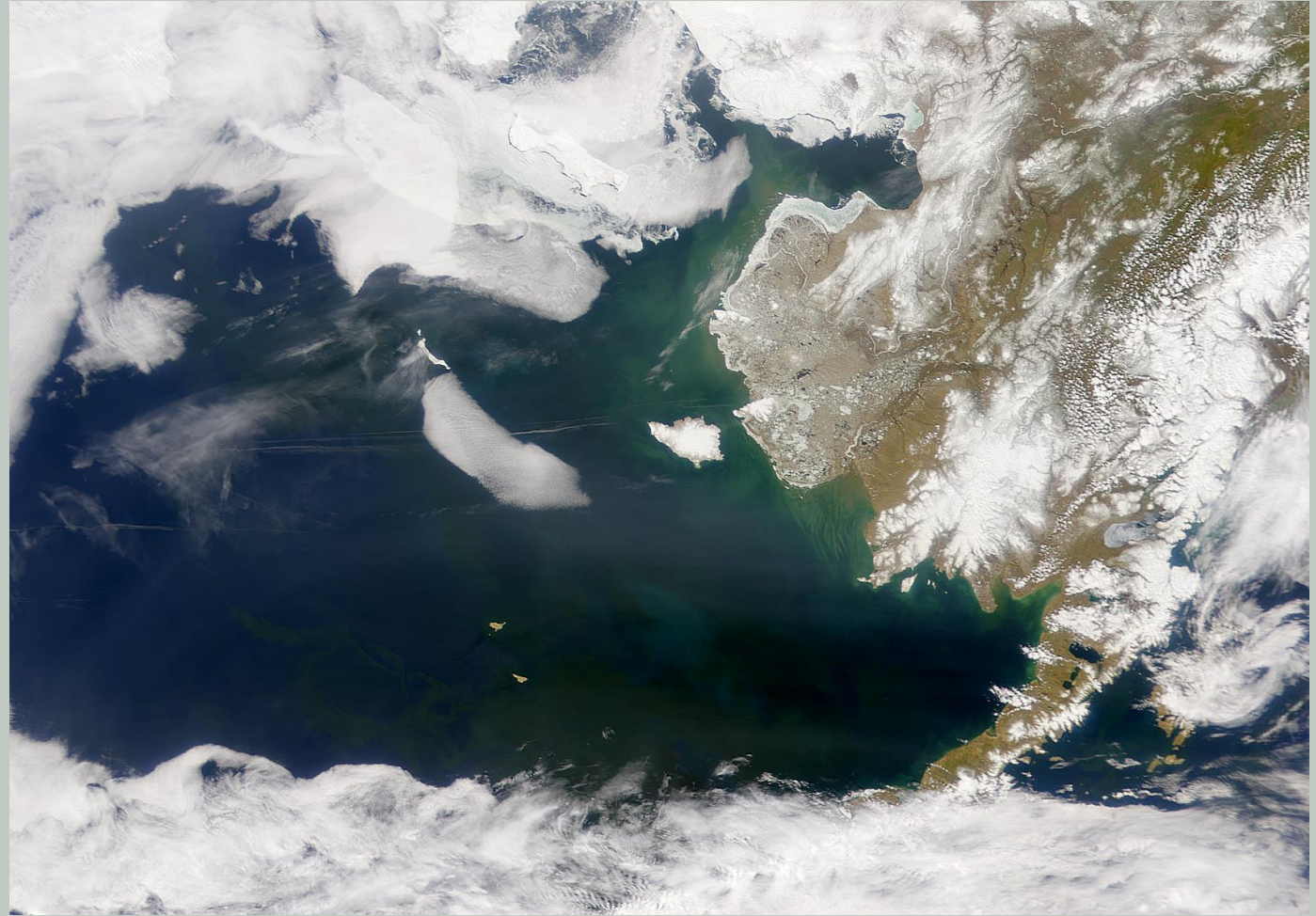# Progress Update

Laura Vary

November 4, 2021

Analytical Workflows

