-06 ***
## factor(year)2013   0.3951     0.2265   1.744 0.081265 .
## factor(year)2014   1.3389     0.1409   9.501  < 2e-16 ***
## factor(year)2015   0.7064     0.2215   3.190 0.001442 **
## factor(year)2016   1.0735     0.1449   7.409 1.72e-13 ***
## ---
## 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)      27.808 28.899 30.905 < 2e-16 ***
## s(doy)           5.523  6.713 22.579 < 2e-16 ***
## s(bottom_depth)  3.537  3.881  3.338 0.00597 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.151   Deviance explained = 53.3%
## -REML = 4865.9  Scale est. = 0.88518   n = 2593
```

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

```
## [1] 9651.877
```

The variable-coefficient geography formulation (in which geographic egg distributions vary differently in relation to regional SST indices). This was the second best-performing model for rex egg variance.

```
##
## 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)       0.00000    0.00000     NaN      NaN
## factor(year)1988  0.72818    0.18940   3.845 0.000124 ***
## factor(year)1991  0.61086    0.17878   3.417 0.000644 ***
## factor(year)1992  0.30852    0.21497   1.435 0.151349
## factor(year)1993 -0.12973    0.16583  -0.782 0.434133
## factor(year)1994 -0.12560    0.15463  -0.812 0.416714
## factor(year)1995  0.52762    0.12811   4.118 3.94e-05 ***
## factor(year)1996  1.18299    0.30031   3.939 8.40e-05 ***
## factor(year)1997  0.55813    0.15500   3.601 0.000323 ***
## factor(year)1998  1.58433    0.29164   5.432 6.09e-08 ***
## factor(year)1999  1.38565    0.15817   8.760  < 