# Marine Geospatial Ecology Tools

## Open Source Geoprocessing for Marine Ecology

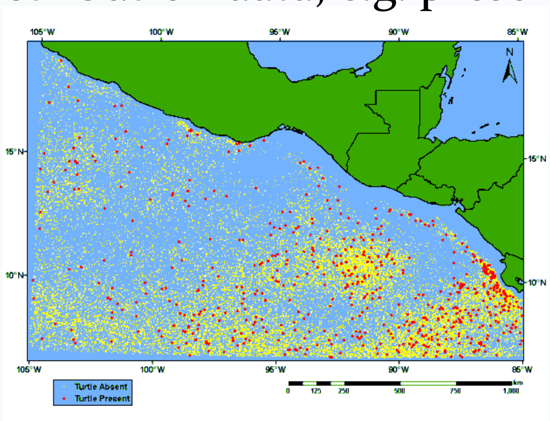
Elliott Hazen, Jason Roberts, Ben Best, Dan Dunn, Pat Halpin
Duke University Marine Geospatial Ecology Lab
Aug 8th, 2010

NICHOLAS SCHOOL OF THE ENVIRONMENT AND EARTH SCIENCES
DUKE UNIVERSITY

NICHOLAS SCHOOL OF THE ENVIRONMENT AND EARTH SCIENCES
DUKE UNIVERSITY

# Modeling habitat (overview)

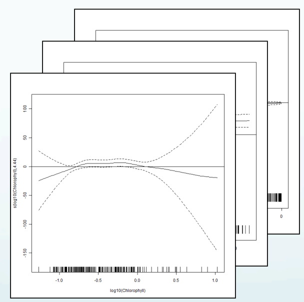
Distribution data, e.g. presence/absence




Fit

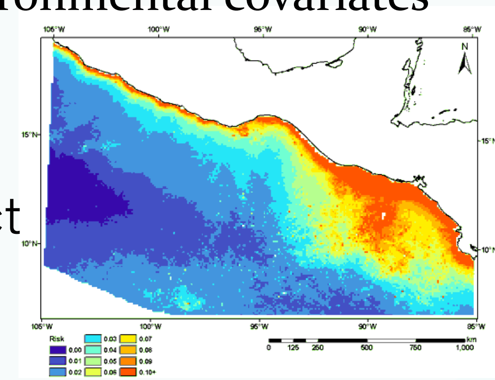
Statistical models

$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

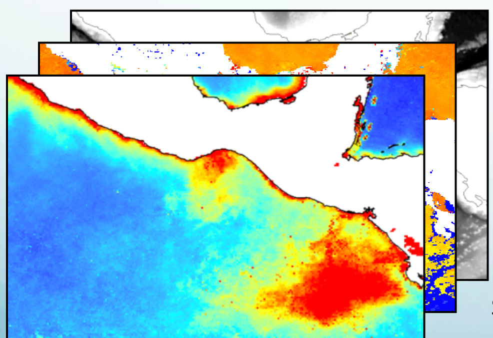


Predict

Probability of occurrence predicted from environmental covariates




Sampled predictive data

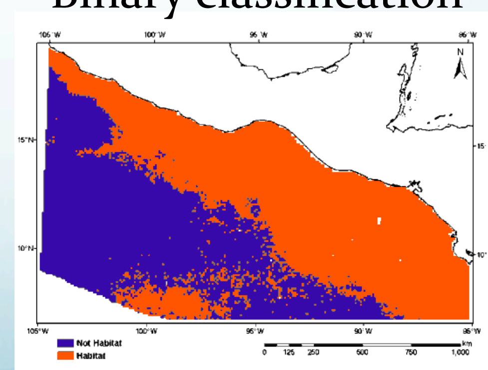


Bathymetry

SST

Chlorophyll

Binary classification




**Warning:** Habitat modeling is complicated! This simplified example is meant to briefly illustrate tools. Consult the literature for best practices!

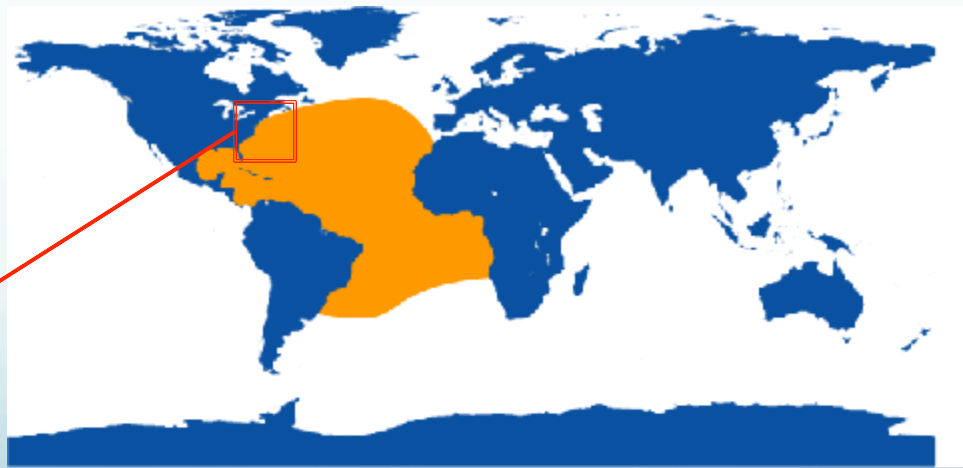
# Focal species: *Stenella frontalis*

Common name:
Atlantic Spotted Dolphin

Photo: Garth Mix

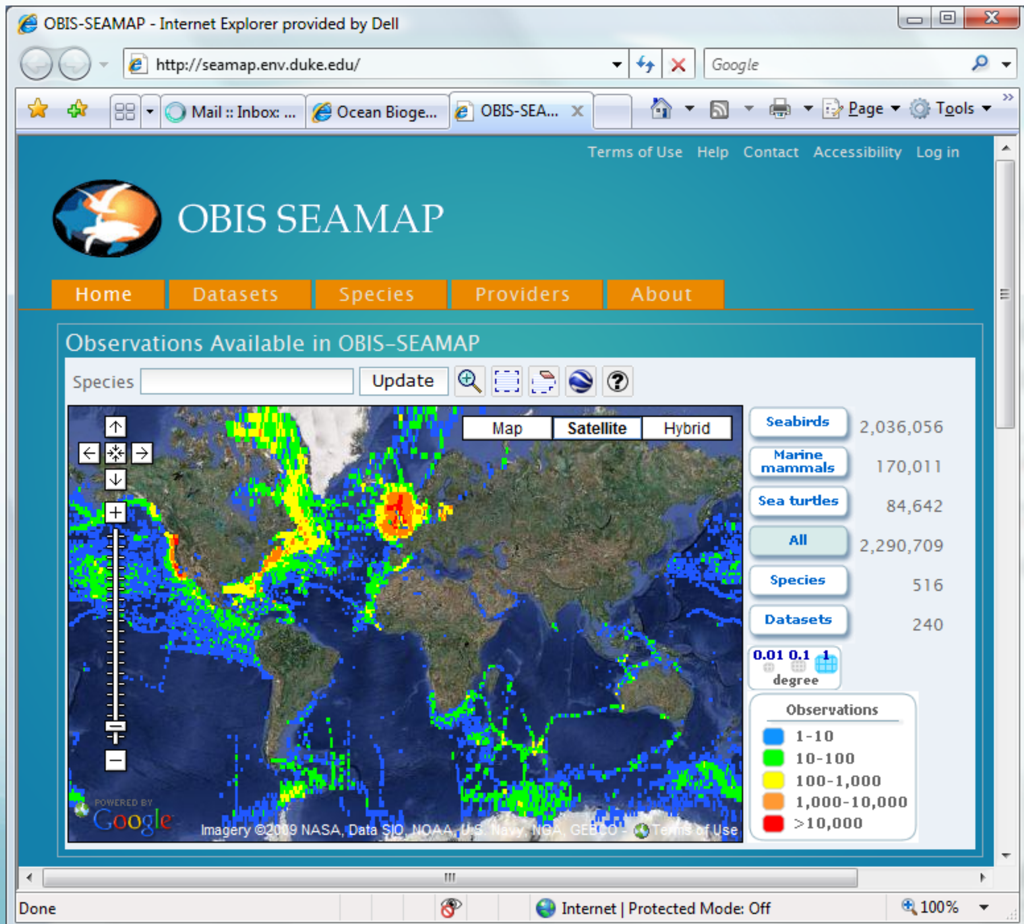
Distribution:
Tropical and warm temperate Atlantic

