

# Streamlined Flathead Sole Analyses & Figures

Laura Vary

1/23/2022

## Contents

<b>Flathead Sole:</b>	<b>1</b>
Loading Data: . . . . .	1
Descriptive Information: . . . . .	1
Generalized Additive Models: Flathead Sole Eggs . . . . .	5
Larval Generalized Additive Models: . . . . .	13
Appendices: . . . . .	20

## Flathead Sole:

### Loading Data:

Flathead sole: both egg and larval data are included for this species. Flathead sole spawn from February to July, live roughly 21 years, and transform to juveniles at standard lengths between 18 to 21 mm.

These data have been trimmed. The egg data are constrained to depths between 54 and 221 meters; temporally, the egg data are constrained to above the 99th day of year and below the 283rd day of the year (temporally centered on the spawning period of flathead sole). The egg data are also joined to regional temperature indices for each year (the reg.sst dataset). The larval data are also constrained to depths between 54 and 221 m, above the 100th day of the year. Larvae are linked to CTD-derived, *in situ* temperature and salinity measurements. Both eggs and larvae are restricted to latitudes below 61 degrees north.

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 flathead 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 Flathead Sole Egg Data

Lat Range	Lon Range	Day of Year Range	Bottom Depth Range
53-61	-178.4 to -159.6	100-282	55-220

Table 2: Descriptive Metrics for Flathead Sole Larval Data

Lat Range	Lon Range	Day of Year Range	Bottom Depth Range
53.4-61	-178.2 to -159.6	100-283	55-220

The following two plots show *the day of year distribution for positive Flathead sole egg catch* (left) and *the year distribution for positive Flathead sole egg catch* (right). Analogous plots for larval data are following.

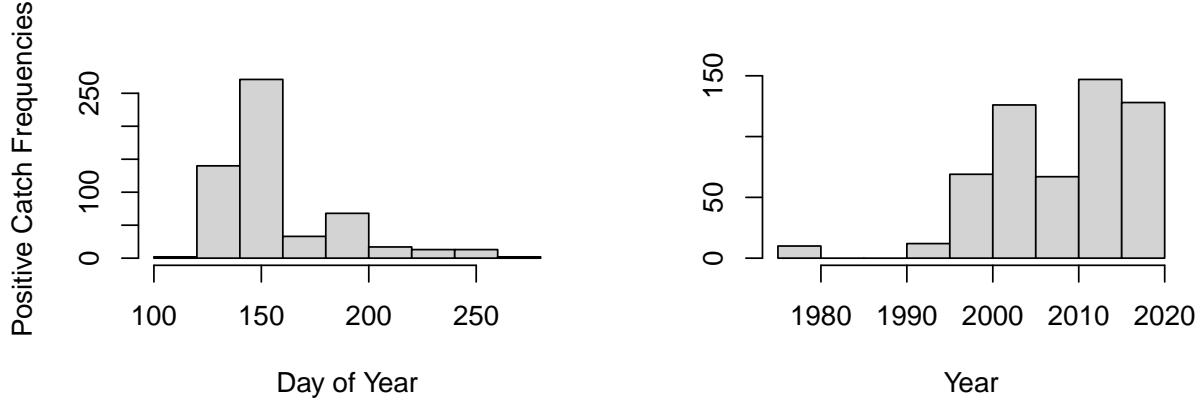
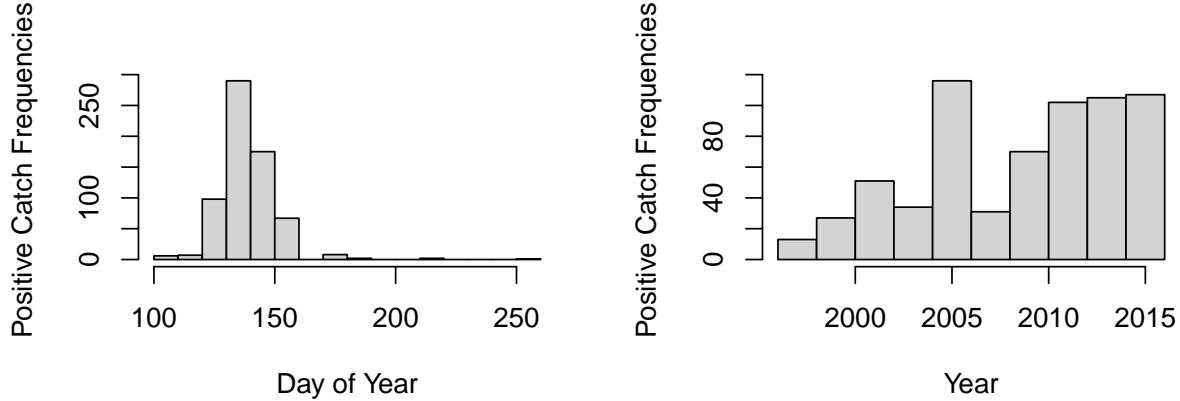
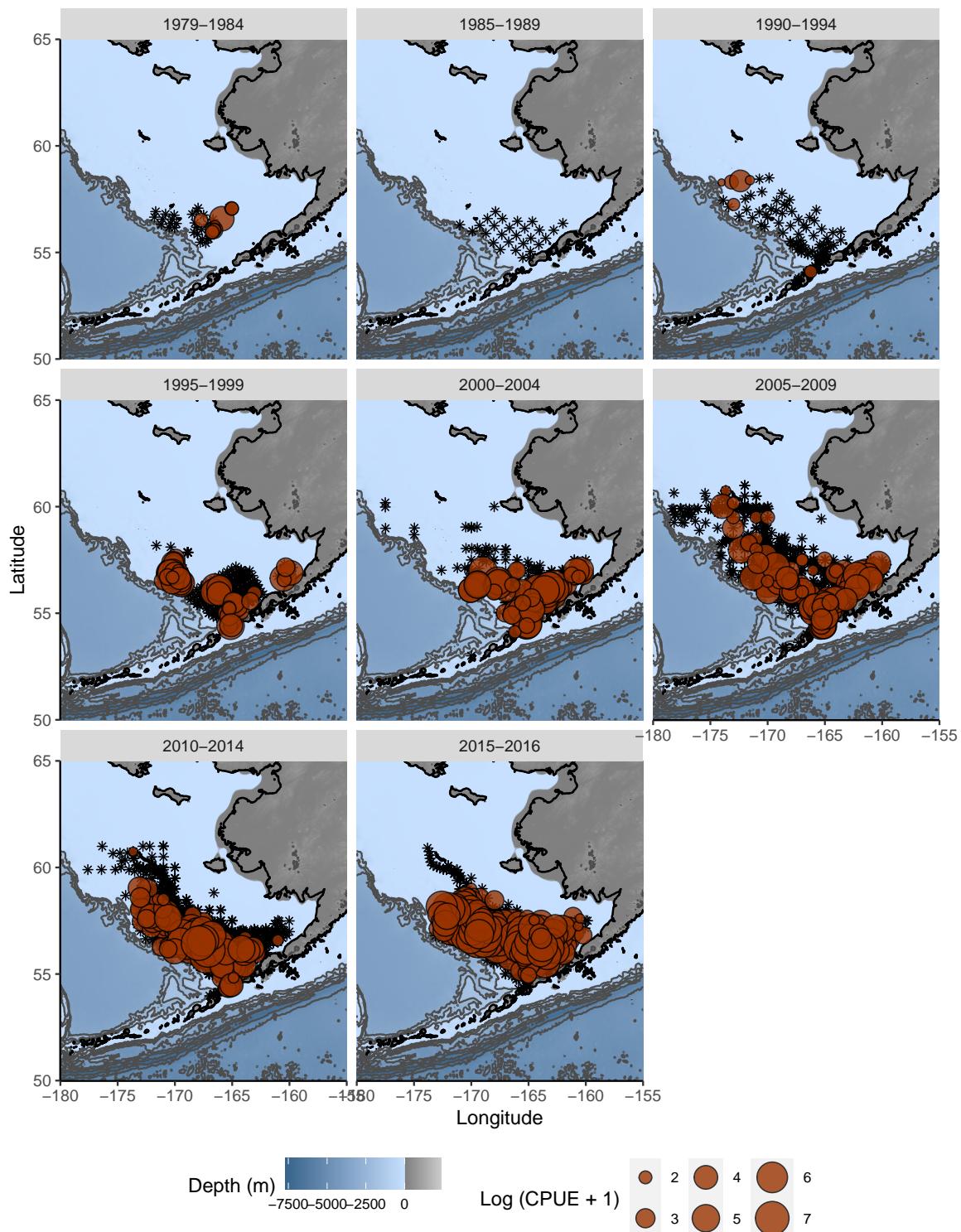