2e-16 ***
## factor(year)2000  2.28699    0.19080  11.986  < 2e-16 ***
## factor(year)2001  0.80267    0.35558   2.257 0.024074 *
## factor(year)2002  1.24401    0.15948   7.800 8.98e-15 ***
## factor(year)2003  1.05371    0.19751   5.335 1.04e-07 ***
## factor(year)2004  1.41247    0.44669   3.162 0.001585 **
## factor(year)2005  1.67082    0.17848   9.361  < 2e-16 ***
## factor(year)2006  1.16664    0.14495   8.049 1.28e-15 ***
## factor(year)2007  1.48129    0.13281  11.154  < 2e-16 ***
## factor(year)2008  0.43025    0.12699   3.388 0.000715 ***
## factor(year)2009  0.42355    0.13753   3.080 0.002095 **
## factor(year)2010  0.68264    0.12091   5.646 1.83e-08 ***
## factor(year)2011  1.27351    0.19606   6.496 9.93e-11 ***
## factor(year)2012  0.49762    0.09685   5.138 2.99e-07 ***
## factor(year)2013  0.12230    0.20269   0.603 0.546311
## factor(year)2014  1.46996    0.13315  11.040  < 2e-16 ***
## factor(year)2015  0.65535    0.23146   2.831 0.004672 **
```

```
## factor(year)2016  1.18501    0.14944   7.930 3.28e-15 ***
## ---
## 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)         20.839 23.345 10.692  <2e-16 ***
## s(doy)              8.025  8.724 19.607  <2e-16 ***
## s(bottom_depth)     3.394  3.809  2.438   0.028 *
## s(lon,lat):reg.SST 21.994 24.454 15.288  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Rank: 99/100
## R-sq.(adj) =  0.193   Deviance explained = 58.8%
## -REML = 4726.5  Scale est. = 0.79547   n = 2593
```