Study area:
Eastern U.S.



Map: OBIS-SEAMAP

# Species observation data

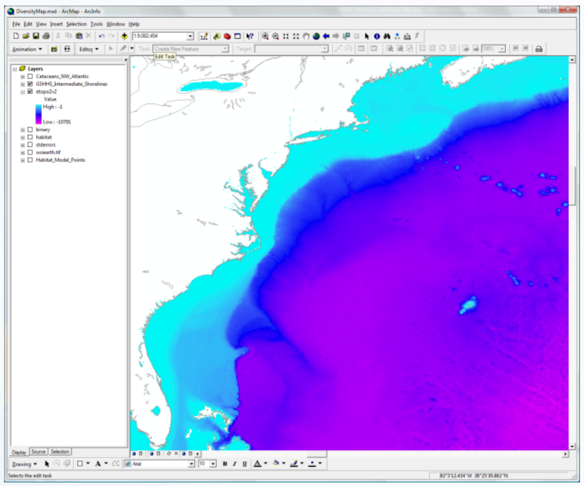


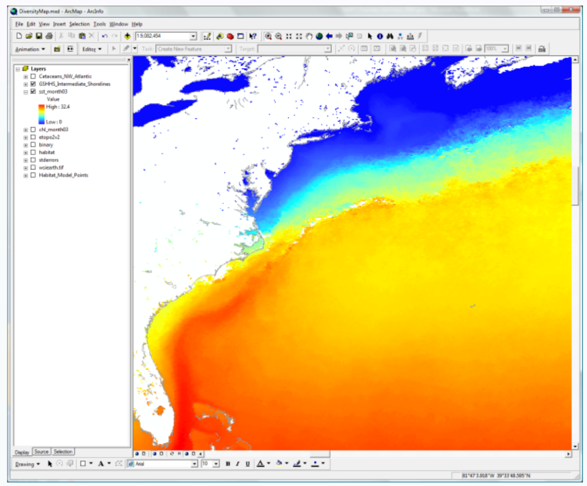

OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

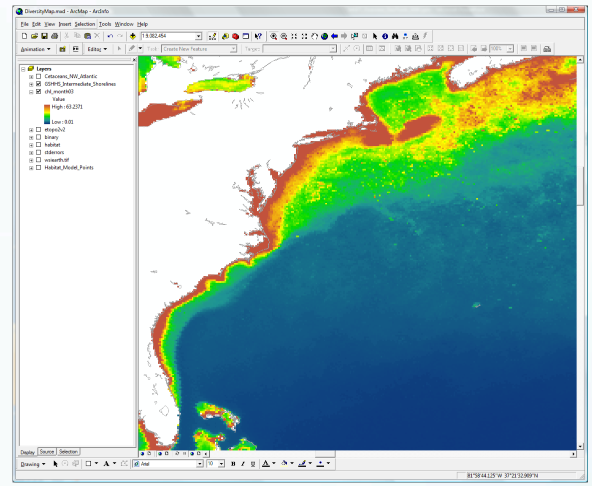
The Ocean Biogeographic Information System (OBIS) is a global database of marine species observations.

The OBIS-SEAMAP system at Duke University holds the records for seabirds, marine mammals, and sea turtles, including records gathered during NOAA cruises.