NASA Earth Observatory, 2002
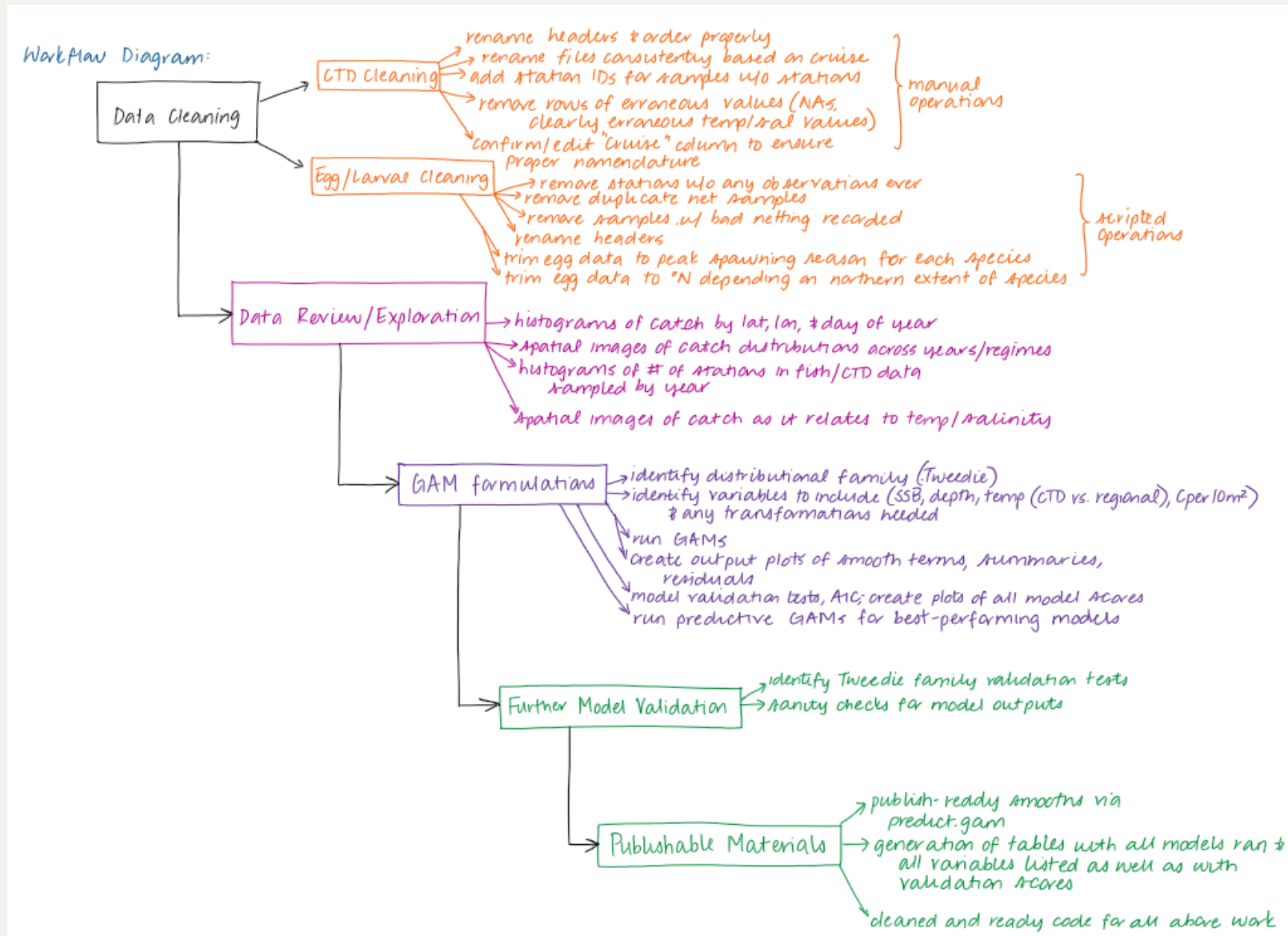
Background of Research

- Research Questions:
  - Spawning Behavior:
    - **How flexible is spawning geography and phenology?**
      - In practice: what GAM best explains variation in spawning location and time and how well does it perform?
  - Larval Biogeography:
    - **What types of water mass characteristics do larvae tend to be associated with?**
      - In practice: what GAM best explains larval biogeography, and where are those best explaining environmental condiitons located?
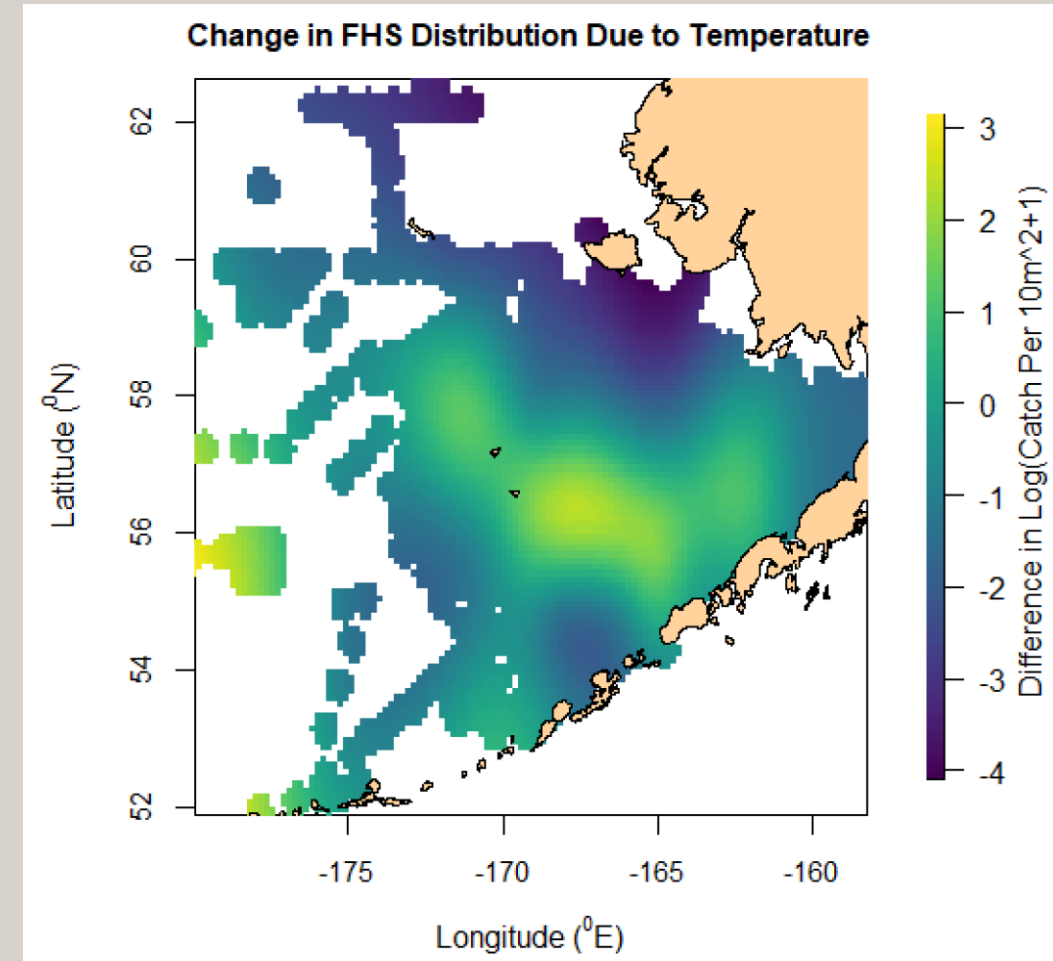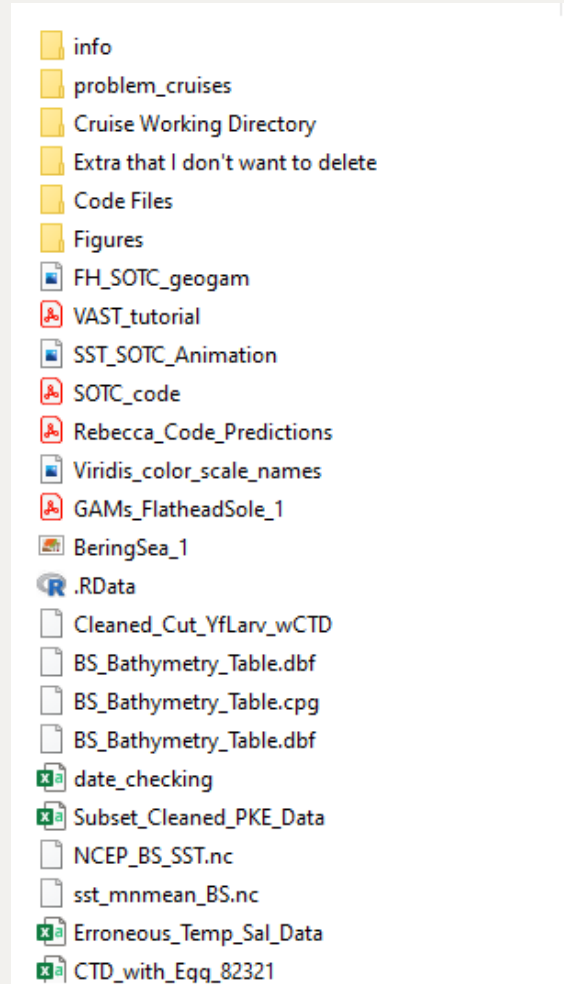

Example of Model Output – Threshold Geography GAM for Flathead Sole

Created a Workflow to Stick To:

Created and Improved GitHub Repo and Project Organization

Improved Readability of Code

```r
1
2    # EGGS: Spawning Behavior Flathead Sole ----------------------------------
3    #the following code creates generalized additive models for eggs and larvae of flathead sole.
4    #these analyses form the basis of my MS thesis.
5    #egg data uses an averaged sea surface temperature for the month of March in the Southeastern Bering Sea
6         #March index was chosen because it is two months before the peak of flathead egg CPUE, and thus March
7         #conditions are likely more relevant to spawning behavior than temperatures in later months.
8
9    # Load in data ----------------------------------------------------
10
11   fhsub<-read.csv(file='../Ichthyo Data/Cleaned_Cut_FhEggs.csv',header=TRUE,check.names=TRUE) #for egg GAMs
12   fhlarv.ctd<-read.csv(file='../Ichthyo Data/Cleaned_Cut_FhLarv_wCTD.csv',header=TRUE,check.names=TRUE) #for larval GAMs
13
14   ##Load in regional temperature index for March (2 mos before peak egg CPUE in May), for egg GAMs
15   reg.sst<-read.csv('../Environmental Data/Mar_SST_RegionalIndex_NCEP_BS.csv',header=TRUE,check.names=TRUE)
16   head(reg.sst) #range of regional average: lon: -180 to -151, lat: 50.5 to 67.5
17
18   for(i in 1:nrow(fhsub)){
19     fhsub$reg.SST[i]<-reg.sst$SST[reg.sst$year==fhsub$year[i]]}
20
21
22   # Spawning Behavior - Egg GAMs ----------------------------------------
23   #the following code generations GAMs that quantify variation in space and time of spawning behavior
24       #(as proxied by egg CPUE) in relation to regional temperature index close to typical time of spawning
25
26   # Base Egg Model ----------------------------------------------
27
28   eg.base<-gam((Cper10m2+1)~factor(year)+s(lon,lat)+s(doy)+s(bottom_depth,k=5),
29               data=fhsub,family=tw(link='log'),method='REML')
30
```

```r
46   # Threshold Phenology Egg GAM ------------------------------------------
47
48   temps<-sort(unique(reg.sst$SST)) #order by unique values of regional SST (March temps here)
49
50   bd<-4 #dictates essentially how many unique temperatures we check. can vary this depending on your system, how long you want the model to take,
51     #and probably some other factors that I don't know as well.. relatively arbitrary. smaller values for bd check more unique temperature values
52
53   temps.in<-temps[bd:(length(temps)-bd)] #vector whose length is dependent on bd; this vector is what the for loop uses to test many models at different thresholds
54
55   aic.pheno<-NA*(temps.in) #create an empty vector with the proper length to fill in later
56
57   thr.pheno<-as.list(1:(length(temps.in))) #another list with the same dimensions to fill in later
58
59   #this function below takes the unique temperature values and creates a model for each temperature as a threshold.
60   #then, we use AIC to ask which model and at what threshold best explains variation in phenology?
```

Continuous Goals

Add comments and introductions to all code files

Add environmental data sources to .readme file and update with any new tools

Create streamlined figure creation workflow (once results are finalized)