Figure 1: Flathead Sole Eggs

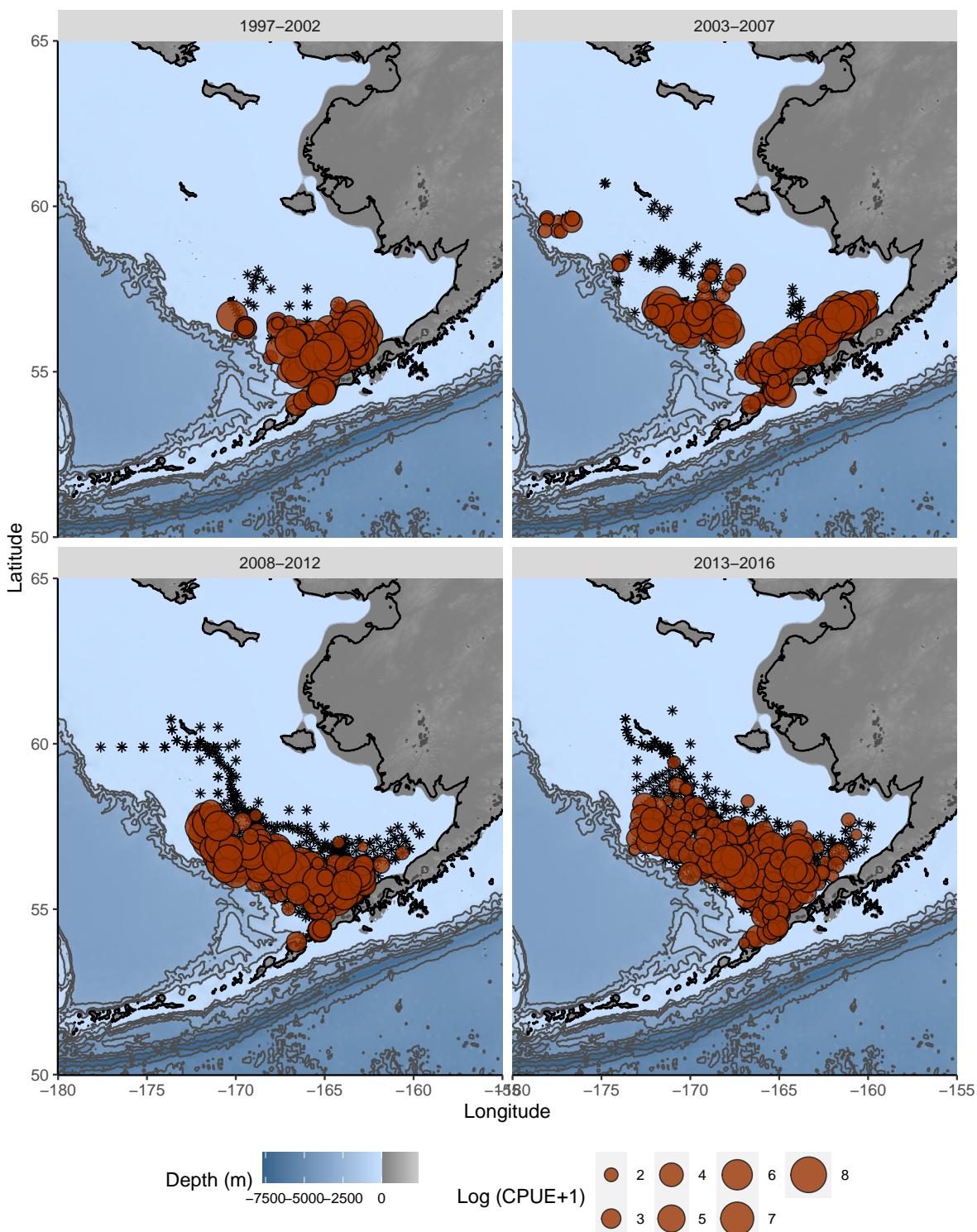


The following plots show flathead sole egg and larval catch distributions (Catch per unit effort, or per  $10m^2$ ) across five year increments from 1979 to 2016.

### Flathead Sole Eggs



### Flathead Sole Larvae



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.**

Flathead sole eggs were best explained by the threshold phenology model, in which the temporal distribution of eggs varied differently below and above 2.29 degrees Celsius.

## Generalized Additive Models: Flathead Sole Eggs

The base model formulation:

```
eg.base<-readRDS("./GAM Models/fh_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) -1.0027    0.1606 -6.244 4.79e-10 ***
## factor(year)1988  1.7991    0.2801  6.424 1.51e-10 ***
## factor(year)1991  1.9807    0.2750  7.203 7.21e-13 ***
## factor(year)1993  2.2811    0.3156  7.229 5.99e-13 ***
## factor(year)1994  2.4807    0.2157 11.498 < 2e-16 ***
## factor(year)1995  1.7017    0.1832  9.291 < 2e-16 ***
## factor(year)1996  1.9960    0.3170  6.295 3.46e-10 ***
## factor(year)1997  2.3863    0.2037 11.715 < 2e-16 ***
## factor(year)1998  1.4476    0.3982  3.635 0.000282 ***
## factor(year)1999  1.8234    0.2279  8.001 1.68e-15 ***
## factor(year)2000  2.4810    0.2167 11.447 < 2e-16 ***
## factor(year)2001  2.9597    0.4111  7.200 7.38e-13 ***
## factor(year)2002  3.3842    0.1963 17.243 < 2e-16 ***
## factor(year)2003  3.2020    0.2128 15.049 < 2e-16 ***
## factor(year)2004  1.8269    0.2357  7.750 1.21e-14 ***
## factor(year)2005  3.5258    0.1848 19.079 < 2e-16 ***
## factor(year)2006  2.3856    0.1763 13.532 < 2e-16 ***
## factor(year)2007  1.8878    0.1787 10.566 < 2e-16 ***
## factor(year)2008  1.6110    0.1841  8.752 < 2e-16 ***
## factor(year)2009  1.6182    0.1805  8.964 < 2e-16 ***
## factor(year)2010  1.5910    0.1763  9.025 < 2e-16 ***
## factor(year)2011  1.9551    0.1858 10.522 < 2e-16 ***
## factor(year)2012  1.2361    0.1737  7.116 1.35e-12 ***
## factor(year)2013  1.8796    0.2803  6.706 2.34e-11 ***
## factor(year)2014  2.9302    0.1682 17.419 < 2e-16 ***
## factor(year)2015  2.0992    0.1976 10.625 < 2e-16 ***
## factor(year)2016  3.7805    0.1709 22.117 < 2e-16 ***
##
## 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) 25.800 28.359  7.326 <2e-16 ***
```

```

## s(doy)          8.793  8.987 240.588 <2e-16 ***
## s(bottom_depth) 3.554  3.870 10.638 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.179   Deviance explained = 70.1%
## -REML = 7876.6   Scale est. = 1.1824    n = 3447

```

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

```
## [1] 15707.05
```

The variable-coefficient phenology formulation (in which temporal (phenological) distribution of eggs vary in relation to regional SST indices). This was the second-best performing model for flathead sole egg variation.

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

```

##
## Family: Tweedie(p=1.99)
## Link function: log
##
## Formula:
## (Cper10m2 + 1) ~ factor(year) + s(lon, lat) + s(doy) + s(bottom_depth,
##     k = 5) + s(doy, by = reg.SST)
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)              0.0000    0.0000    NaN      NaN
## factor(year)1988         1.7248    0.2456   7.024 2.60e-12 ***
## factor(year)1991         1.9875    0.2414   8.233 2.57e-16 ***
## factor(year)1993         2.7719    0.2896   9.572 < 2e-16 ***
## factor(year)1994         2.7808    0.1888  14.732 < 2e-16 ***
## factor(year)1995         1.8034    0.1533  11.761 < 2e-16 ***
## factor(year)1996         2.3806    0.3310   7.193 7.78e-13 ***
## factor(year)1997         2.6140    0.1855  14.093 < 2e-16 ***
## factor(year)1998         2.6578    0.4482   5.930 3.33e-09 ***
## factor(year)1999         1.7866    0.1948   9.173 < 2e-16 ***
## factor(year)2000         2.5138    0.1887  13.320 < 2e-16 ***
## factor(year)2001         3.3977    0.4053   8.383 < 2e-16 ***
## factor(year)2002         3.2767    0.1781  18.397 < 2e-16 ***
## factor(year)2003         3.0070    0.2164  13.896 < 2e-16 ***
## factor(year)2004         2.5170    0.2262  11.128 < 2e-16 ***
## factor(year)2005         3.3568    0.1933  17.367 < 2e-16 ***
## factor(year)2006         2.8134    0.1644  17.114 < 2e-16 ***
## factor(year)2007         1.8101    0.1455  12.438 < 2e-16 ***
## factor(year)2008         1.1785    0.1336   8.823 < 2e-16 ***
## factor(year)2009         1.5273    0.1462  10.444 < 2e-16 ***
## factor(year)2010         1.2986    0.1277  10.172 < 2e-16 ***
## factor(year)2011         2.4310    0.1711  14.212 < 2e-16 ***
## factor(year)2012         0.9773    0.1113   8.785 < 2e-16 ***
## factor(year)2013         1.8018    0.2397   7.516 7.20e-14 ***

```

```

## factor(year)2014  2.9196      0.1532   19.055 < 2e-16 ***
## factor(year)2015  3.4918      0.2134   16.365 < 2e-16 ***
## factor(year)2016  3.4871      0.1732   20.128 < 2e-16 ***
## ---
## 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)    26.068 28.457  7.632 <2e-16 ***
## s(doy)        7.668  8.336 69.094 <2e-16 ***
## s(bottom_depth) 3.602  3.895 12.200 <2e-16 ***
## s(doy):reg.SST  9.122  9.567 123.714 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Rank: 78/79
## R-sq.(adj) =  0.203  Deviance explained = 75.9%
## -REML = 7477.2  Scale est. = 0.98083 n = 3447

```

```
AIC(vc.pheno)
```

```
## [1] 14852.47
```

The threshold phenology model formulation (in which the temporal (phenological) distribution of eggs vary differently above and below a threshold temperature. *This is the best model to explain Flathead sole egg variation across years, as of 1/10/2021.*

```

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

for(i in 1:length(temps.in)){
  fhsub$th<-factor(fhsub$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=fhsub,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]]

```

```
summary(thr.pheno)
```

```

##
## 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)           -0.9655    0.1569  -6.152 8.52e-10 ***
## factor(year)1988      1.5372    0.2614   5.880 4.50e-09 ***
## factor(year)1991      1.5886    0.2579   6.159 8.18e-10 ***
## factor(year)1993      1.9031    0.2918   6.523 7.93e-11 ***
## factor(year)1994      1.9054    0.2048   9.303 < 2e-16 ***
## factor(year)1995      1.7115    0.1758   9.735 < 2e-16 ***
## factor(year)1996      1.6320    0.3090   5.281 1.37e-07 ***
## factor(year)1997      2.5309    0.1953  12.960 < 2e-16 ***
## factor(year)1998      1.3295    0.4858   2.737  0.00624 **  
## factor(year)1999      1.8165    0.2155   8.429 < 2e-16 ***
## factor(year)2000      1.7940    0.2117   8.473 < 2e-16 ***
## factor(year)2001      2.9755    0.3983   7.470 1.02e-13 ***
## factor(year)2002      2.7133    0.1905  14.241 < 2e-16 ***
## factor(year)2003      2.4997    0.2050  12.193 < 2e-16 ***
## factor(year)2004      2.1995    0.2270   9.689 < 2e-16 ***
## factor(year)2005      2.9392    0.1808  16.252 < 2e-16 ***
## factor(year)2006      2.5640    0.1699  15.089 < 2e-16 ***
## factor(year)2007      1.4266    0.1724   8.274 < 2e-16 ***
## factor(year)2008      1.1653    0.1765   6.601 4.71e-11 ***
## factor(year)2009      1.1551    0.1744   6.624 4.05e-11 *** 
## factor(year)2010      1.1925    0.1708   6.982 3.50e-12 *** 
## factor(year)2011      2.2509    0.1793  12.555 < 2e-16 ***
## factor(year)2012      1.0735    0.1691   6.349 2.46e-10 *** 
## factor(year)2013      1.5798    0.2625   6.018 1.95e-09 *** 
## factor(year)2014      2.4708    0.1633  15.126 < 2e-16 ***
## factor(year)2015      2.5614    0.1898  13.497 < 2e-16 *** 
## factor(year)2016      3.2820    0.1656  19.818 < 2e-16 *** 
## --- 
## 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)          26.279 28.528  7.816 <2e-16 ***
## s(bottom_depth)     3.572  3.881 13.288 <2e-16 *** 
## s(doy):thFALSE     8.818  8.990 409.768 <2e-16 *** 
## s(doy):thTRUE       8.300  8.854  54.858 <2e-16 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## R-sq.(adj) =  0.2 Deviance explained = 76.2%
## -REML = 7453.3 Scale est. = 0.97022 n = 3447

```

```
AIC(thr.pheno)
```

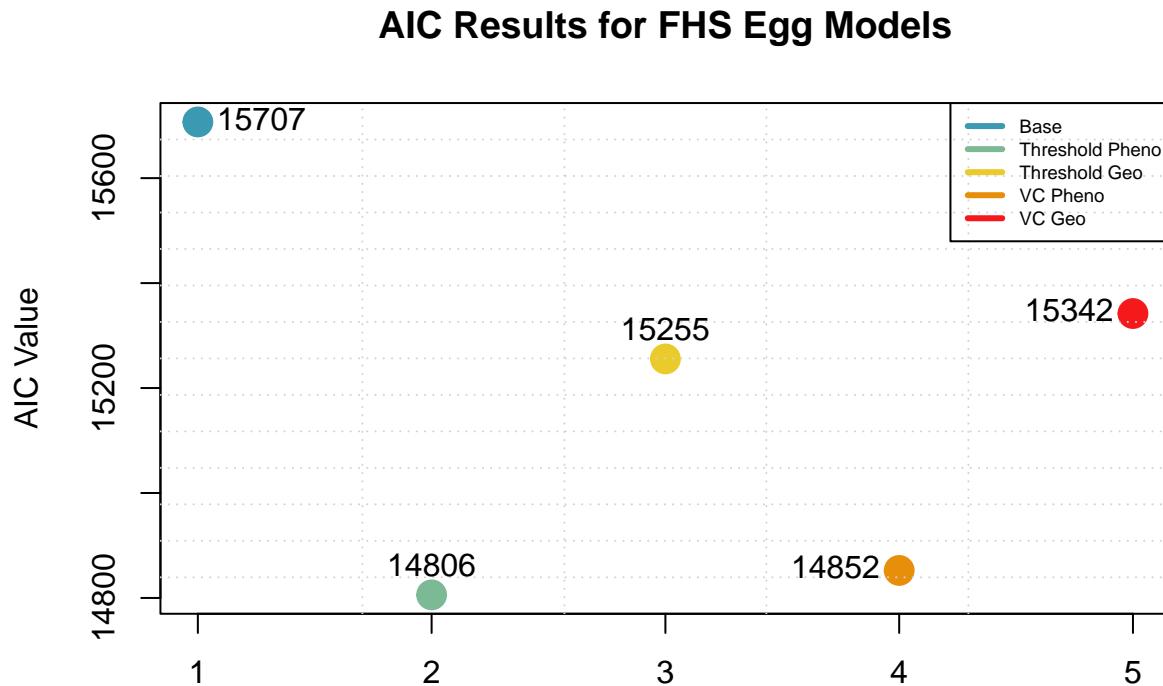
```
## [1] 14805.92
```

```
temp<-temps.in[[best.index.phe]]
```

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

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

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



This is the below threshold and above threshold temporal distribution of flathead eggs (based on the threshold phenology model).

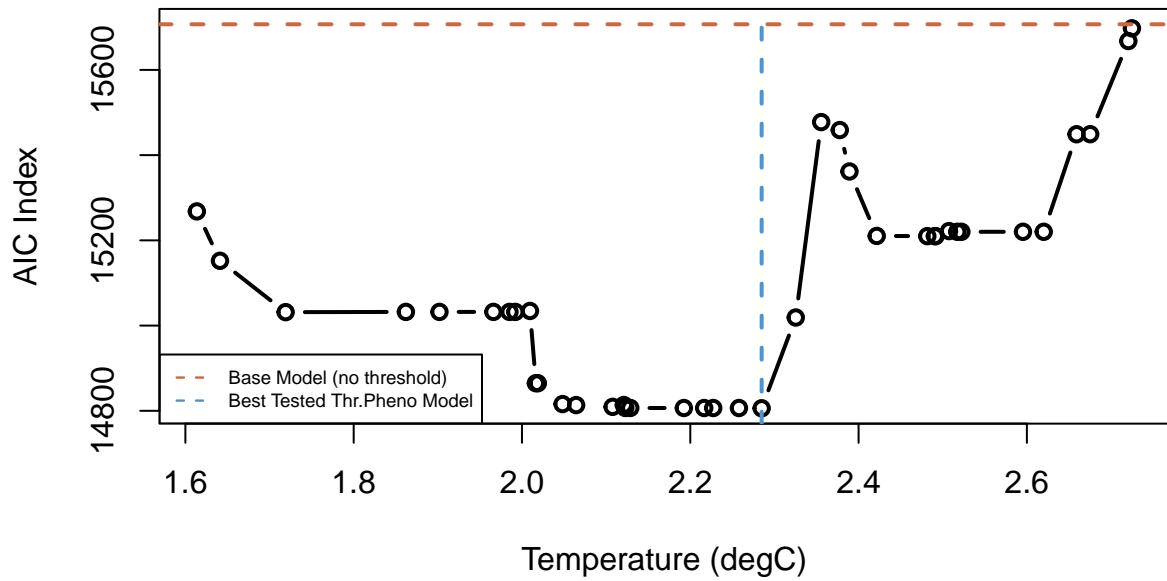
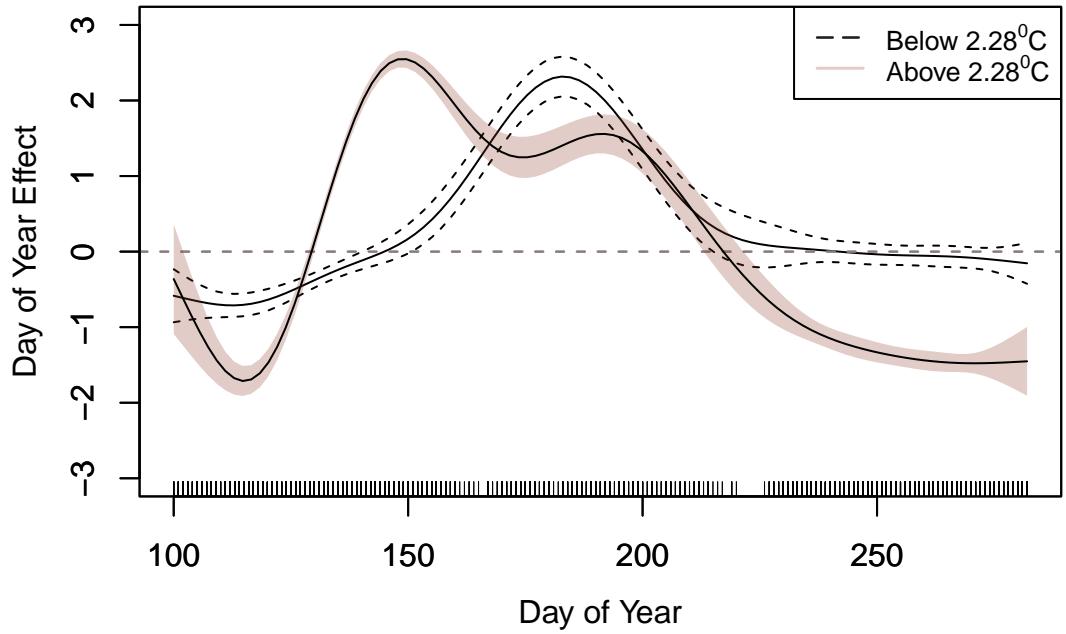
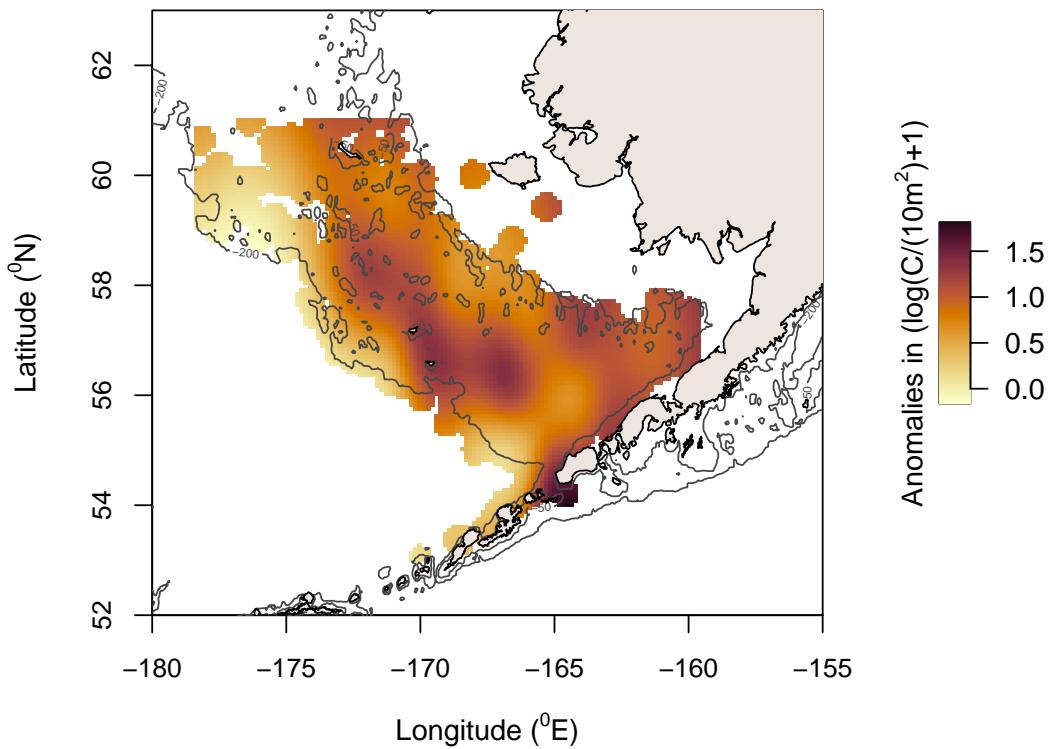


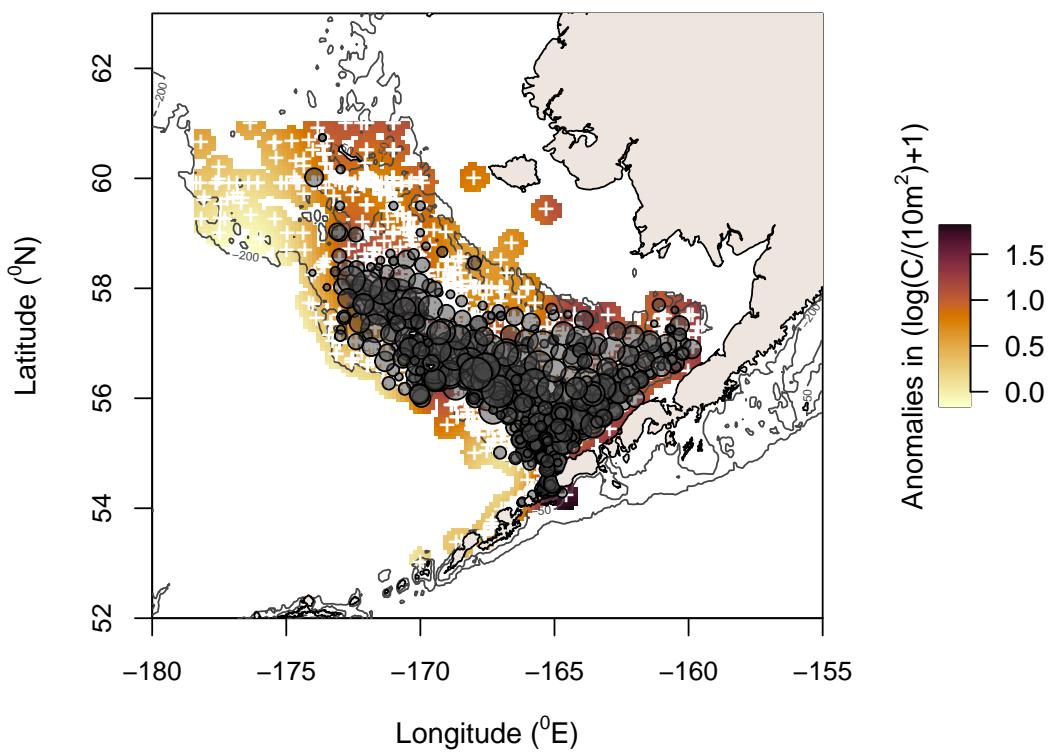
Figure 2: Thr. Pheno AIC scores across varying threshold temperatures, Flathead Sole.

With the threshold phenology model, we can model the spatial distribution of flathead sole eggs through

predictions based on this model. Observed, log-transformed  $(n+1)$  egg catches are displayed as well.



**Flathead Sole Distribution, Th.Ph, Eggs**



Reduction in MSE (%):

```
## [1] 9.685161
```

## 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:

Then additive temperature and salinity, in individual additive terms. This is the second-best performing model.

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

```
lv.2d<-readRDS("./GAM Models/fh_larv_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) + te(salinity, temperature)
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)            2.17016   0.34547   6.282 4.71e-10 ***
## factor(year)1998      0.02491   0.91989   0.027 0.978399
## factor(year)1999      0.43739   0.60634   0.721 0.470829
## factor(year)2000      1.47667   0.49086   3.008 0.002683 **
## factor(year)2002      2.71844   0.40805   6.662 4.14e-11 ***
## factor(year)2003      1.31768   0.44732   2.946 0.003286 **
## factor(year)2005      1.37570   0.41822   3.289 0.001034 **
## factor(year)2006      2.21639   0.38403   5.771 1.01e-08 ***
## factor(year)2007      0.73732   0.42322   1.742 0.081744 .
## factor(year)2008      0.69810   0.55057   1.268 0.205062
## factor(year)2009     -1.53006   0.38555  -3.968 7.67e-05 ***
## factor(year)2010      0.10721   0.37021   0.290 0.772188
## factor(year)2011      0.40568   0.50073   0.810 0.418003
## factor(year)2012      1.39874   0.38994   3.587 0.000348 ***
## factor(year)2013     -0.04662   0.46402  -0.100 0.919985
## factor(year)2014      0.44153   0.39122   1.129 0.259296
## factor(year)2015     -0.63426   0.44941  -1.411 0.158420
## factor(year)2016      0.79185   0.44040   1.798 0.072426 .
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                               edf Ref.df      F p-value
```

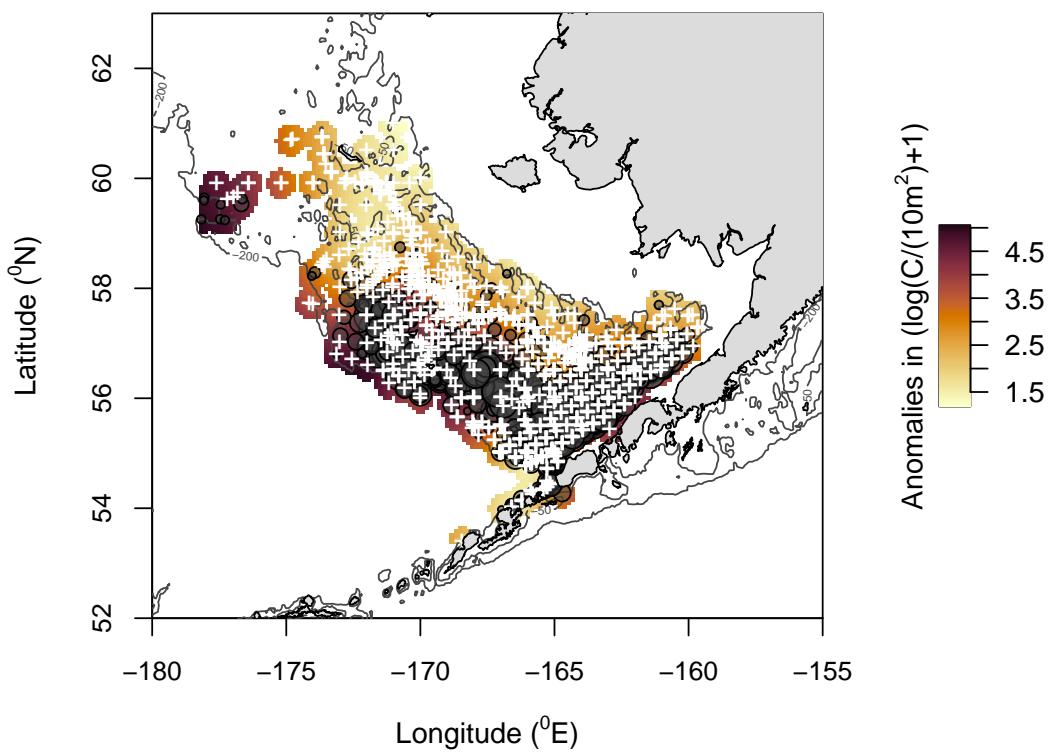
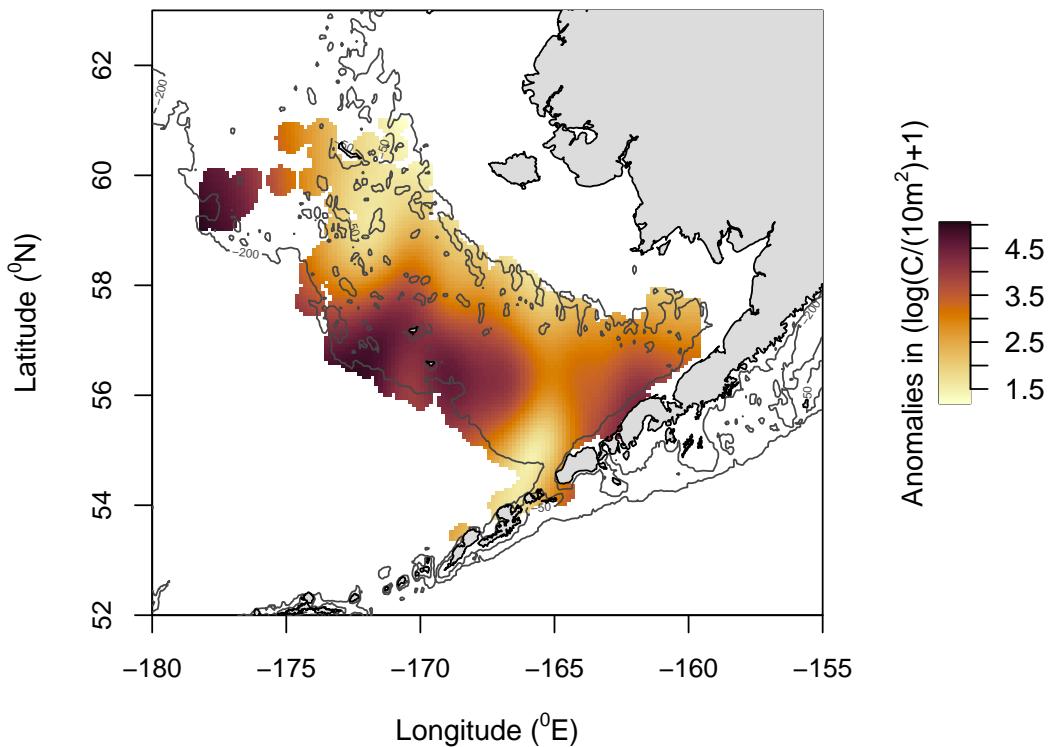
```
## s(lon,lat)           24.586 27.791 11.47  <2e-16 ***
## s(doy)                4.472  5.172 23.78  <2e-16 ***
## s(bottom_depth)       6.918  7.925 19.78  <2e-16 ***
## te(salinity,temperature) 17.015 19.328 10.80  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.139   Deviance explained = 60.6%
## -REML = 5036.7   Scale est. = 1.7035    n = 1244
```

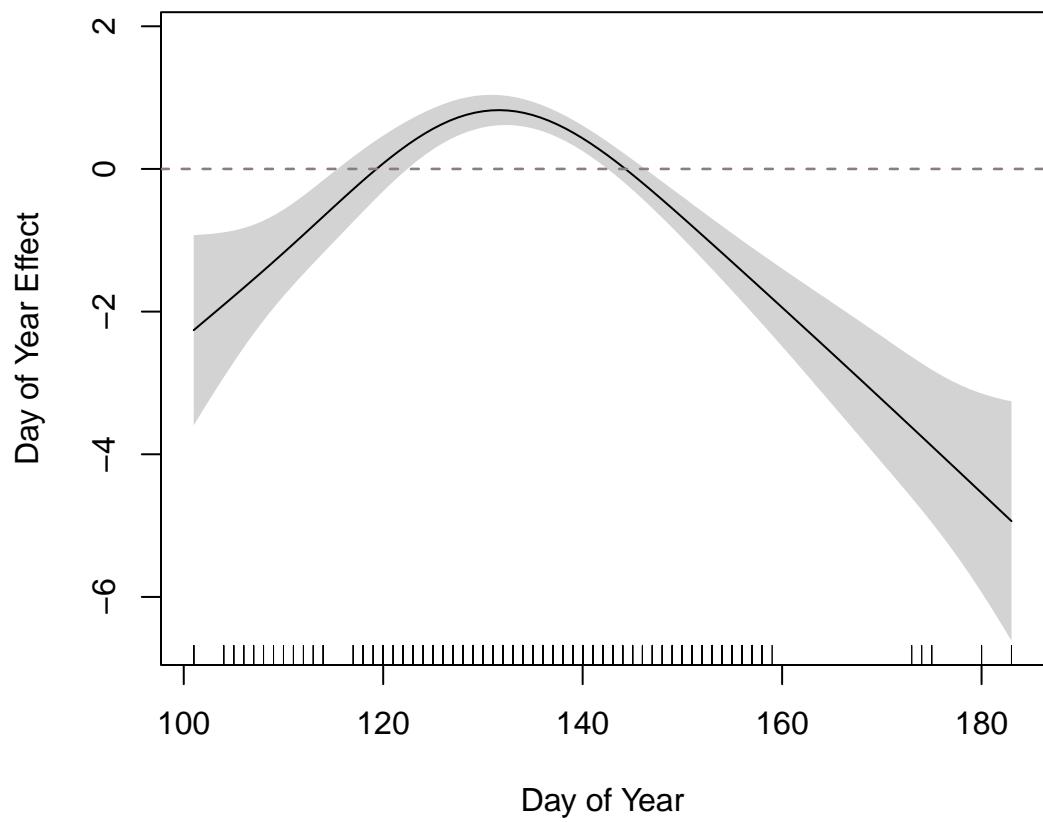
```
AIC(lv.2d)
```

```
## [1] 10049.01
```

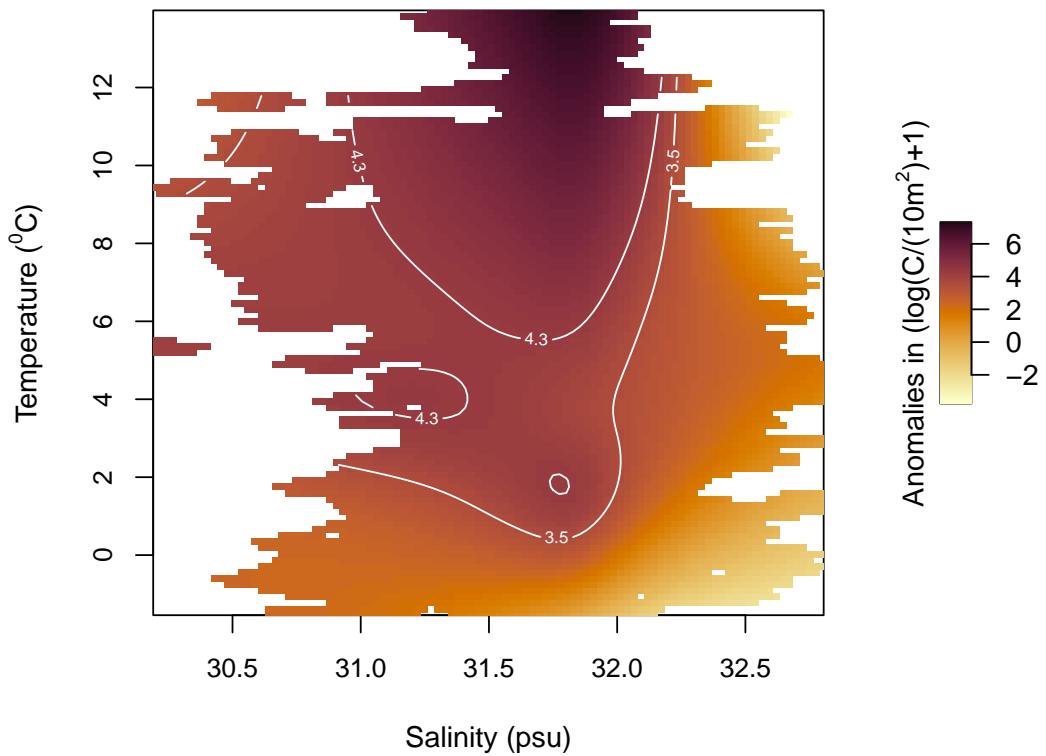
The following plot is the predicted Flathead sole larval biogeography based on the best performing model, the bivariate salinity-temperature GAM. Observations (log transformed, n+1) are shown as well.

### Predicted FHS Larval Biogeography, 2D Model

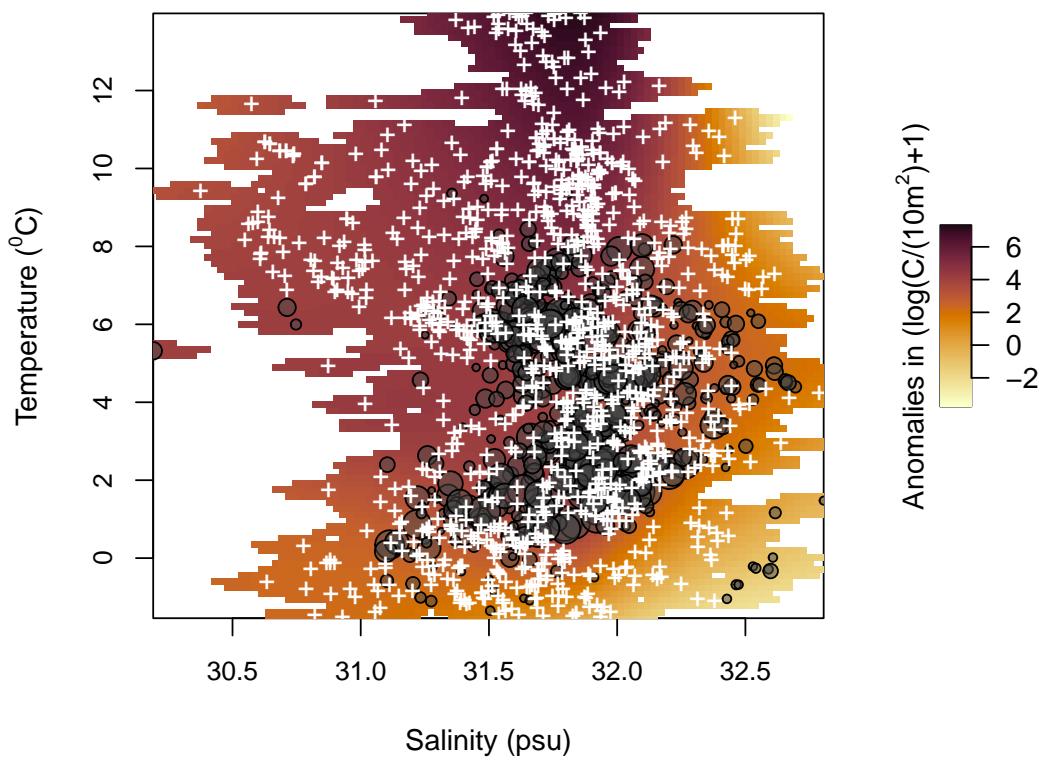




With this bivariate model, we can also calculate the predicted anomalous larval catch (more or less than expected) on a salinity-temperature plot. This figure shows that prediction, with observed larval catch ( $\log(n=1)$ ) overlaid.