# Environmental predictor variables







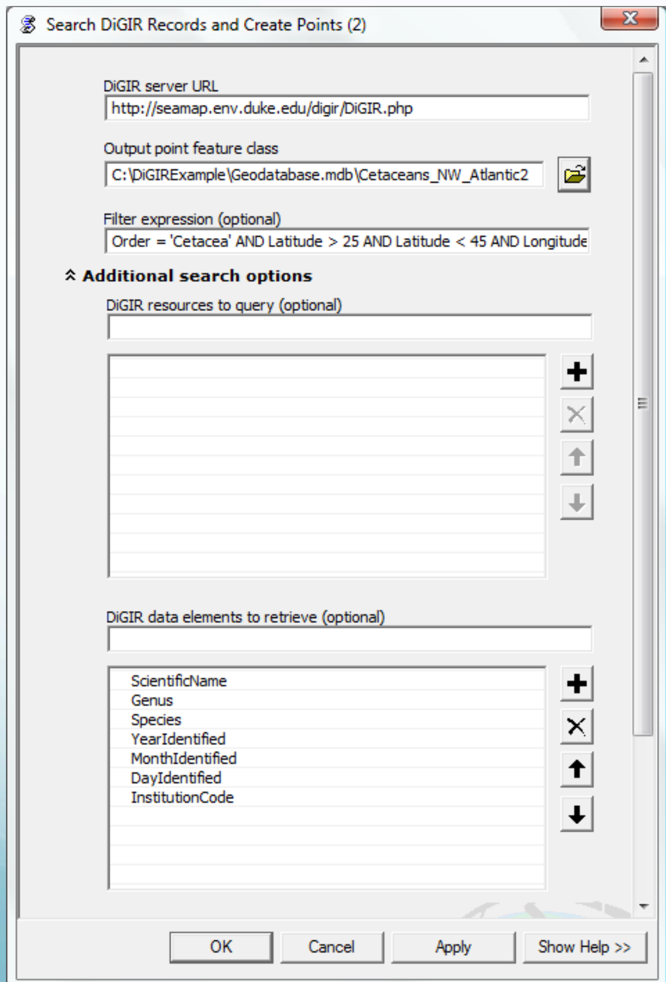
**Bathymetry:**
ETOPO2V2 from NOAA NGDC

**SST:**
Monthly climatological 4km AVHRR Pathfinder from NOAA NODC

**Chlorophyll:**
Monthly climatological SeaWiFS chlorophyll-*a* from NASA GSFC

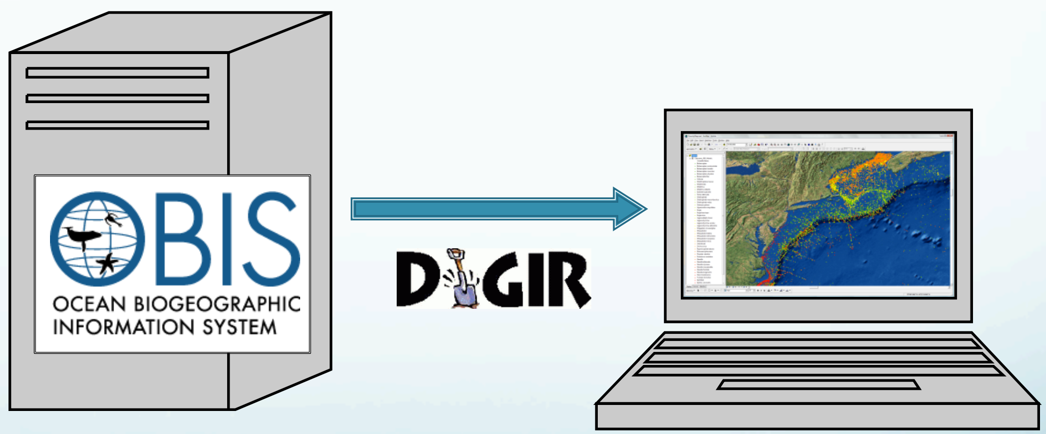
Images shown above are for month of March

# Step 1: Download species points

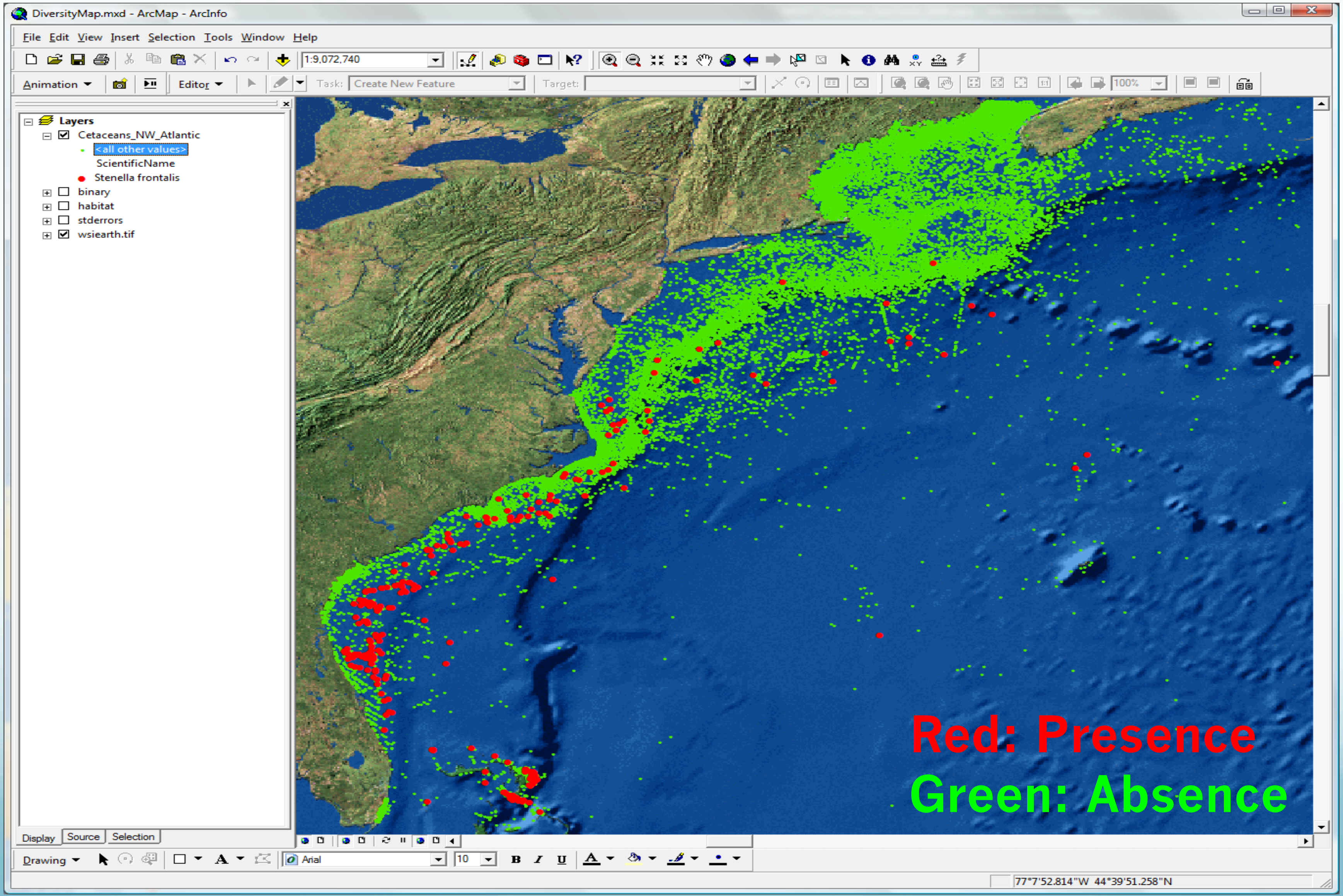



Download points using MGET tool:

- Presence: Records of *Stenella frontalis*
- Absence: Records of other cetaceans

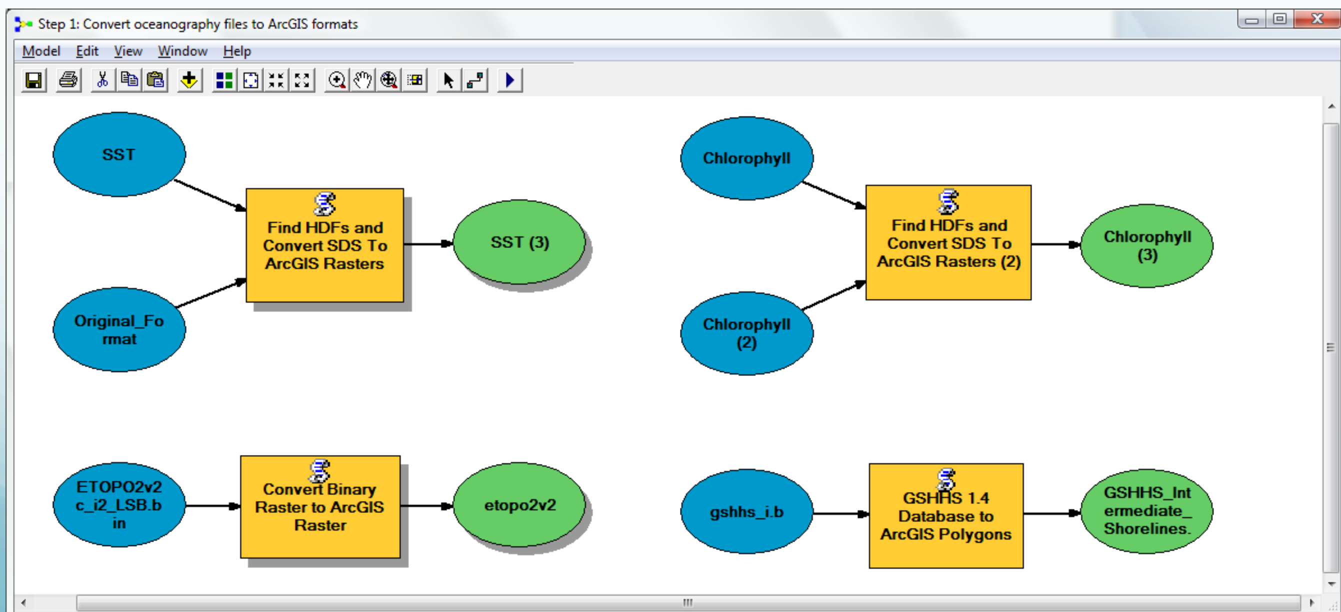



The tool uses the DiGIR protocol to retrieve data from OBIS servers


DiversityMap.mxd - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:9,072,740
Animation
Editor
Task: Create New Feature
Target:
Layers
Cetaceans_NW_Atlantic
<all other values>
ScientificName
Stenella frontalis
binary
habitat
stderrors
wsiearth.tif
Display
Source
Selection
Drawing
Arial
10
77°7'52.814"W 44°39'51.258"N
Red: Presence
Green: Absence