The threshold geography model formulation (in which the geographic distribution of eggs vary differently above and below a threshold temperature of 2.28 degrees Celsius): *This is the best model to explain rex egg distribution as of 1/10/2022*

```
##
## 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)       -0.1552     0.1229  -1.263 0.206777
## factor(year)1988   0.9051     0.1962   4.612 4.18e-06 ***
## factor(year)1991   0.7932     0.1860   4.265 2.08e-05 ***
## factor(year)1992   0.8138     0.2203   3.693 0.000226 ***
## factor(year)1993   0.6738     0.1664   4.049 5.29e-05 ***
## factor(year)1994   0.6179     0.1600   3.863 0.000115 ***
## factor(year)1995   0.7062     0.1408   5.015 5.68e-07 ***
## factor(year)1996   1.3160     0.2772   4.747 2.18e-06 ***
## factor(year)1997   0.8497     0.1575   5.396 7.45e-08 ***
## factor(year)1998   2.0036     0.2847   7.037 2.53e-12 ***
## factor(year)1999   1.7367     0.1689  10.280  < 2e-16 ***
## factor(year)2000   2.0494     0.1958  10.464  < 2e-16 ***
## factor(year)2001   1.0029     0.3355   2.989 0.002823 **
## factor(year)2002   1.0907     0.1646   6.626 4.21e-11 ***
## factor(year)2003   1.1223     0.1821   6.161 8.38e-10 ***
## factor(year)2004   1.5434     0.4279   3.607 0.000316 ***
## factor(year)2005   1.8336     0.1632  11.238  < 2e-16 ***
## factor(year)2006   1.2651     0.1452   8.714  < 2e-16 ***
## factor(year)2007   1.2496     0.1487   8.402  < 2e-16 ***
## factor(year)2008   0.5006     0.1513   3.308 0.000952 ***
## factor(year)2009   0.4828     0.1506   3.206 0.001365 **
## factor(year)2010   0.9401     0.1456   6.456 1.28e-10 ***
## factor(year)2011   1.2177     0.1931   6.306 3.37e-10 ***
```

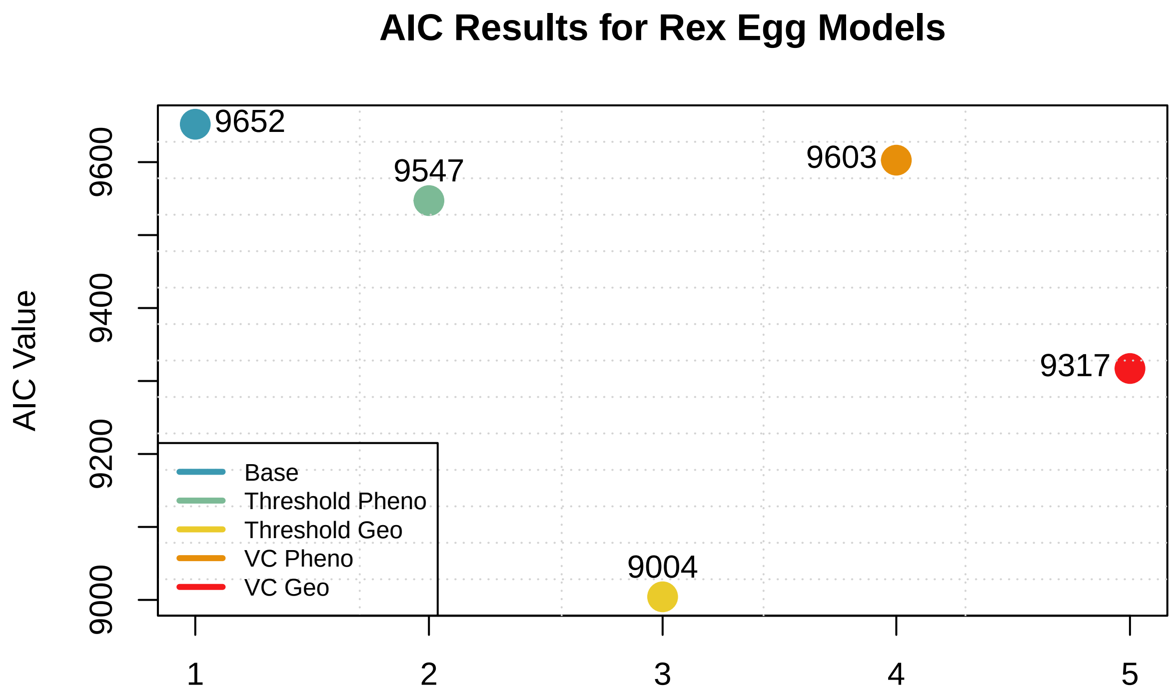
```
## factor(year)2012   0.6137     0.1376   4.459 8.59e-06 ***
## factor(year)2013   0.2311     0.2127   1.086 0.277363
## factor(year)2014   1.3391     0.1377   9.727  < 2e-16 ***
## factor(year)2015   0.8137     0.2081   3.910 9.47e-05 ***
## factor(year)2016   1.0882     0.1397   7.789 9.85e-15 ***
## ---
## 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)               7.887  8.653 21.993  <2e-16 ***
## s(bottom_depth)      3.417  3.815  2.634  0.0212 *
## s(lon,lat):thFALSE 26.940 28.706 41.907  <2e-16 ***
## s(lon,lat):thTRUE  27.109 28.756 13.625  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.241   Deviance explained = 63.2%
## -REML = 4580.7  Scale est. = 0.71907   n = 2593
```

```
aic.pheno<-NA*(temps.in)
thr.pheno<-as.list(1:(length(temps.in)))

for(i in 1:length(temps.in)){
  rxsub$th<-factor(rxsub$reg.SST<=temps.in[i])
  thr.pheno[[i]]<-gam((Cper10m2+1)~factor(year)+
                        s(lon,lat)+
                        s(bottom_depth,k=5)+
                        s(doy,by=th),
                      data=rxsub,family=tw(link='log'),method='REML')
  aic.pheno[i]<-AIC(thr.pheno[[i]])
}

best.index.phe<-order(aic.pheno)[1]
thr.pheno<-thr.pheno[[best.index.phe]]
```

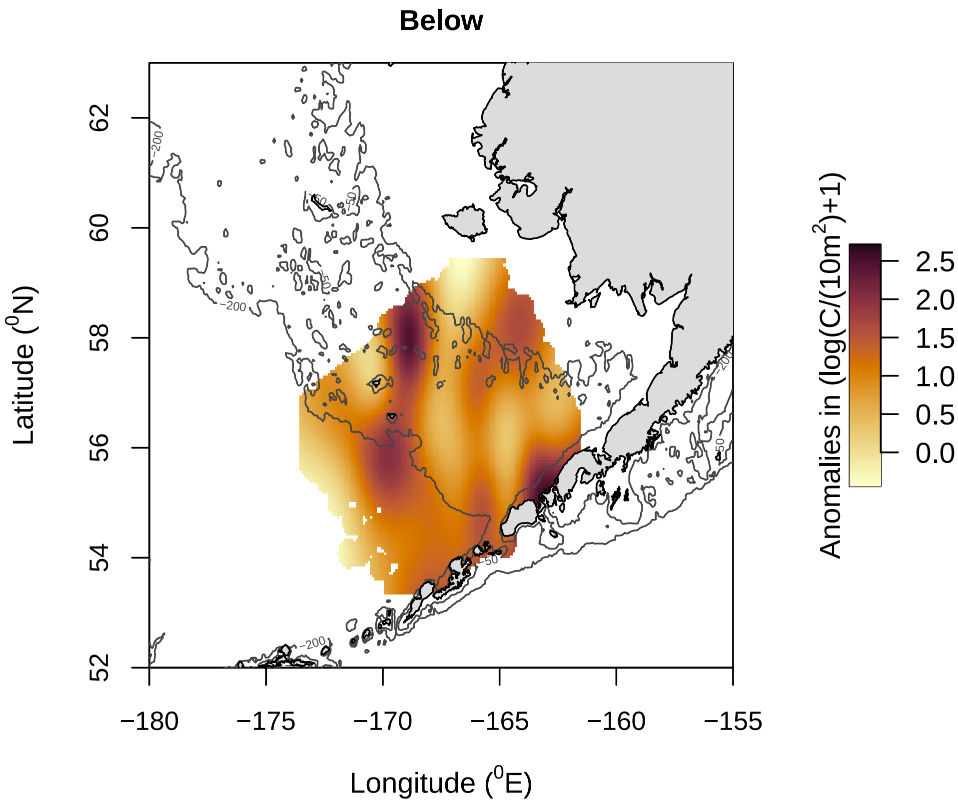
Best threshold temperature:

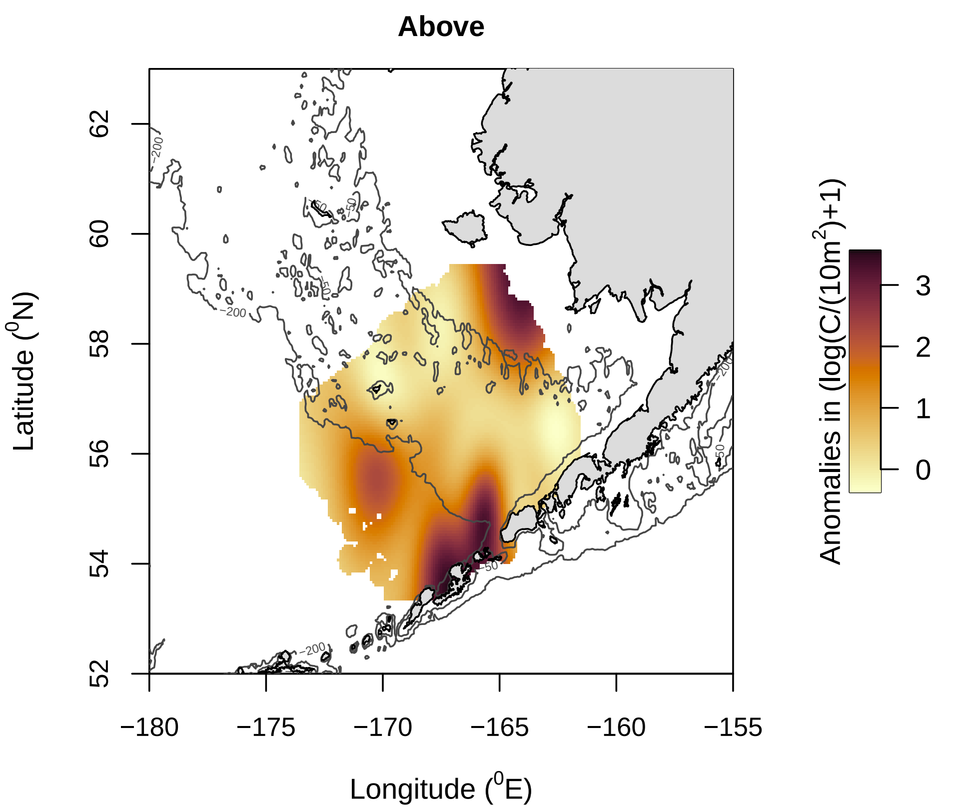
```
## [1] 2.284812
```

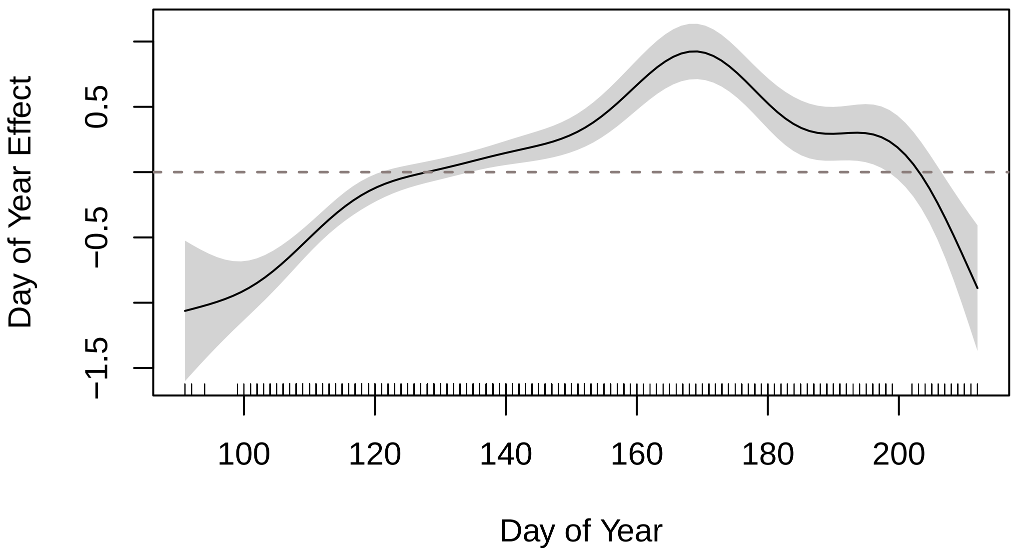
To confirm that the threshold geography model is indeed the best model, we can compare AIC values across all five tested models.

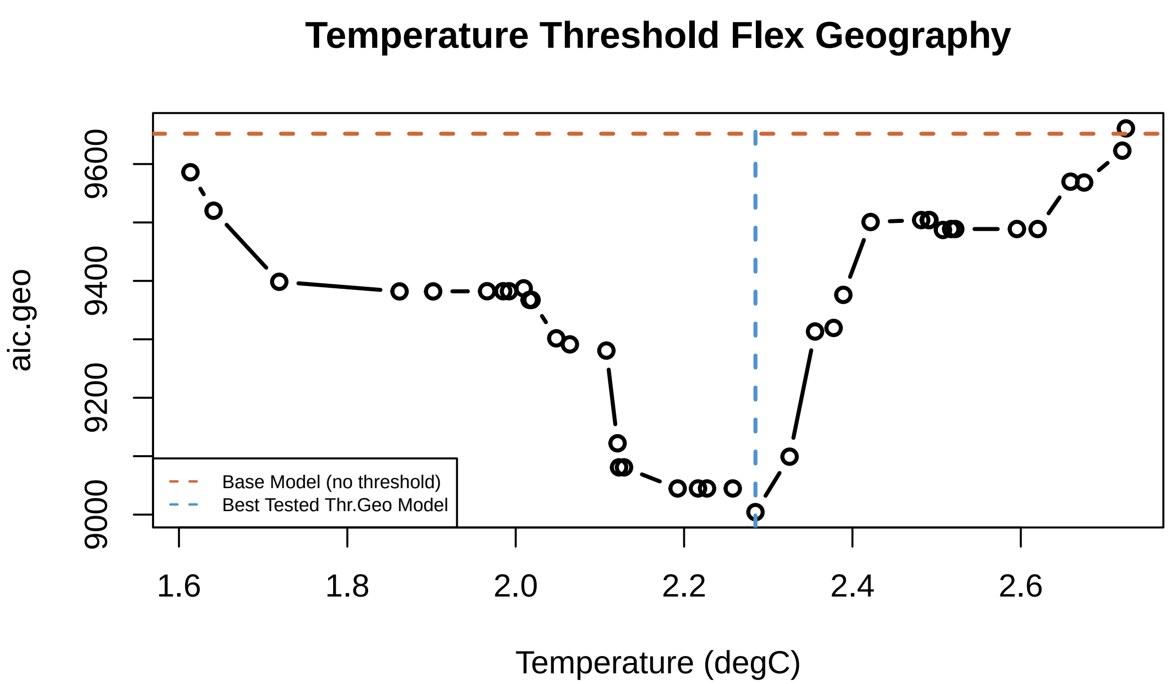


This is the below threshold (2.28 deg C) and above threshold geographic distribution of Rex eggs.

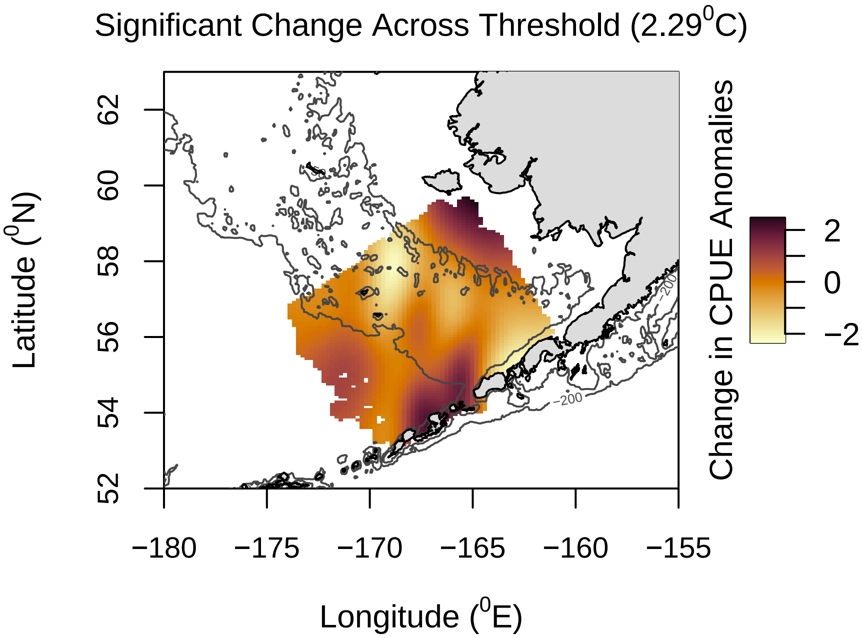

Below
Latitude (⁰N)
62
60
58
56
54
52
−180
−175
−170
−165
−160
−155
Longitude (⁰E)
Anomalies in (log(C/(10m²)+1)
2.5
2.0
1.5
1.0
0.5
0.0


Above
Latitude (⁰N)
62
60
58
56
54
52
−180
−175
−170
−165
−160
−155
Longitude (⁰E)
Anomalies in (log(C/(10m²)+1)
3
2
1
0





With the threshold geography model, we can also see the *significant* differences across the temperature threshold in predicted catch anomalies. This figure shows the prediction of egg catch anomalies below the threshold subtracted from those above the threshold. Thus, more positive values indicate a higher predicted egg catch anomaly above the threshold compared with below the threshold.



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

Table 2: Model Power through AIC Comparisons, Alaska plaice

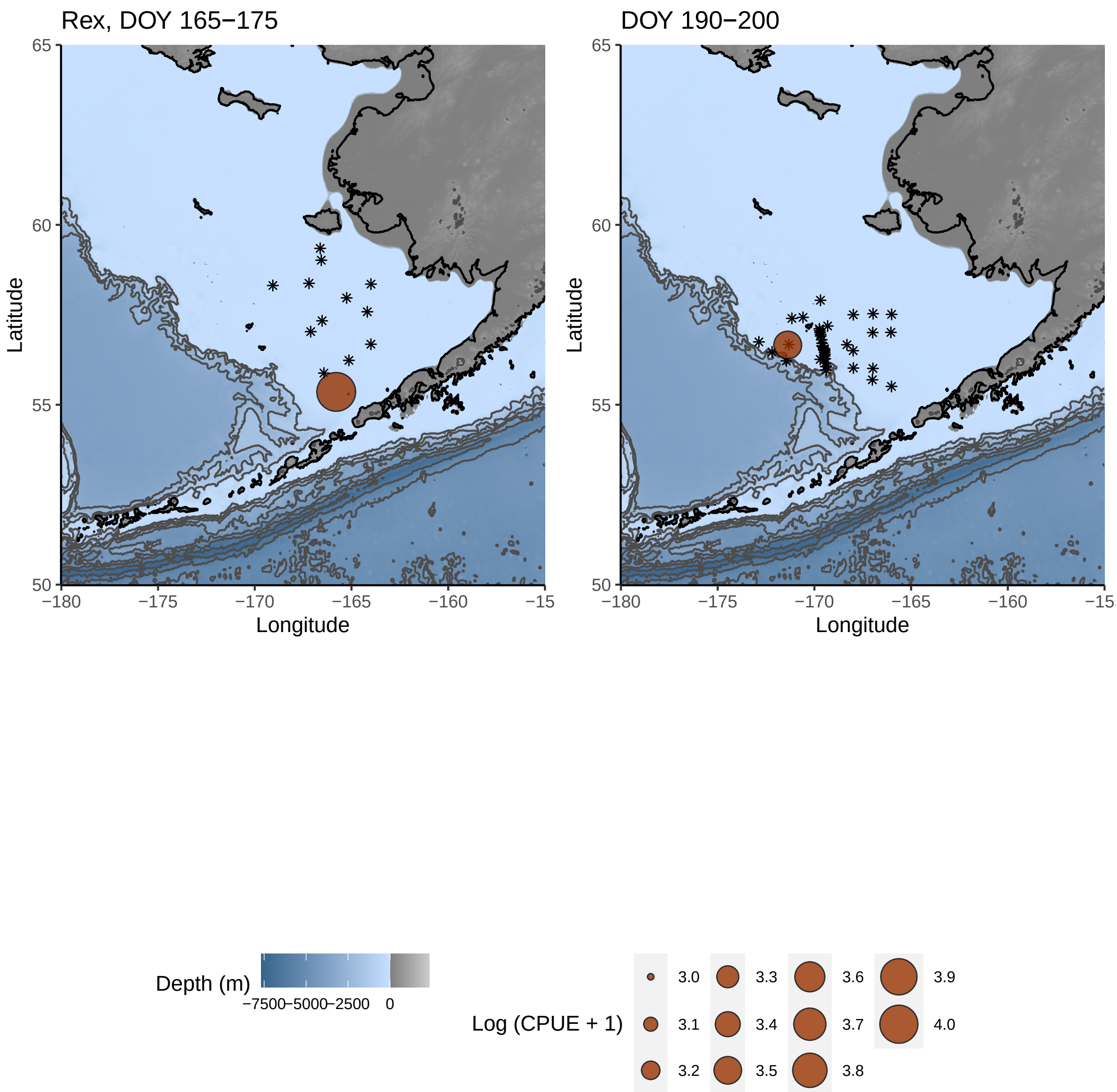
	Best Divided By Base	Best Divided By Second Best
Eggs	0.9329015	0.9664177

Reduction in MSE (%):

```
## [1] 18.76552
```

Appendices:

The following two plots investigate the two distinct phenological peaks of rex sole eggs to evaluate spatial differences in distribution between the two peaks.


Rex, DOY 165−175
DOY 190−200
65
60
55
50
−180
−175
−170
−165
−160
Latitude
Longitude
Depth (m)
−7500
−5000
−2500
0
Log (CPUE + 1)
3.0
3.1
3.2
3.3
3.4
3.5
3.6
3.7
3.8
3.9
4.0