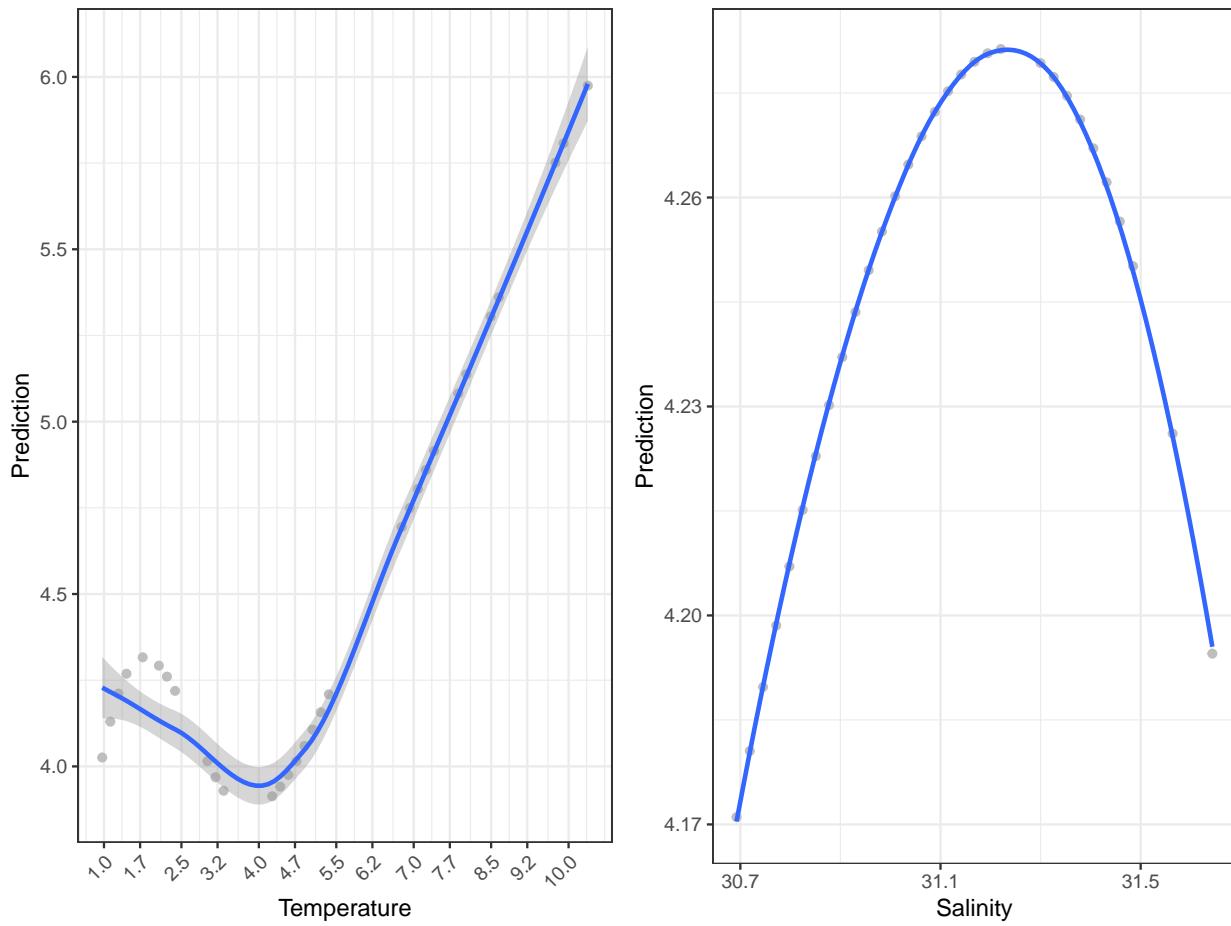
**Larval Biogeography By Temperature and Salinity**



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.

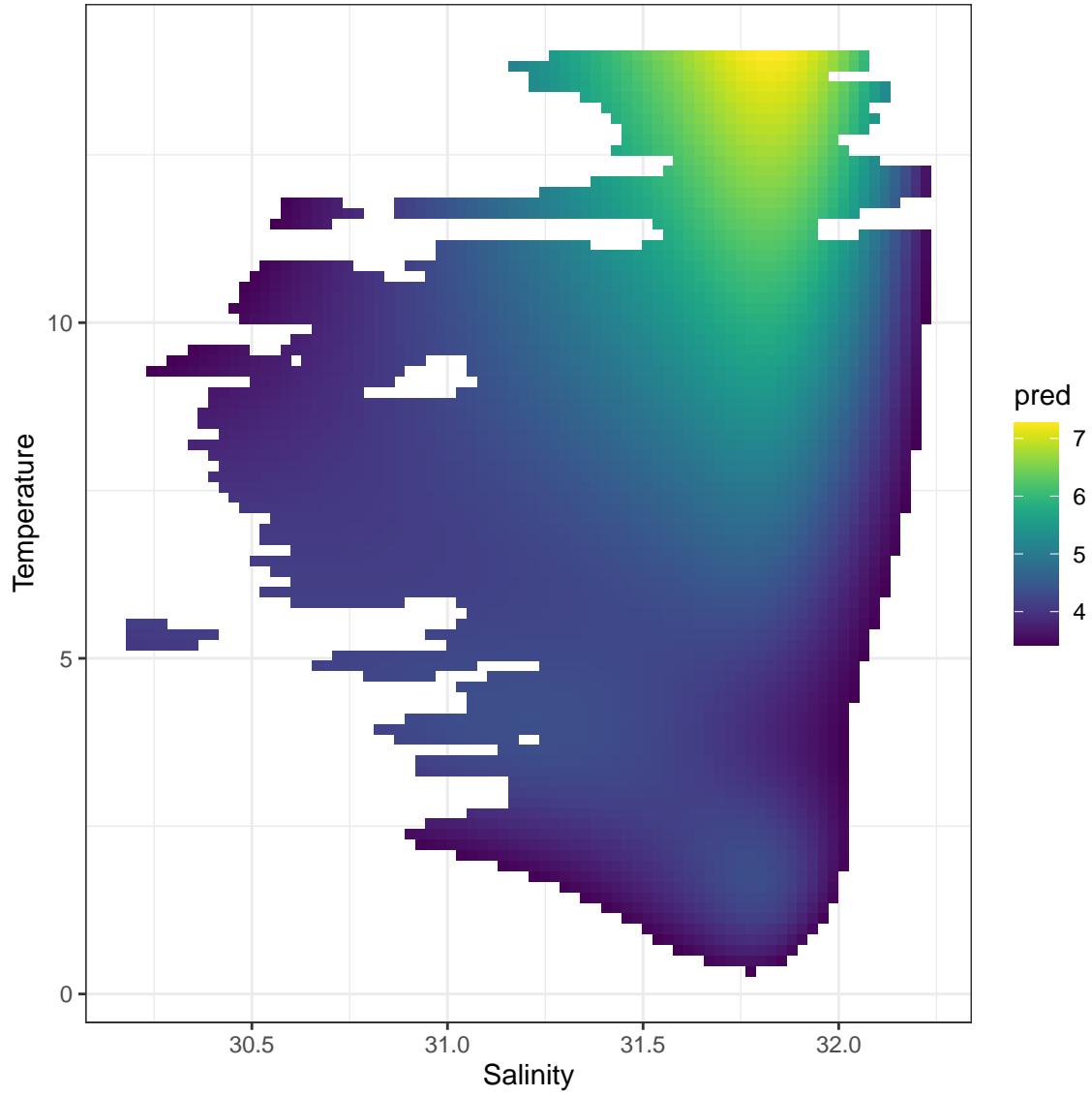
```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



For the bivariate analysis:

```
##   salinity temperature      dist year      lon      lat doy bottom_depth
## 1 31.82809     13.97930 0.0302099971 2005 -166.275 56.742 144          80
## 2 31.80168     13.97930 0.0262841445 2005 -166.275 56.742 144          80
## 3 31.85450     13.97930 0.0090287168 2005 -166.275 56.742 144          80
## 4 31.77528     13.97930 0.0001248852 2005 -166.275 56.742 144          80
## 5 31.82809     13.82251 0.0409729756 2005 -166.275 56.742 144          80
## 6 31.88091     13.97930 0.0254020714 2005 -166.275 56.742 144          80
##           pred
## 1 7.273352
## 2 7.262877
## 3 7.256758
## 4 7.228993
```

```
## 5 7.216647
## 6 7.210754
```



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, Flathead Sole

	Best Divided By Base	Best Divided By Second Best
Eggs	0.9712500	0.9943591
Larvae	0.9840347	0.9971243

Reduction in MSE (%):

```
## [1] 9.416975
```

## Appendices:

The following plots investigate the diverging phenological peaks in egg density below and above the threshold temperature in order to evaluate whether different peaks correlate to different spatial areas in the BS.