# Step 2: Convert oceanography to Arc rasters

1. Download with FTP from NOAA and NASA:
   - ETOPO2 bathymetry – 1 binary file
   - AVHRR Pathfinder monthly climatological SST – 12 HDF files
   - SeaWiFS monthly climatological chlorophyll – 12 HDF files
2. Convert to ArcGIS rasters using MGET tools:



# Step 3: Sample oceanography at points

Attributes of Habitat_Model_Points

| OBJECTID * | Shape * | ScientificName | YearIdentified | MonthIdentified | DayIdentified | Presence | Depth | SST | Chlorophyll |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Point | Tursiops truncatus | 1998 | 8 | 18 | 0 | | | |
| 2 | Point | Tursiops truncatus | 1998 | 8 | 19 | 0 | | | |
| 3 | Point | Physeter catodon | 1998 | 7 | 10 | 0 | | | |
| 4 | Point | Physeter catodon | 1998 | 7 | 10 | 0 | | | |
| 5 | Point | Physeter catodon | 1998 | 7 | 10 | 0 | | | |
| 6 | Point | Balaenoptera acutorostrata | 1998 | 7 | 10 | 0 | | | |

Record: 1 Show: All Selected Records (0 out of *4000 Selected) Options

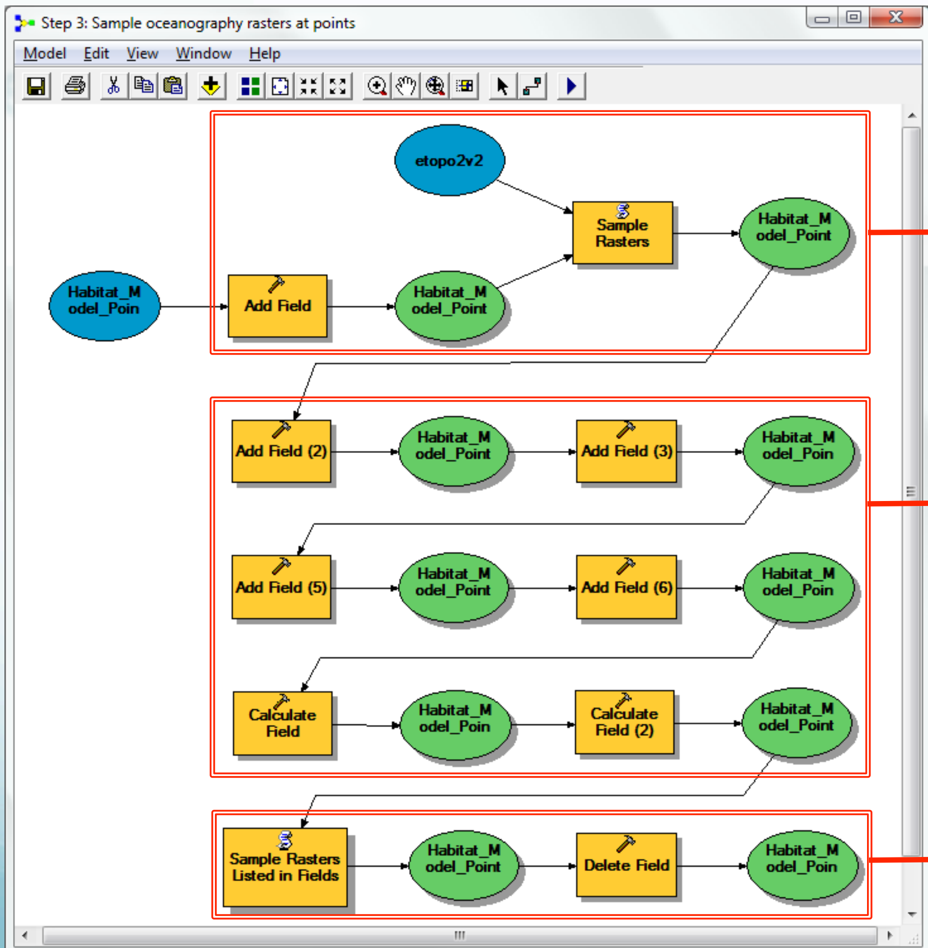
- Need to sample rasters and populate fields
- Must sample SST and chlorophyll by date

Attributes of Habitat_Model_Points

| OBJECTID * | Shape * | ScientificName | YearIdentified | MonthIdentified | DayIdentified | Presence | Depth | SST | Chlorophyll |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Point | Tursiops truncatus | 1998 | 8 | 18 | 0 | -567 | 29.700001 | 0.094755 |
| 2 | Point | Tursiops truncatus | 1998 | 8 | 19 | 0 | -541 | 29.85 | 0.096544 |
| 3 | Point | Physeter catodon | 1998 | 7 | 10 | 0 | -3187 | 26.25 | 0.209108 |
| 4 | Point | Physeter catodon | 1998 | 7 | 10 | 0 | -3283 | 26.775 | 0.187317 |
| 5 | Point | Physeter catodon | 1998 | 7 | 10 | 0 | -3477 | 27 | 0.183823 |
| 6 | Point | Balaenoptera acutorostrata | 1998 | 7 | 10 | 0 | -3633 | 26.924999 | 0.173819 |

Record: 1 Show: All Selected Records (0 out of *4000 Selected) Options

# Step 3: Sample oceanography at points

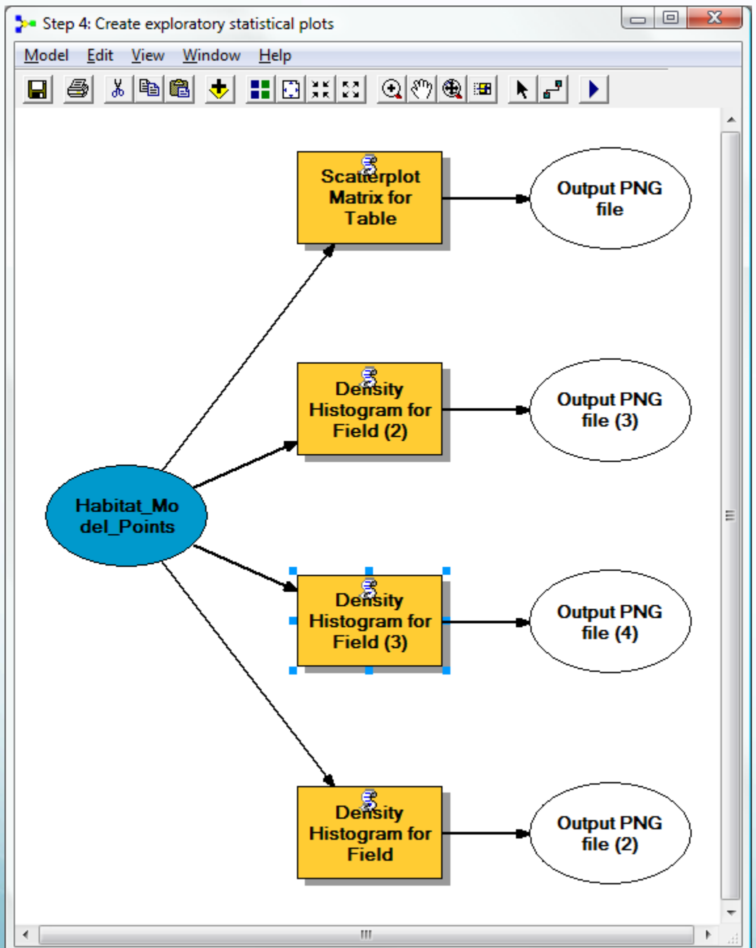


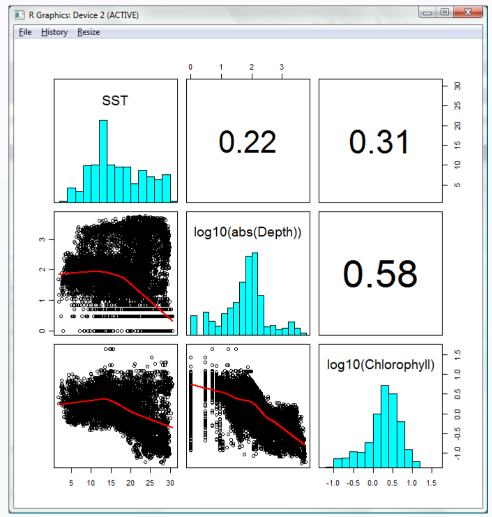

Sampling bathymetry is easy because it is static

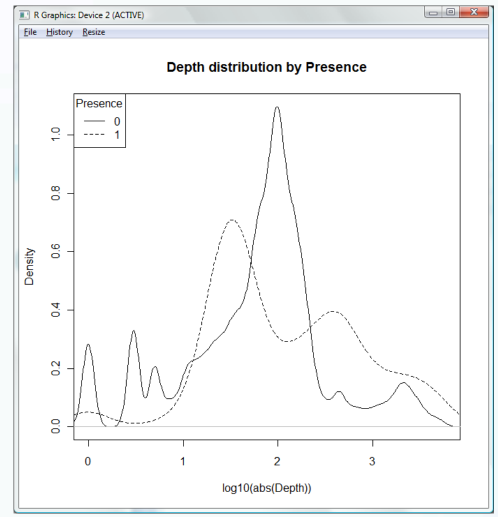
To sample dynamic data such as SST and chlorophyll, you must first calculate the paths to rasters to sample from the points' dates

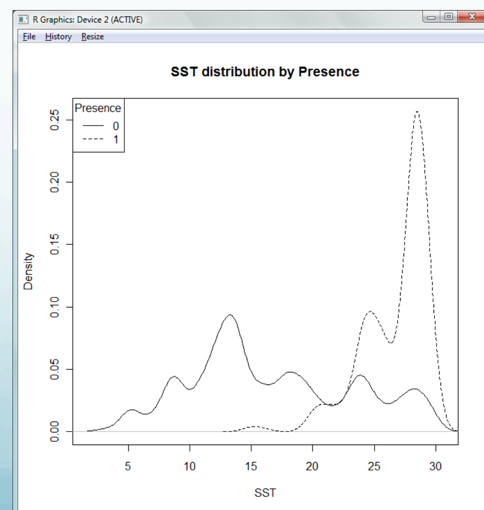
Then use an MGET batch sampling tool

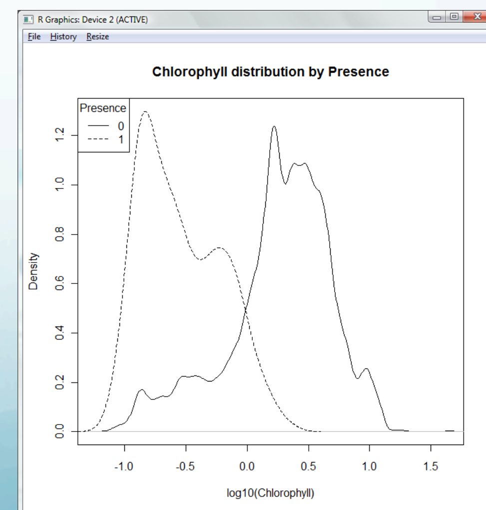
# Step 4: Create exploratory plots



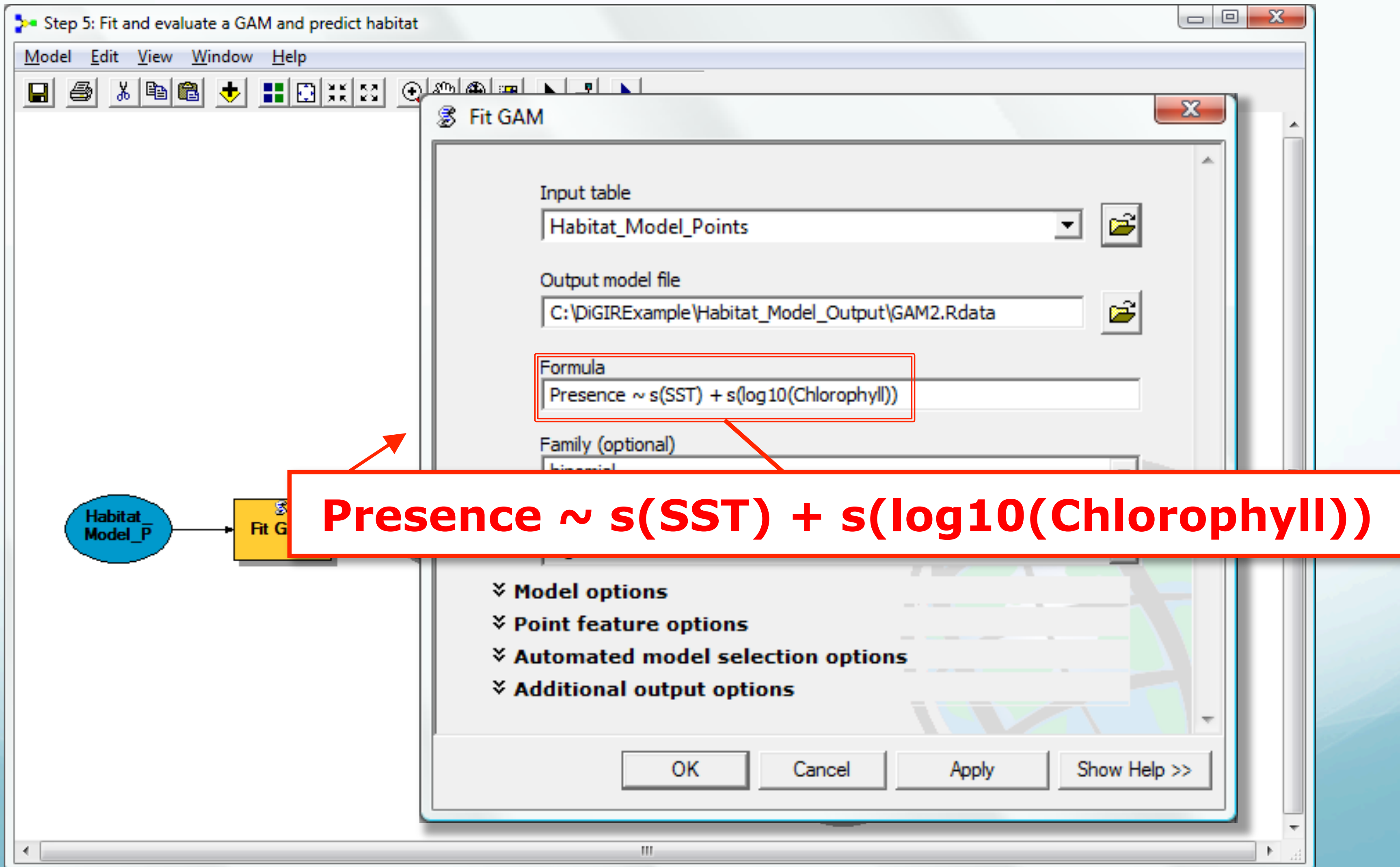



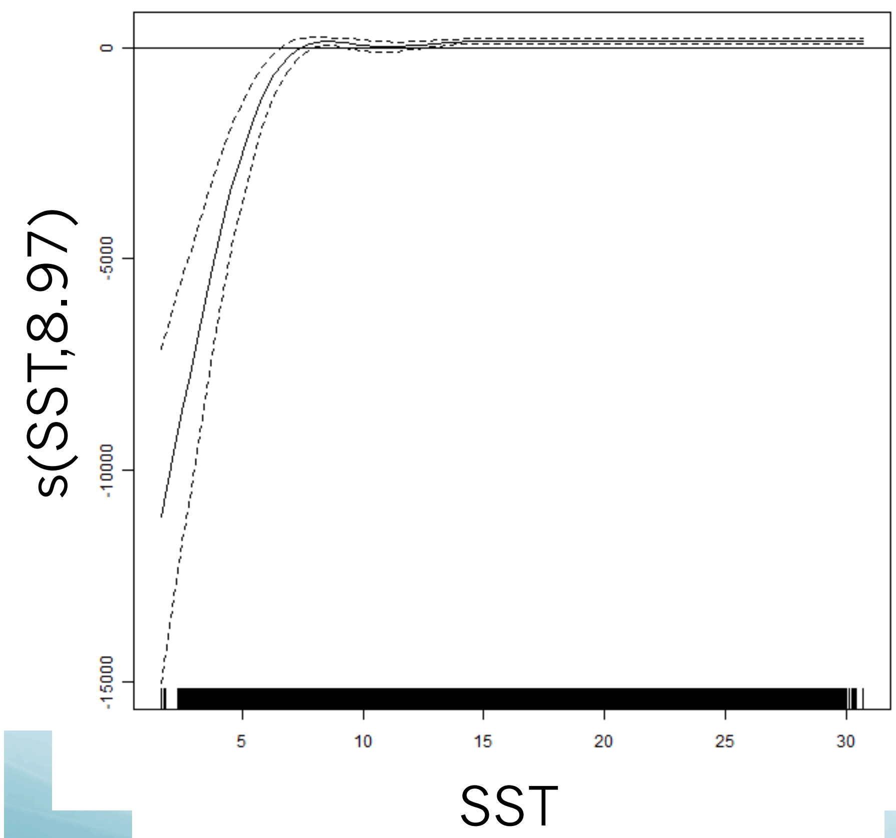



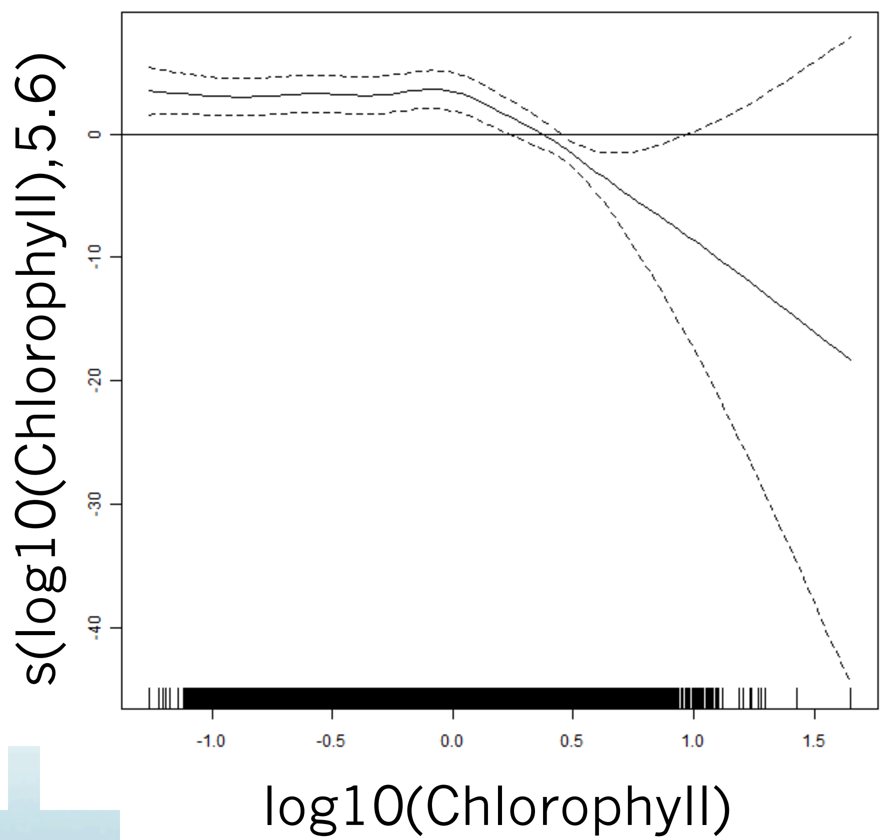

**Best predictors: SST and Chl**

# Step 5: Fit, evaluate, and predict model



# Partial plots produced by the Fit GAM tool

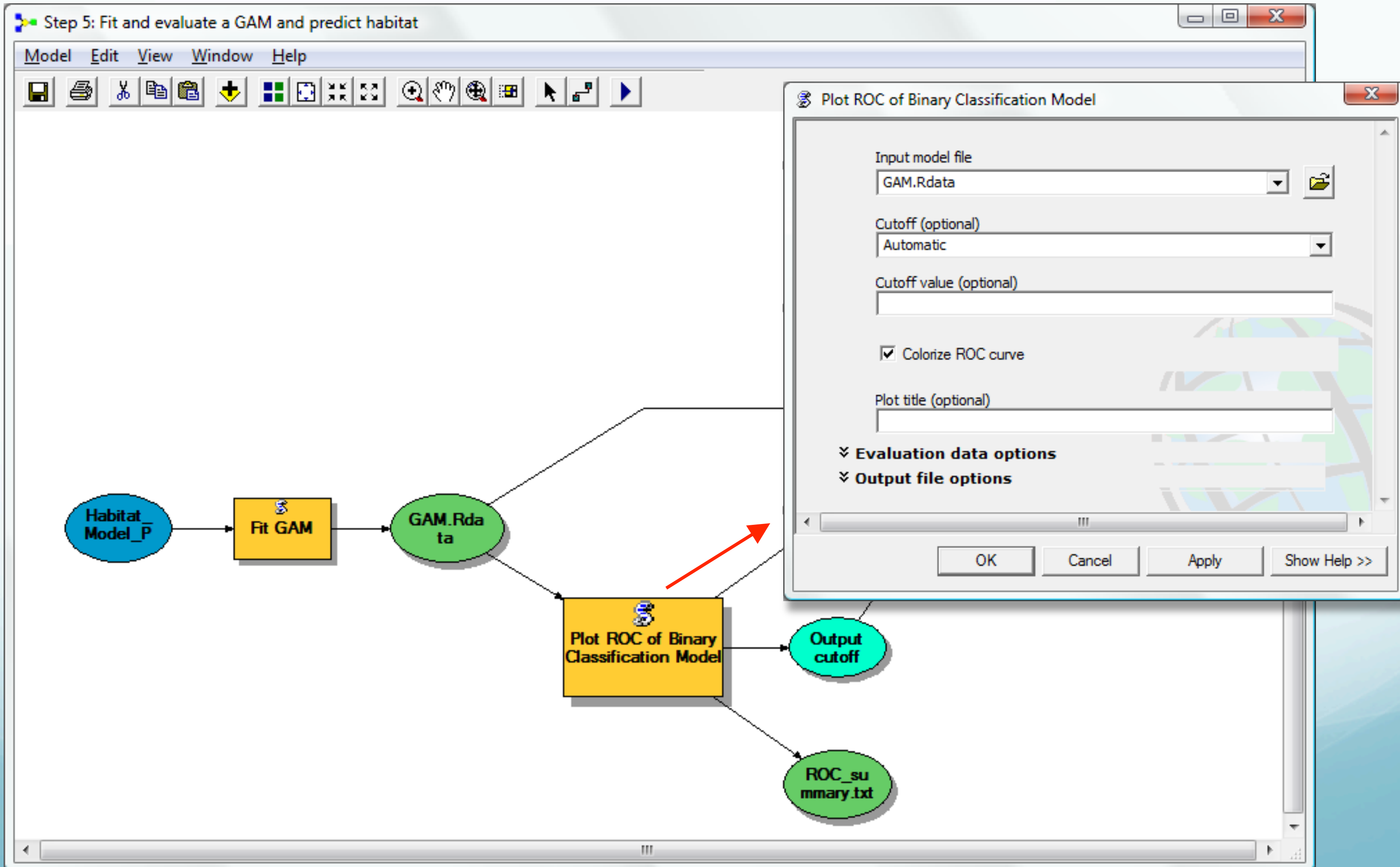


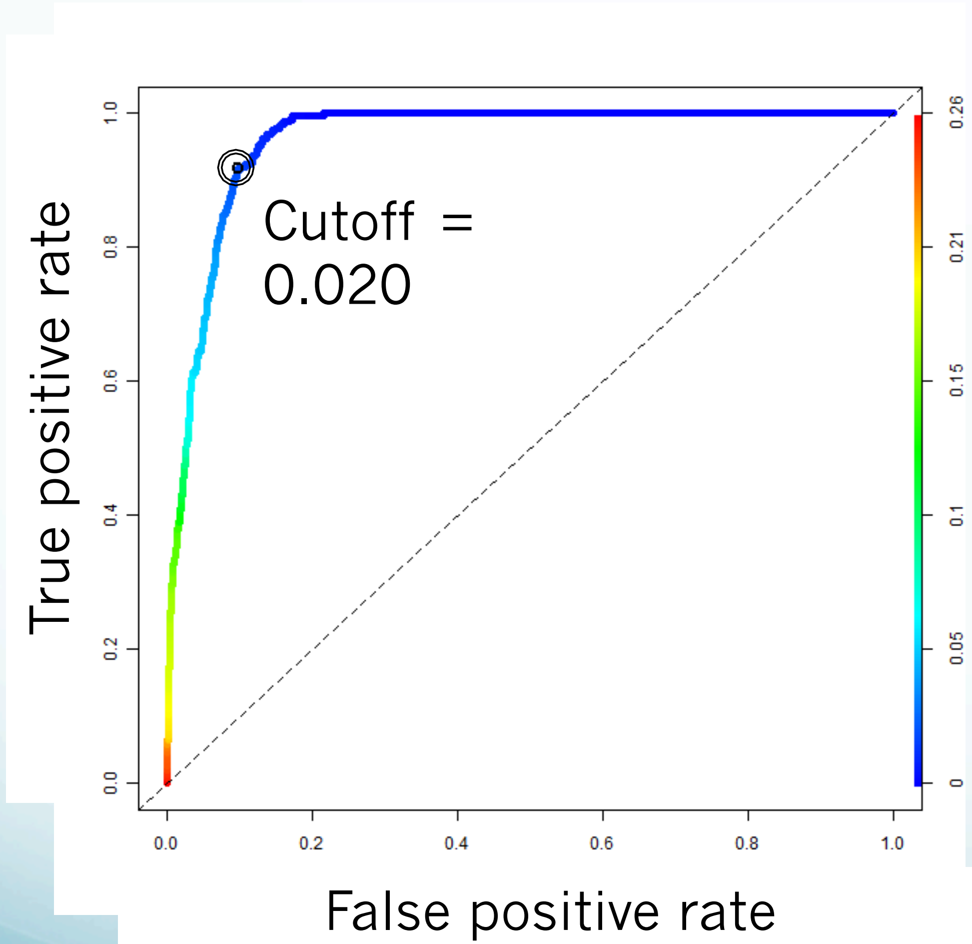

Presence more likely at higher SST

Presence more likely at lower Chl

# Plotting a receiver operating characteristic curve



# The ROC plot

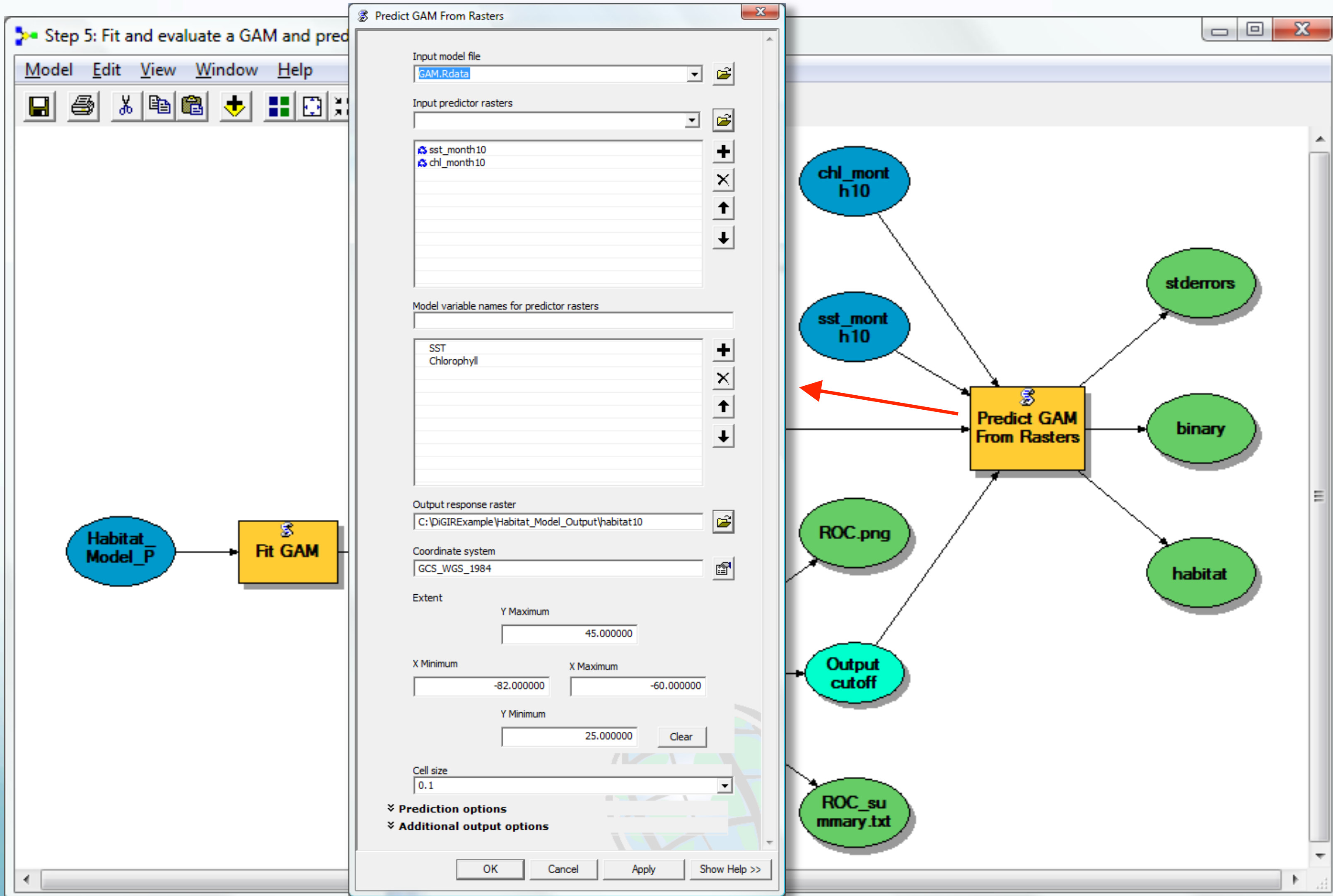



ROC summary stats for cutoff:

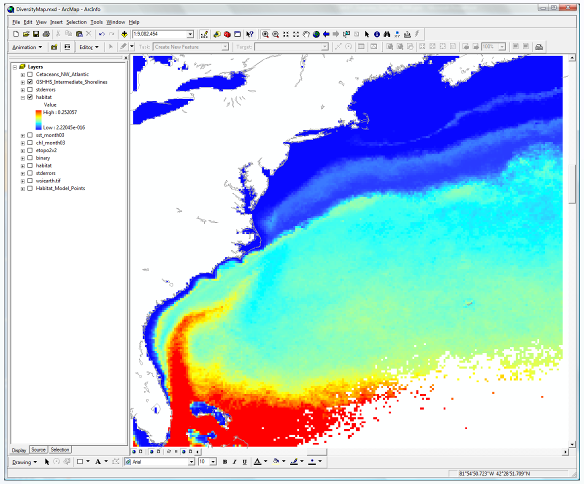
```
Model summary statistics:

Area under the ROC curve (auc)            = 0.960779
Mean cross-entropy (mxe)                  = 0.030566
Precision-recall break-even point (prbe) = 0.001866
Root-mean square error (rmse)             = 0.087781

Contingency table for cutoff = 0.019638:

              Actual P  Actual N     Total
Predicted P        287      3541      3828
Predicted N         26     32408     32434
      Total        313     35949     36262

Accuracy (acc)                                     = 0.901633
Error rate (err)                                   = 0.098367
Rate of positive predictions (rpp)                 = 0.105565
Rate of negative predictions (rnp)                 = 0.894435

True positive rate (tpr, or sensitivity)           = 0.916933
False positive rate (fpr, or fallout)              = 0.098501
True negative rate (tnr, or specificity)           = 0.901499
False negative rate (fnr, or miss)                 = 0.083067

Positive prediction value (ppv, or precision) = 0.074974
Negative prediction value (npv)                    = 0.999198
Prediction-conditioned fallout (pcfall)            = 0.925026
Prediction-conditioned miss (pcmiss)               = 0.000802

Matthews correlation coefficient (mcc)             = 0.246384
Odds ratio (odds)                                  = 101.026394
SAR                                                = 0.650065
```

By default, tool selects the cutoff closest to the point of perfect classification (0, 1)

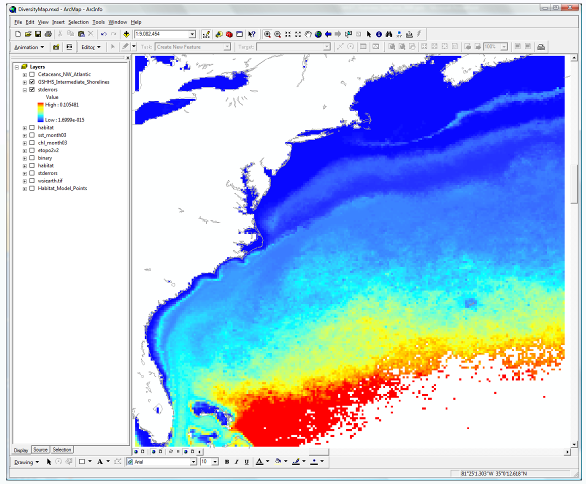
# Predicting presence for oceanographic rasters



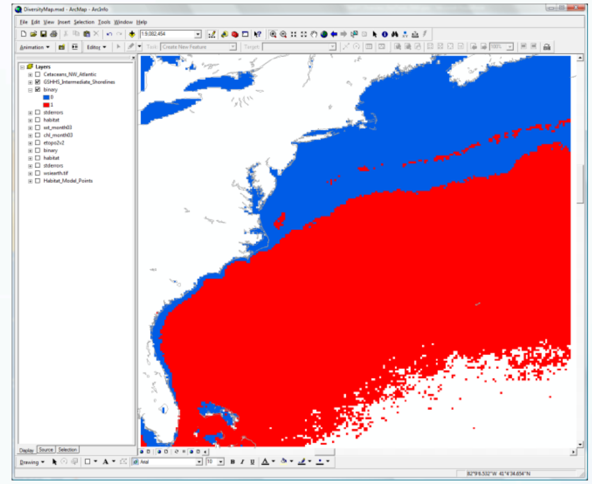
# Rasters output by the Predict GAM tool



**Predicted presence:**
Range: 0 - 0.25



**Standard errors:**
Range: 0 - 0.11



**Binary classification:**
Species range map produced by classifying presence into 0 or 1 according to ROC cutoff

Similar to OBIS-SEAMAP range map?

Predictions for October

# Acknowledgements

A special thanks to the many developers of the open source software that MGET is built upon! Also, folks that have helped with this talk:

MGEL lab, NOAA-SWFSC-ERD, and many, many others

Thanks to our funders: