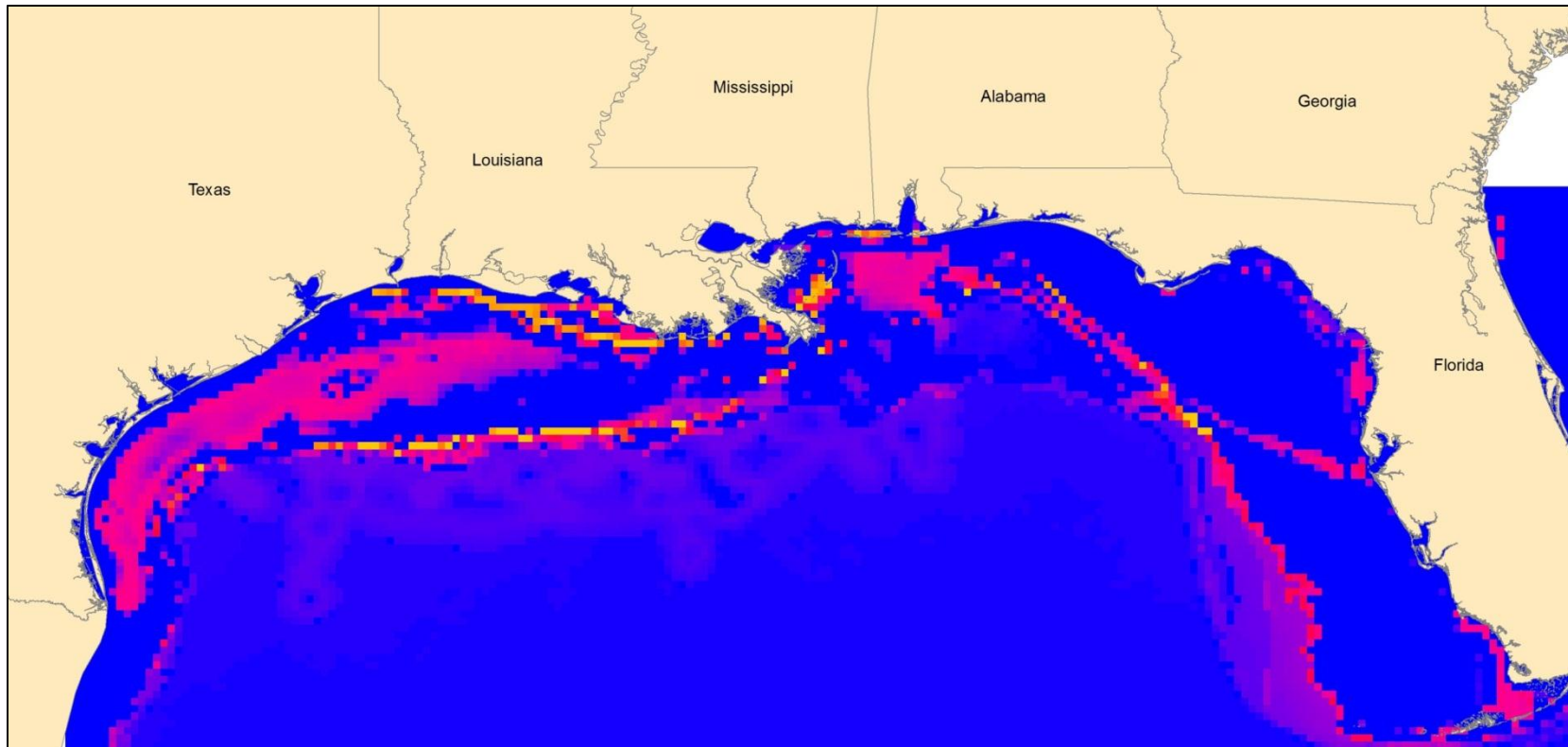
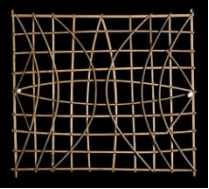


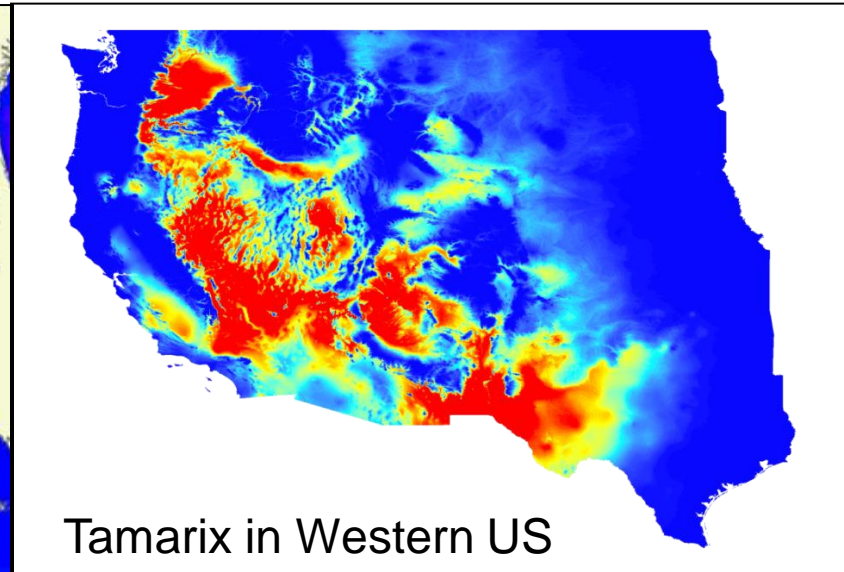
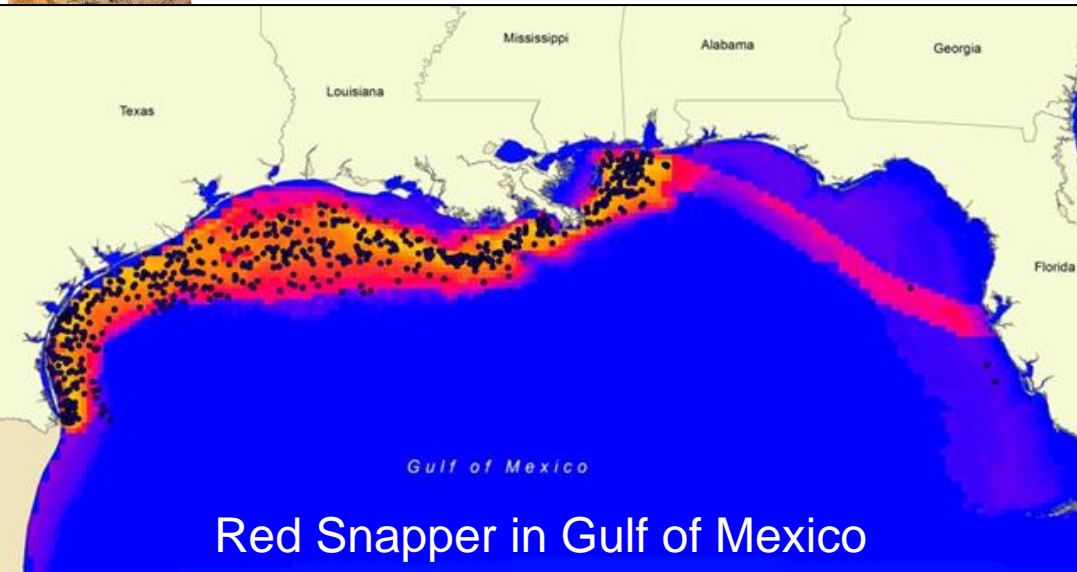
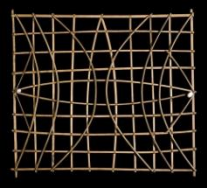
Managing Uncertainty in Habitat Suitability Models

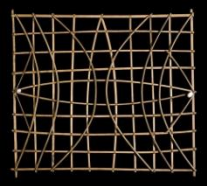
Jim Graham and Jake Nelson
Oregon State University



Habitat Suitability Modeling (HSM)

- Also known as:
 - Ecological Niche Modeling (ENM)
 - Species Distribution Modeling (SDM)





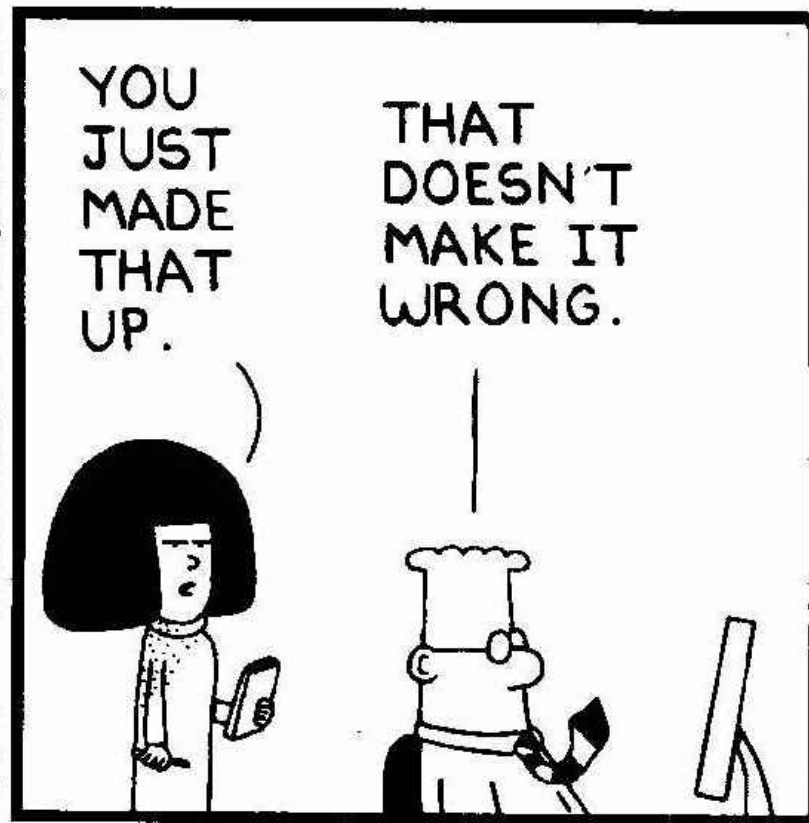
Uncertainty



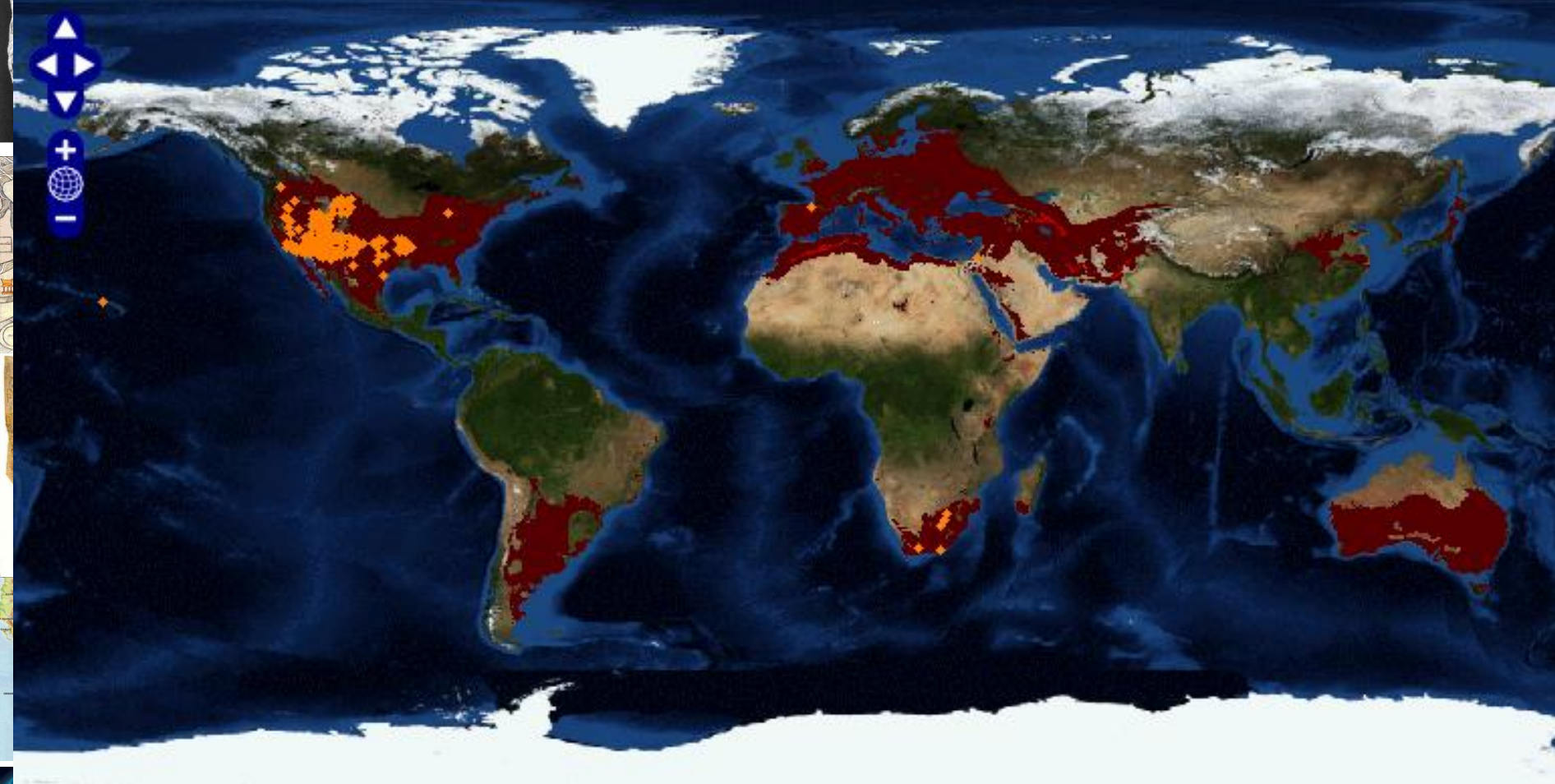
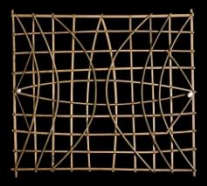
- How “certain” of the data are we?
- How much “error” does it contain?
- How well does the model match reality?
- Goal:
 - Understand and document uncertainties from data collection to publication



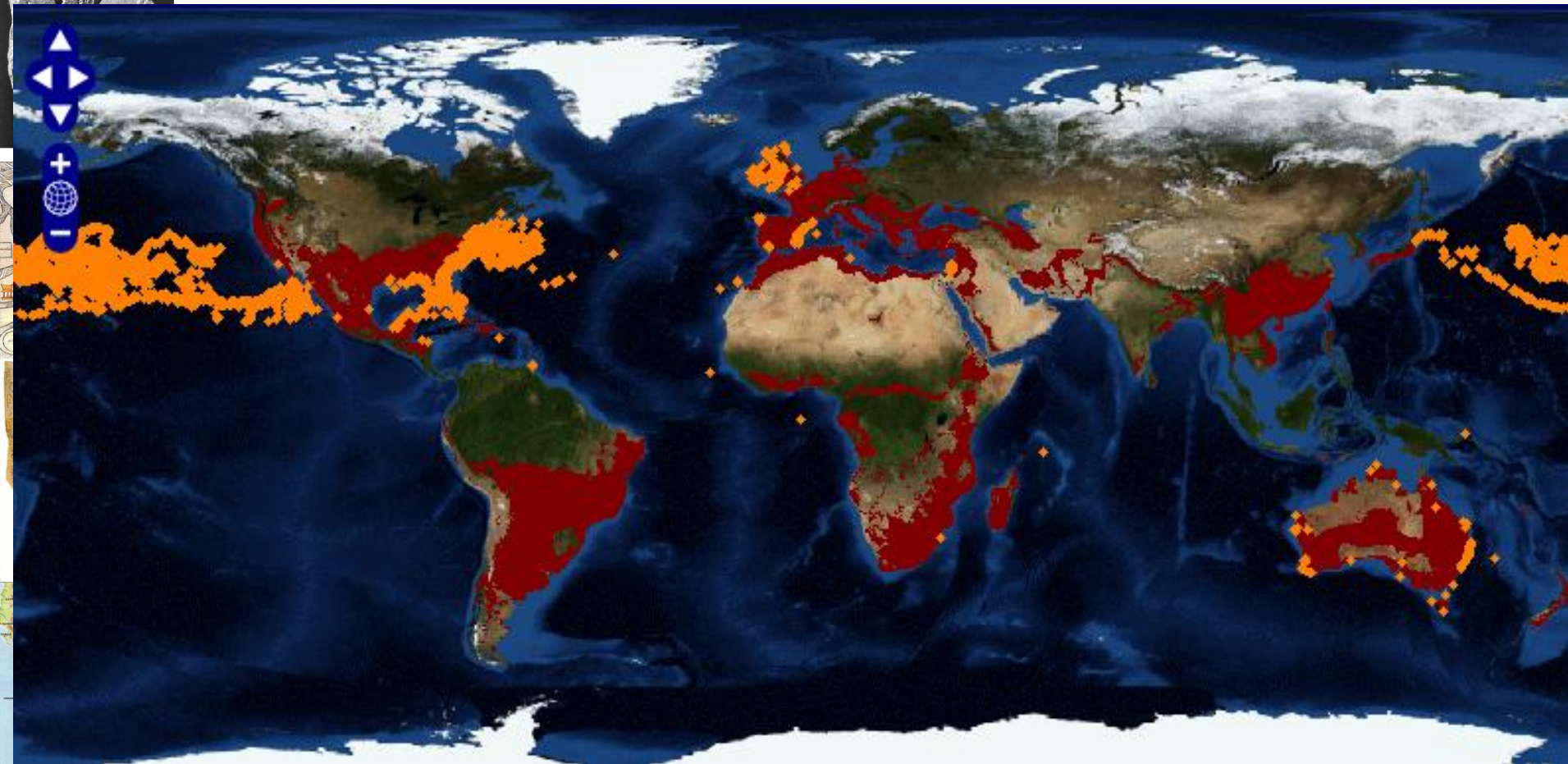
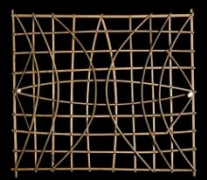
8/6/03 © 2003 United Feature Syndicate, Inc.



LifeMapper: *Tamarix chinensis*

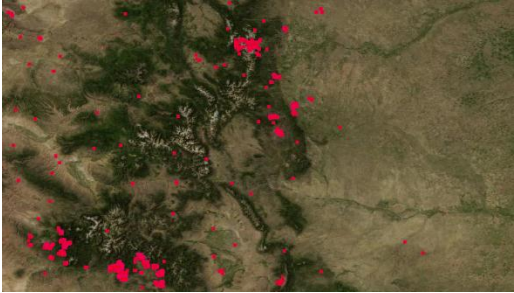


LifeMapper: Loggerhead Turtles



HSM Process

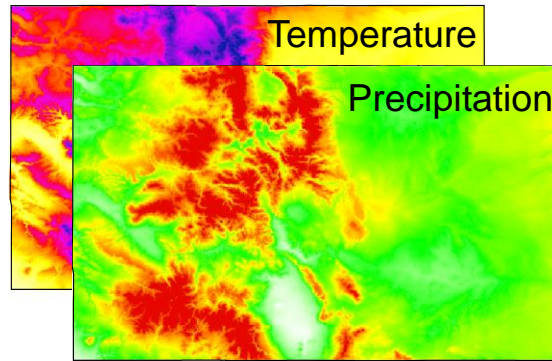
Occurrences (Sample Data)



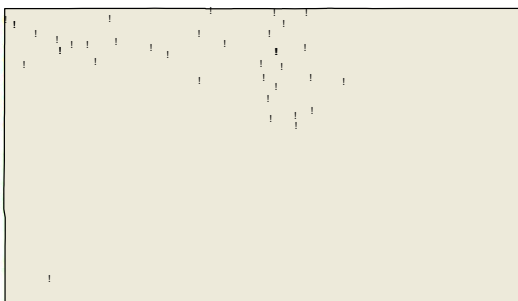
Spreadsheets

Lat	Lon	Temp	Precip
-105.504	40.35819	5.32	58.4
-107.472	40.498	6.31	47.6

Environmental Layers (Predictors)



Habitat Suitability Map

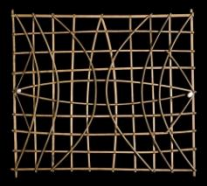


Map
Generation

Model Parameters

Variable	Param1	Param2
Annual Precip	-0.05	0.0
Annual Temp.	0.61	0.0

Modeling
Method

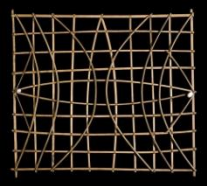


Input Data

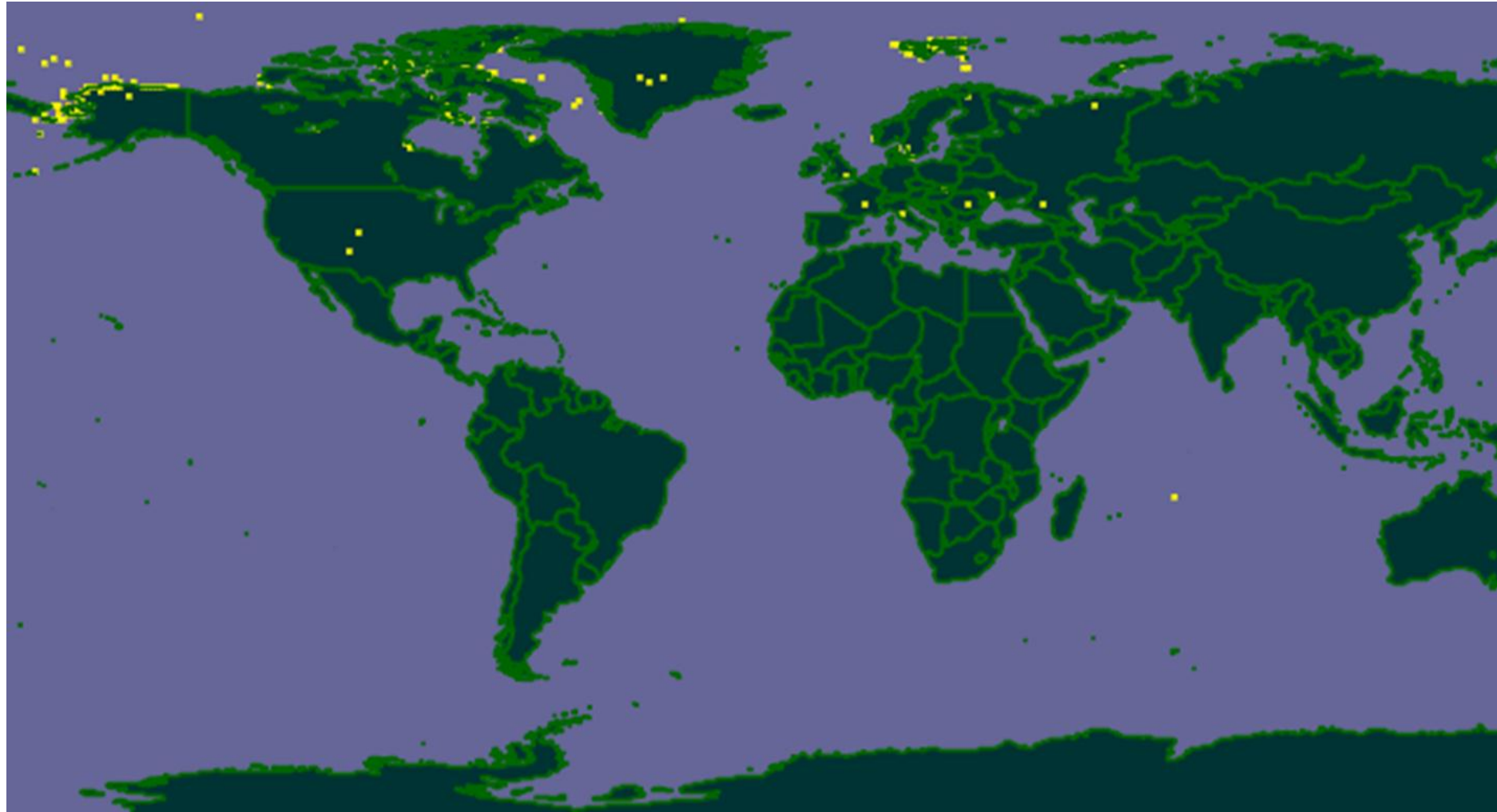


- Sample Data:
 - Typically points
 - Usually “Occurrences” or “Observations”
 - Often opportunistic, integrated data
 - Typically of unknown uncertainty
- Predictor layers
 - Environmental factors/characteristics
 - Typically rasters
 - Huge variety
 - Typically of unknown uncertainty

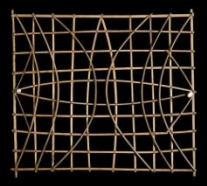




Occurrences of Polar Bears



From The Global Biodiversity Information Facility (www.gbif.org, 2011)



January 1st Dates



- If you put just a “year”, like 2011, into a relational database, the database will return:

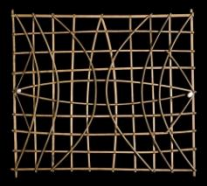
- Midnight, January 1st, of that year

- In other words:

- 2011 becomes:

- 2011-01-01 00:00:00.00



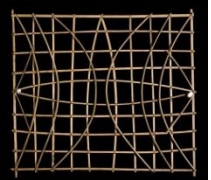


Predictor Layers

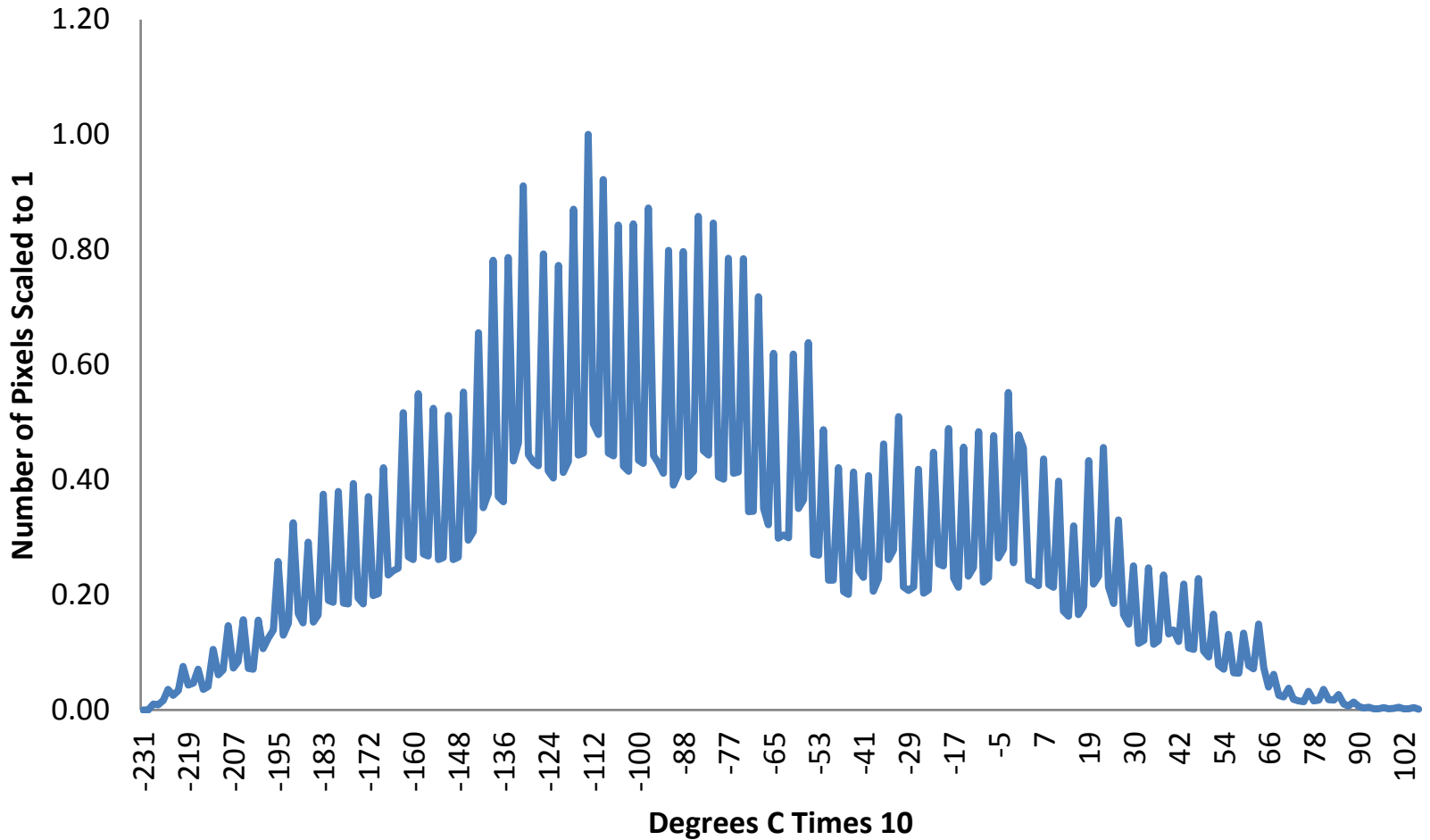


- Remotely sensed:
 - DEMs, Visual, IR, NIR, SST, NPP, Sea Height
- Integrated from multiple data sets:
 - Bathymetry
- Interpolated:
 - Temp, precipitation, wind
 - DO, Sub-surface temp, salinity, bottom type
- Processed from other layers:
 - Slope, aspect, distance to shore
- From other models:
 - Climate

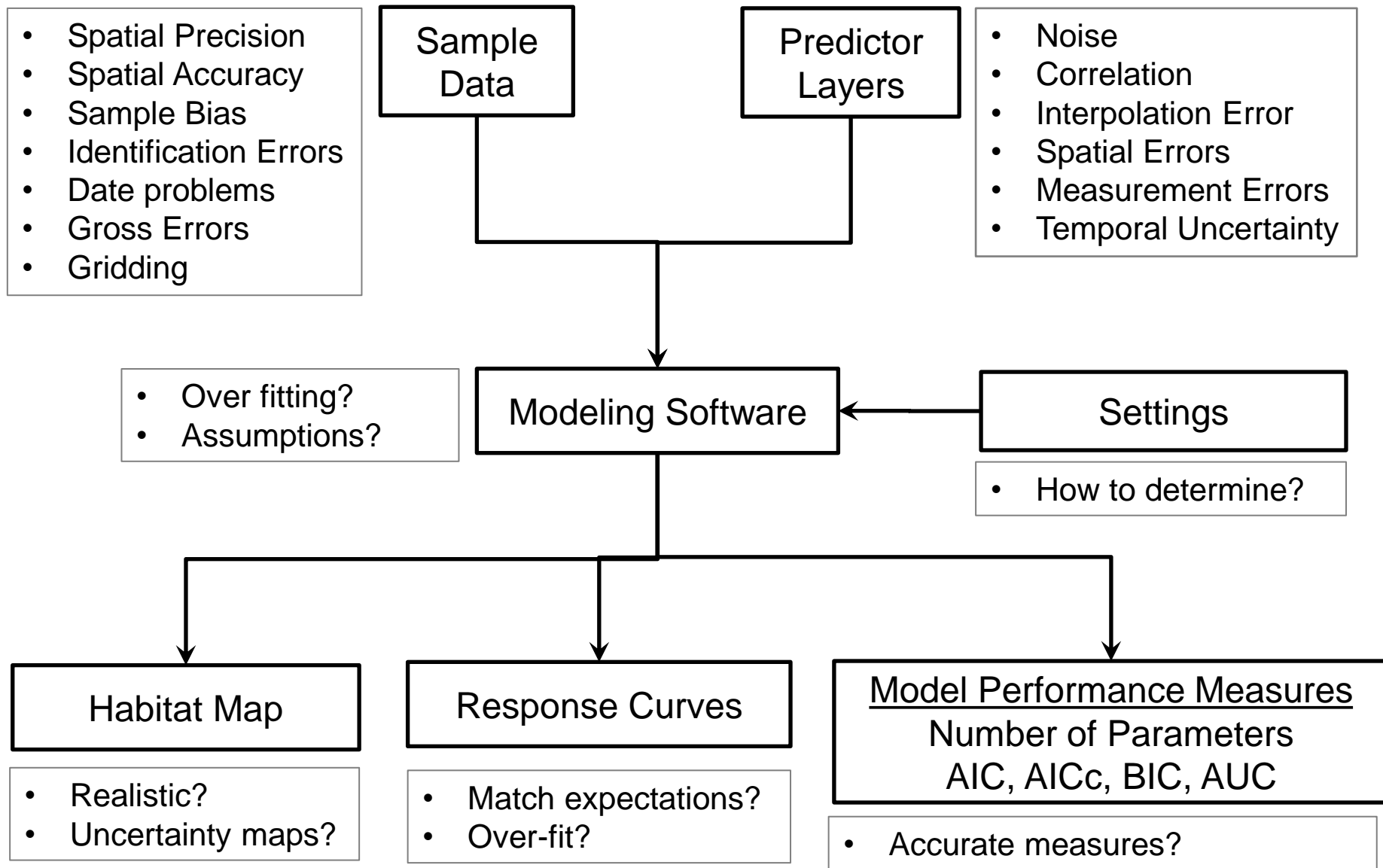
BioClim/WorldClim Data



Min Temp of Coldest Month

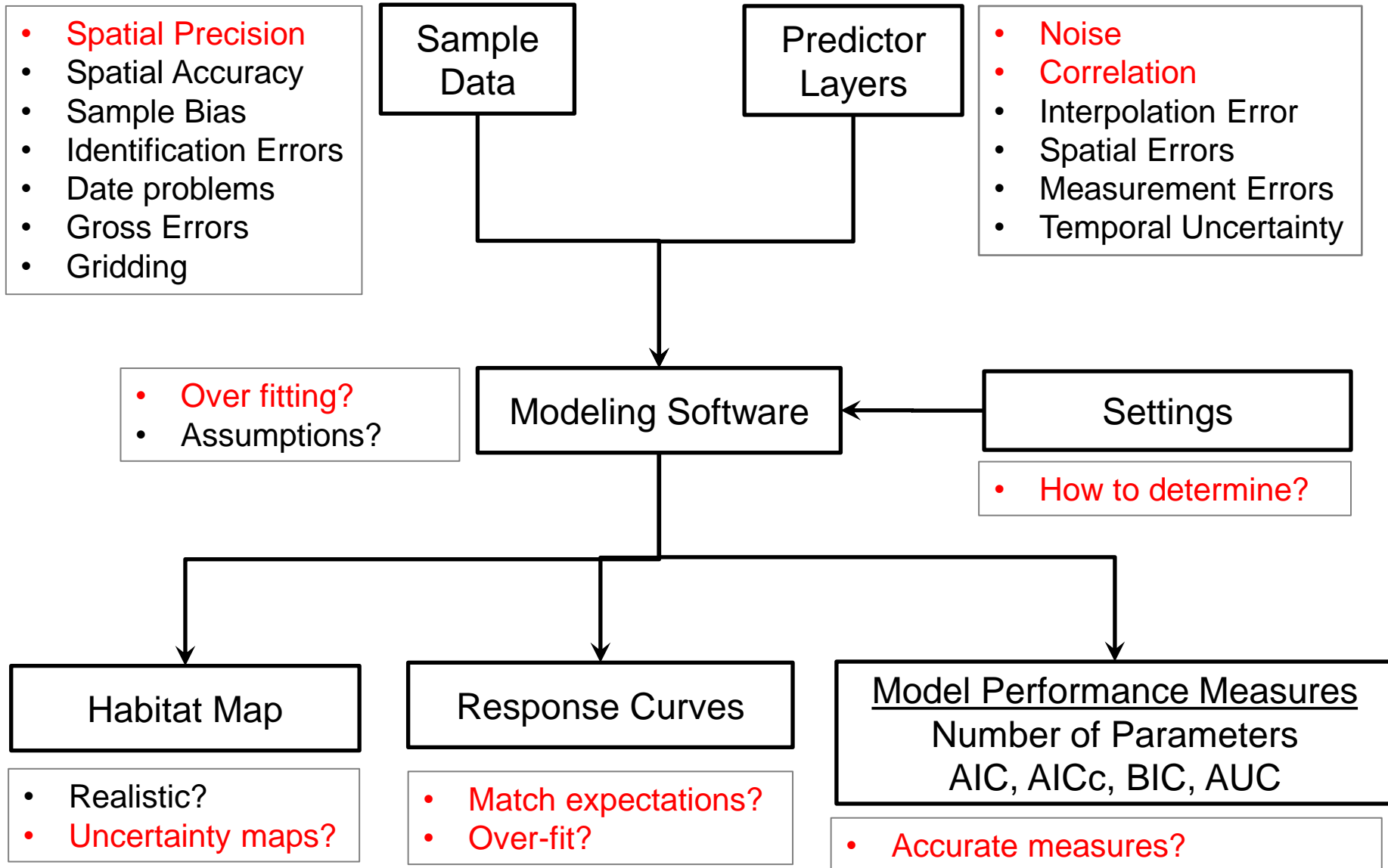


Road Map of Uncertainty

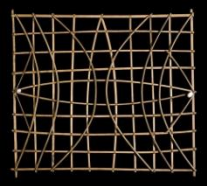


What is the best model?

Road Map of Uncertainty



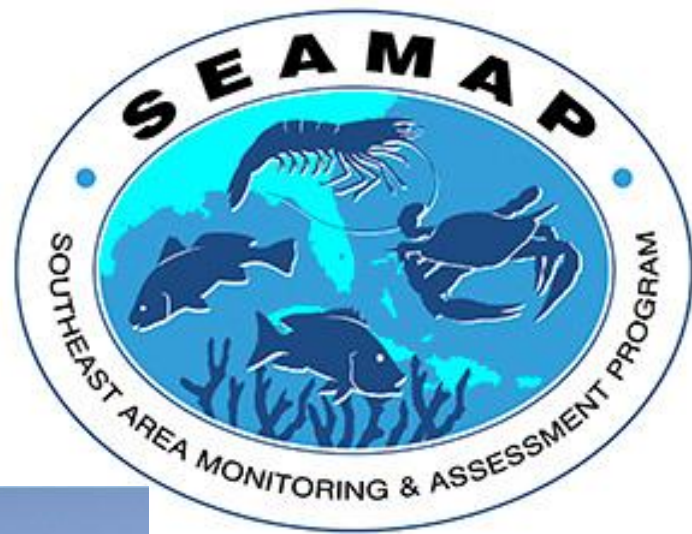
What is the best model?



SEAMAP



- Southeast Area Monitoring and Assessment Program (SEAMAP)
- Over 50 years of data
- Over 40,000 trawls

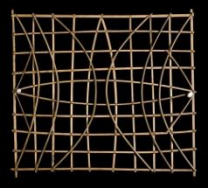


Red Snapper

- An important recreational and commercial species
- \$7 - 70 million/year

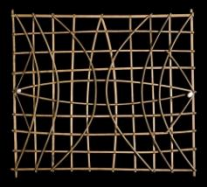


www.safmc.net



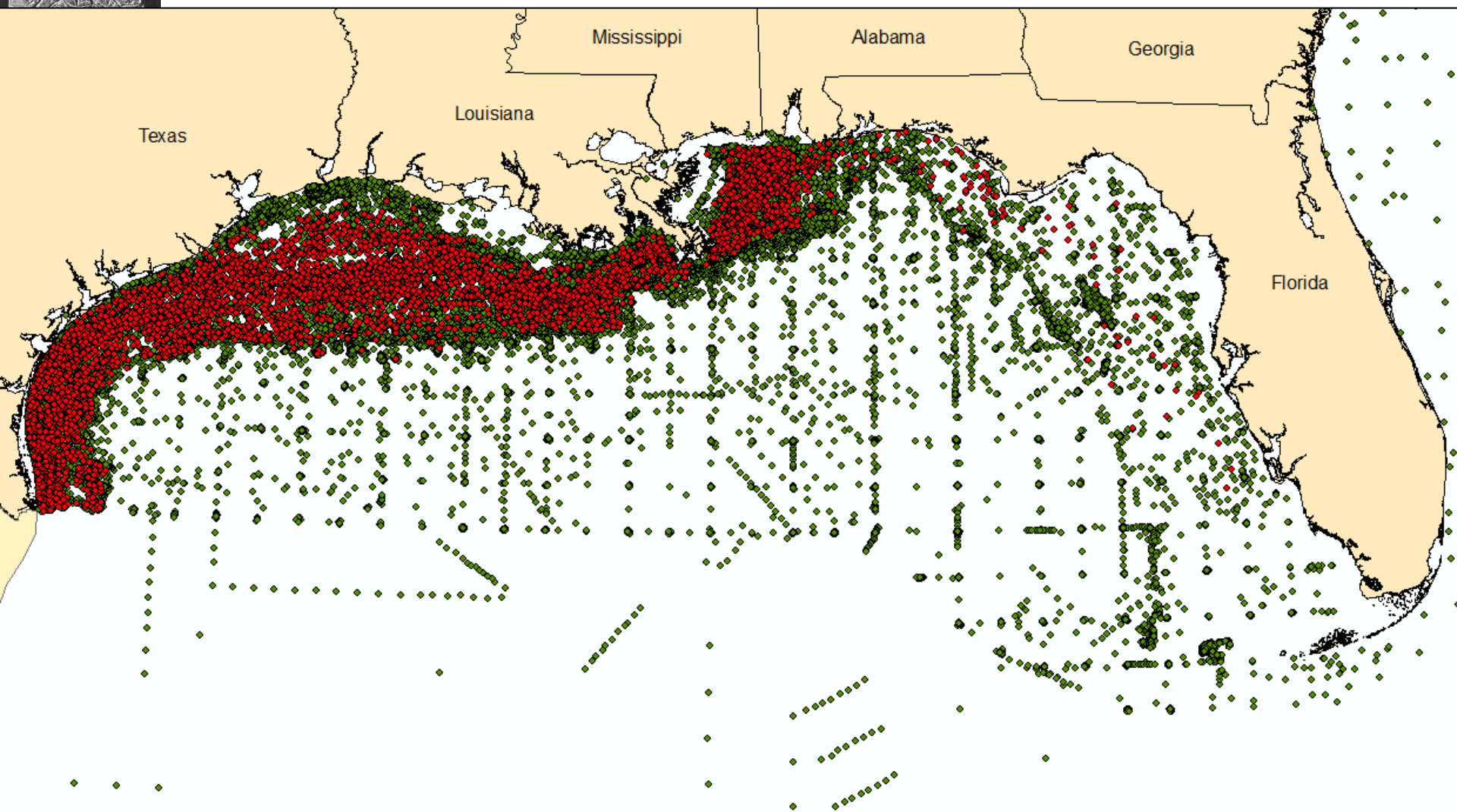
lickyourownbowl.wordpress.com

outdooralabama.com



● SEAMAP Trawls (>47,000 records)

● Red Snapper Occurrences (>6,000 records)



Sea Surface Temperature (SST)

NOAA

AVHRR Pathfinder Satellite

Spatial: 9km

Measured: <0.4 K

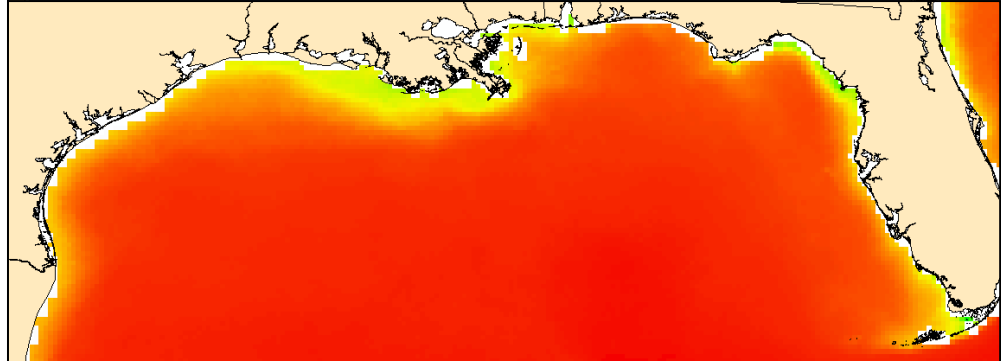


Net Primary Production (NPP)

Milligrams of Carbon per
Meter Squared per Day

OSU Ocean Productivity

Uncertainty?

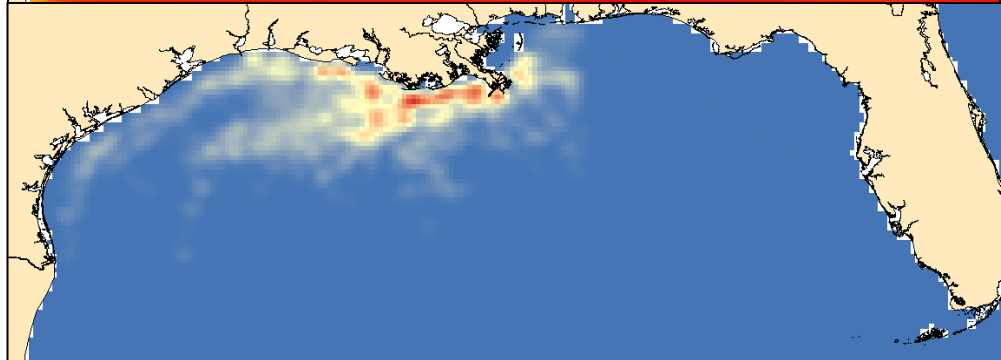


Density of

Platforms and Pipelines

Created from Bureau of Ocean Energy
Management (BOEM) Point Data Set

Uncertainty?



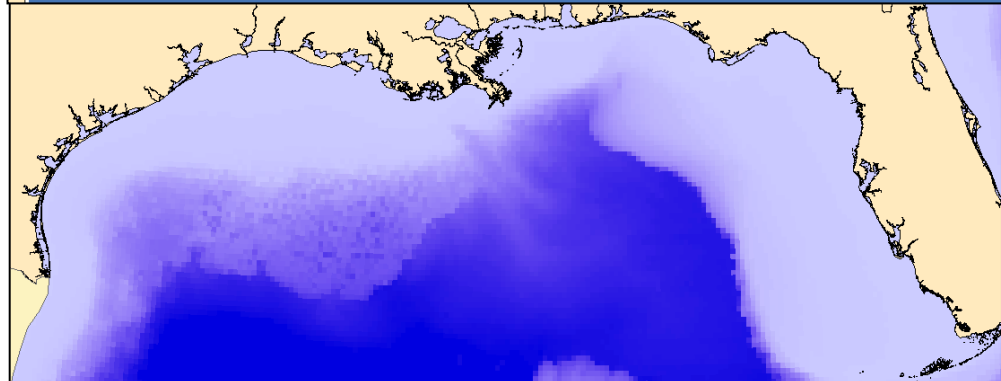
Bathymetry

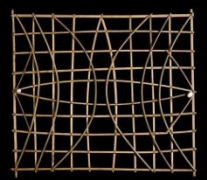
Resampled from 90m

NOAA Coastal Inundation Dataset

And others

Uncertainty < 9km





Predictor Resolution



- Should model at the lowest resolution of the predictor layers or lower



- Lower resolution:

- Can reduce problems from sample data uncertainty



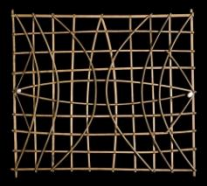
- In some cases can combine sample data into measures such as CPUE, density, abundance



- Reduces detail of final model



- For these tests: 9km



Sample Data

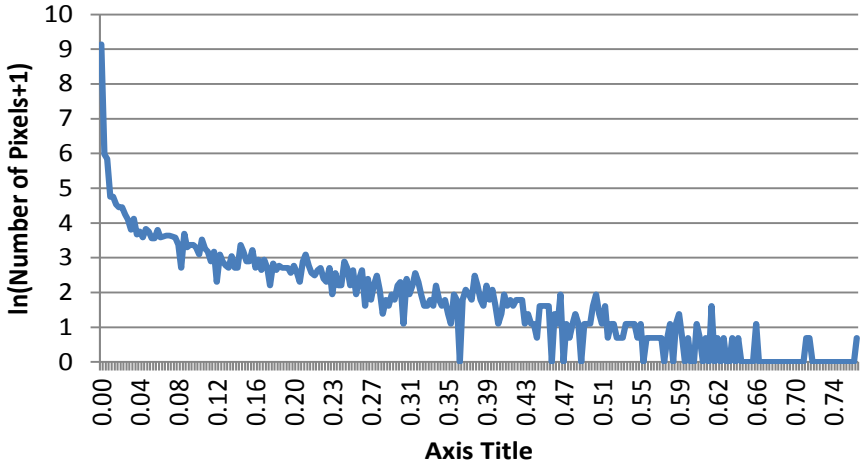


- Spatial Precision
 - Standard Deviation of about 2 km
- Spatial Accuracy:
 - < 1km
- Sampling Bias:
 - Some areas more heavily sampled
- Identification Errors:
 - Unknown but red snapper are pretty easy to identify
- Date problems:
 - Not a temporal model
- Gridding:
 - Not at 9 kilometers

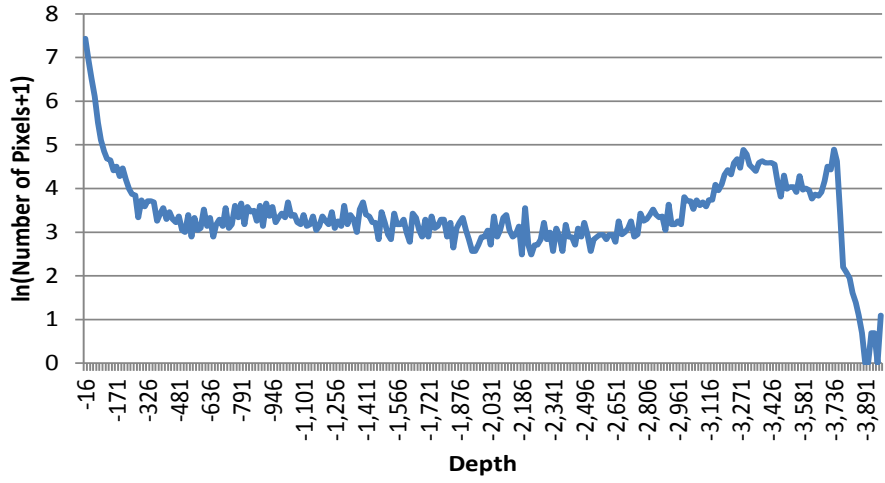


Histograms of Predictor Layers

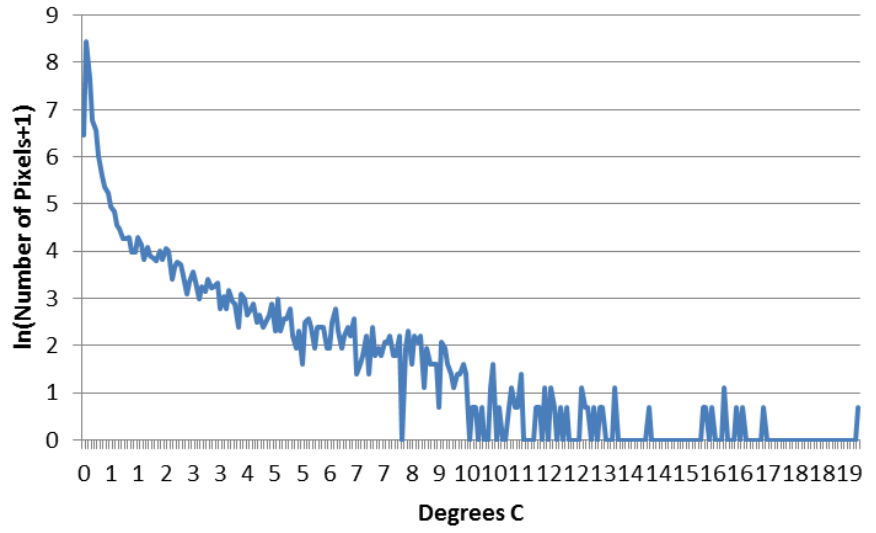
Density of Infrastructure



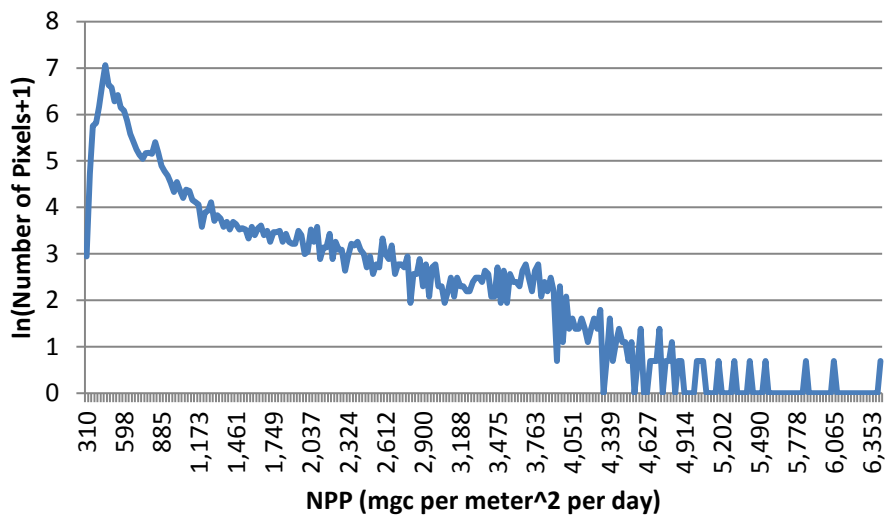
Bathymetry Histogram

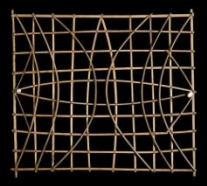


Sea Surface Temperature



Net Primary Productivity Histogram

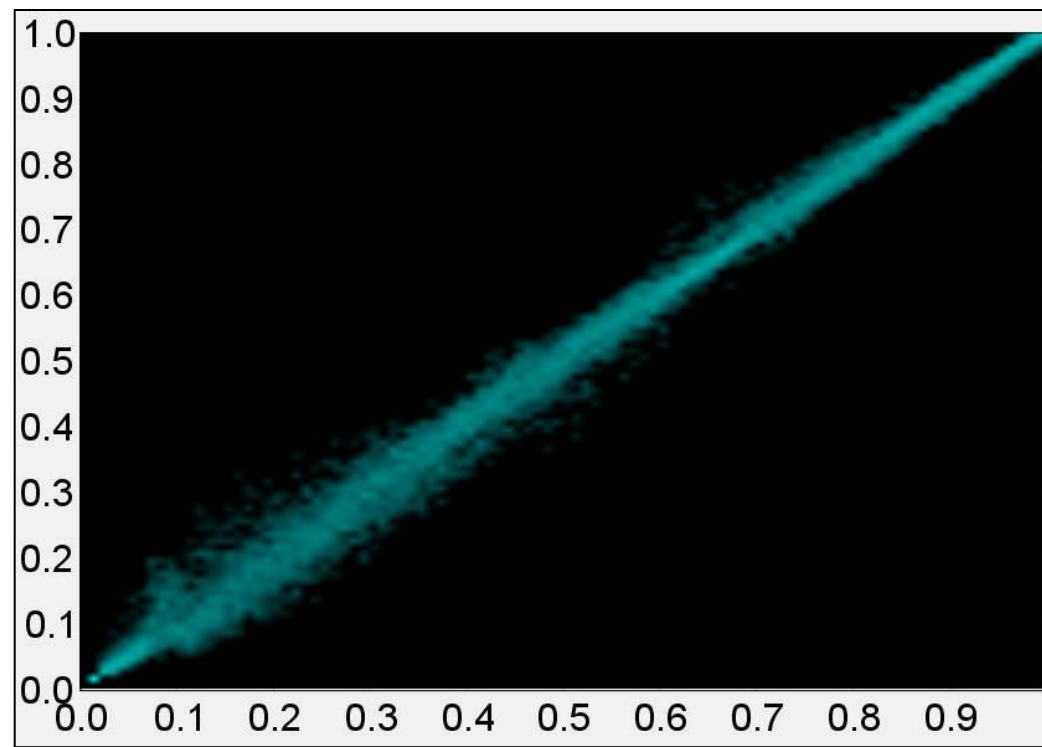
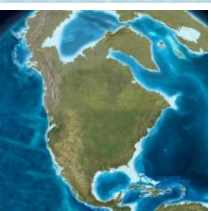


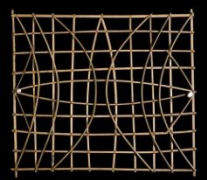


Predictor Layer Uncertainty



- Noise: Minimal
- Correlation: Eliminate NPP
- Spatial Errors: Unknown, < 9km?
- Measurement errors: Minimal to Unknown



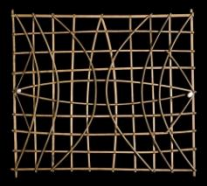
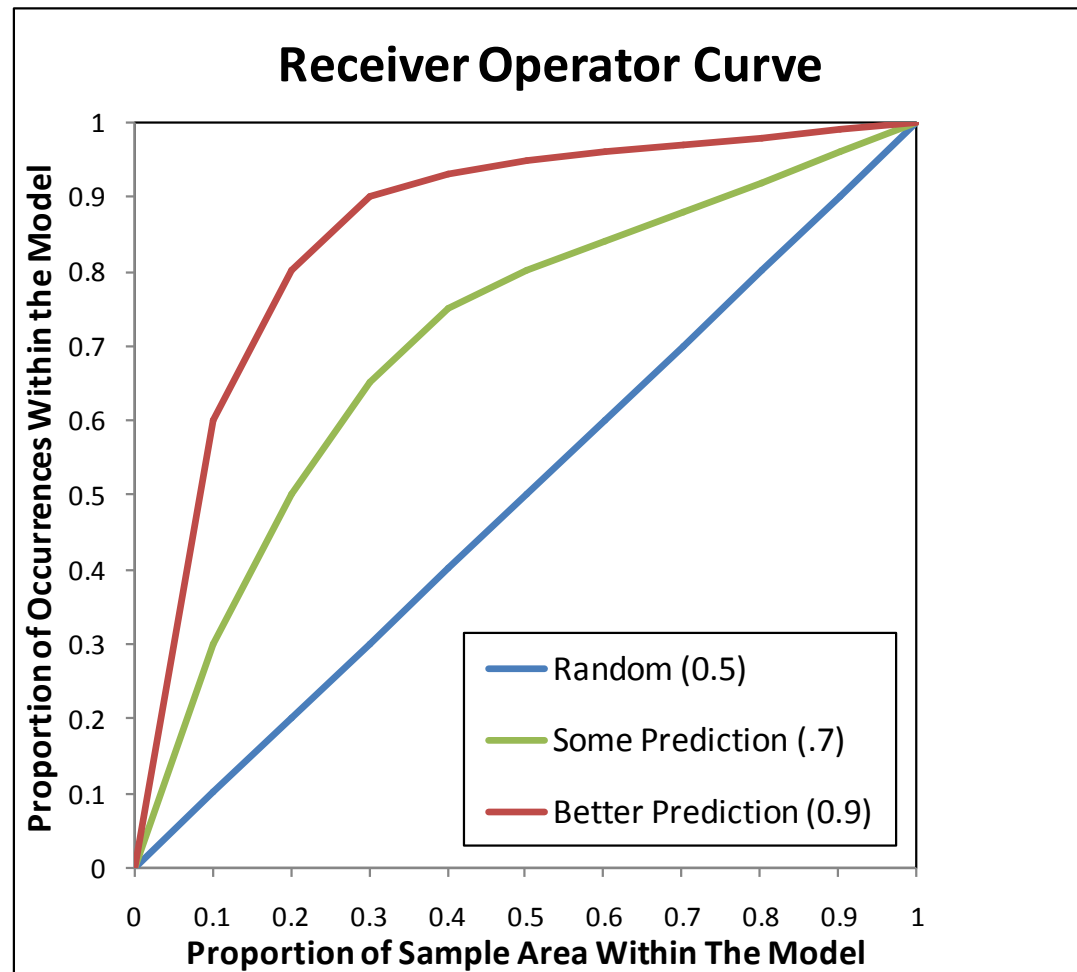


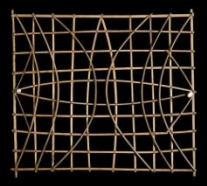
Modeling Software: Maxent

- Popular, relatively easy to use
- Only requires presence points
- **Known for over-fitting**
- Performs “piece-wise” regression
- Assumes:
 - Predictors are error free
 - Independence of errors in response
 - Lack of correlation in predictors
 - Constant variance
 - Random data collection over sample area

Area Under the Curve (AUC)

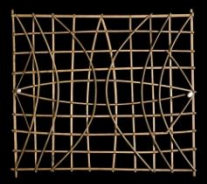
- Area under a Receiver Operator Curve (ROC)
- Popular for HSM
- Encourages over-fitting





Akaike Information Criterion (AIC)

- Balance between complexity:
 - Over fitting or modeling the residuals (errors)
 - Lots of parameters
- And bias
 - Under fitting or the model is missing part of the phenomenon we are trying to model
 - Too few parameters
- Smaller is “better”
- Must be used with the same samples and predictors between models

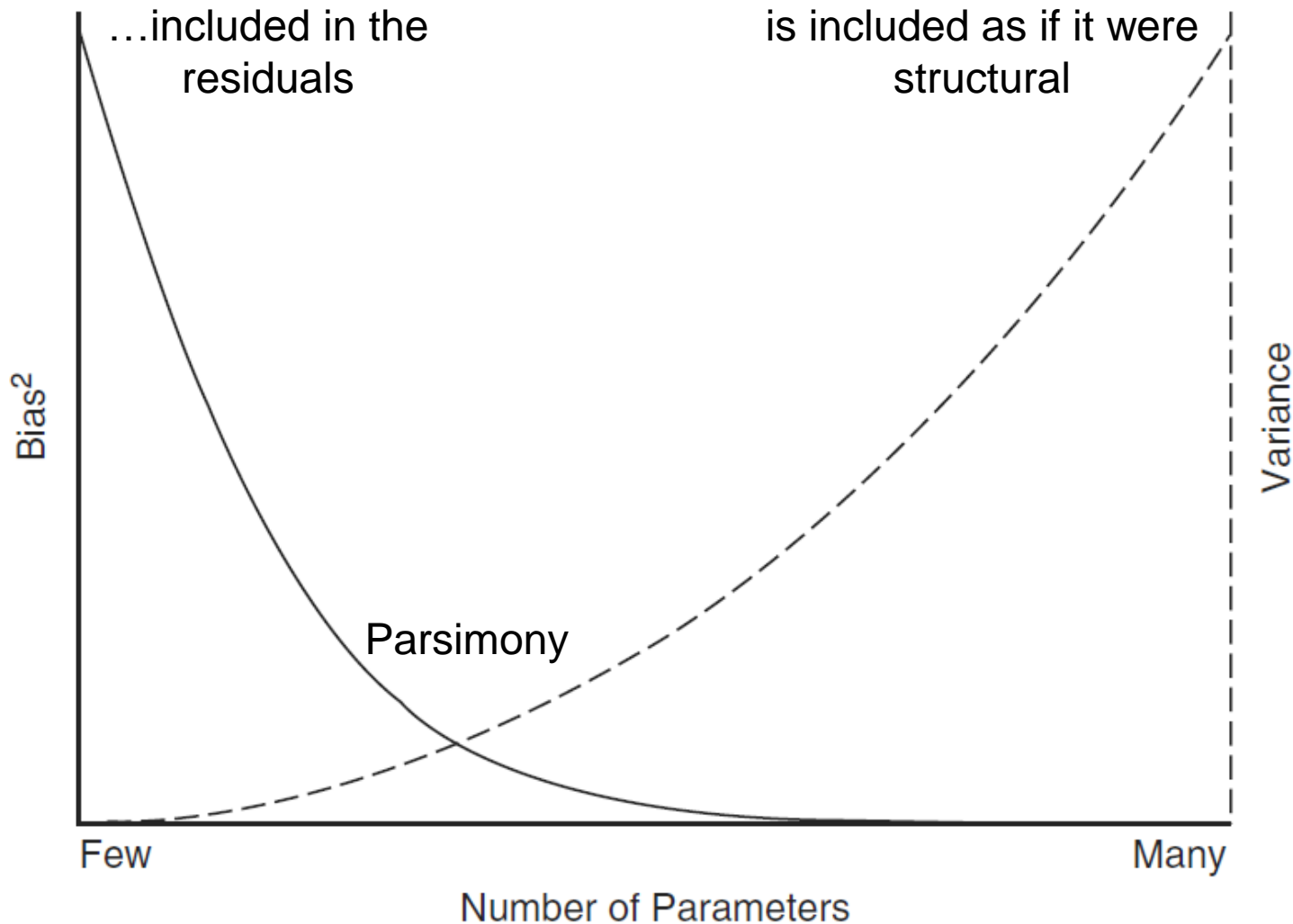


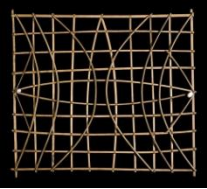
Parsimony



Under fitting
model structure
...included in the
residuals

Over fitting
residual variation
is included as if it were
structural



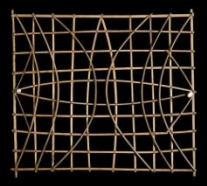


Experiments



- Different regularization multipliers
 - Select regularization for other tests
- Try different numbers of samples
 - Determine the number of samples to use
- “Jiggle” the sample data spatially
 - Introduce different amounts of spatial error



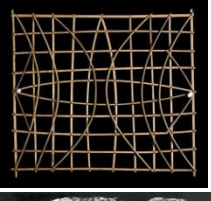


BlueSpray



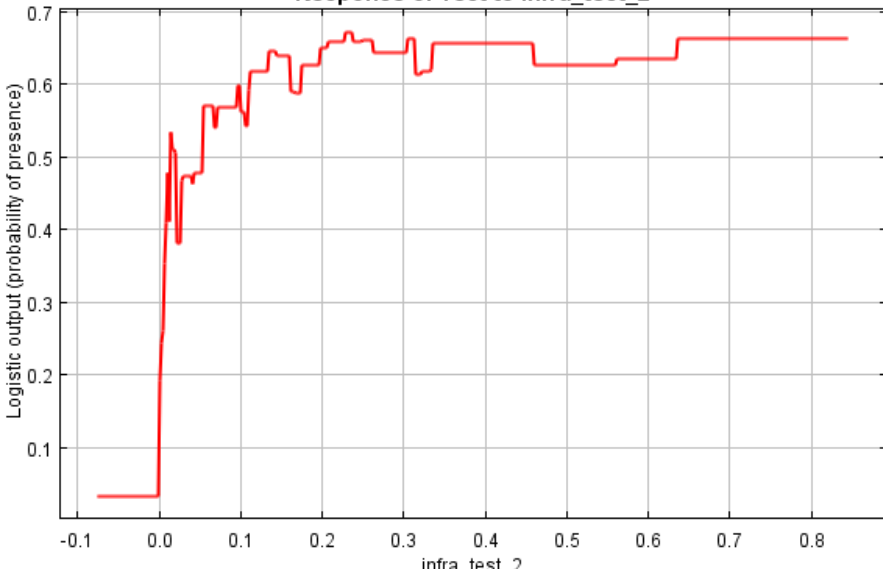
- GIS Application for Natural Resources and Environmental Research
 - Specifically HSM
- Available at www.schoonerturtles.com
- Free for beta testers in environmental research and conservation
- Contact Jim at: jimg@schoonerturtles.com for the passcode



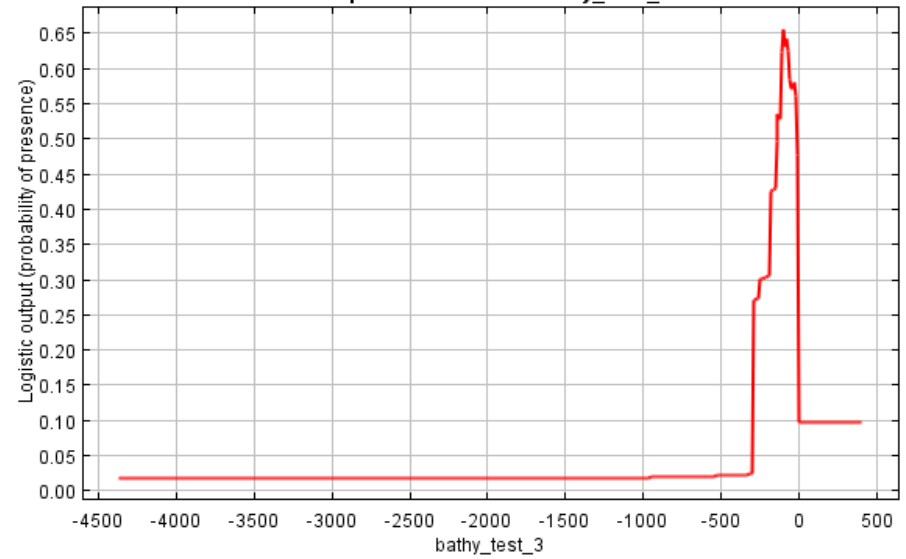


Regularization=0.1

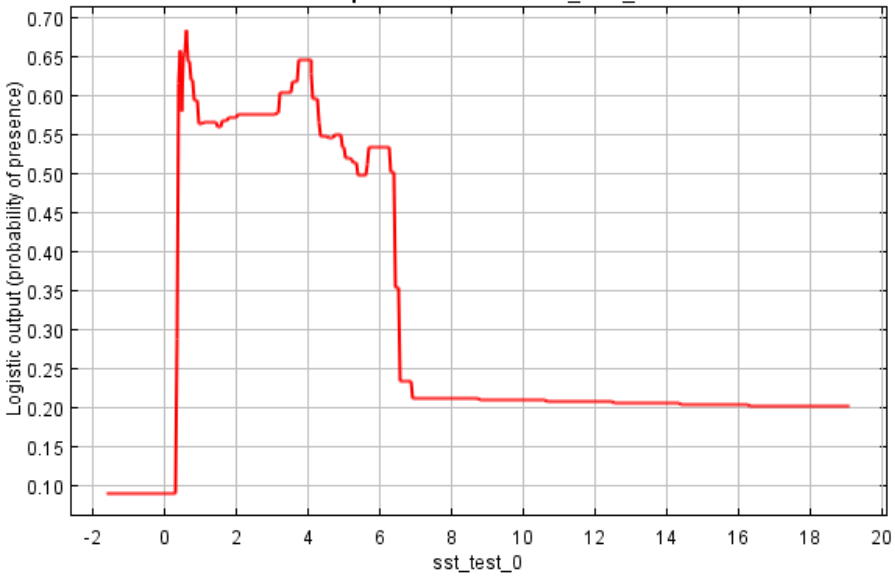
Response of Test to infra_test_2



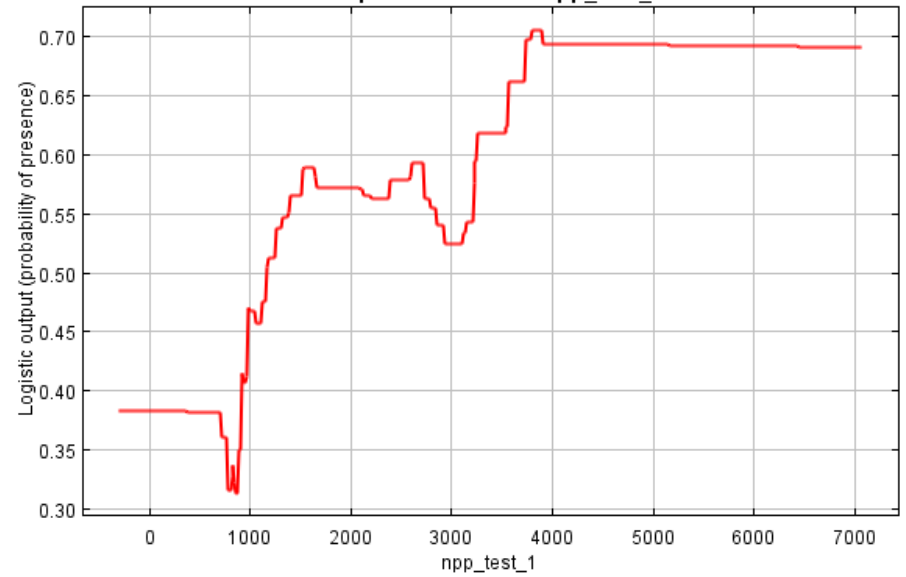
Response of Test to bathy_test_3

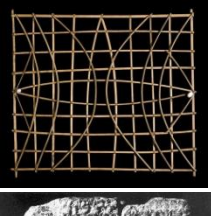


Response of Test to sst_test_0



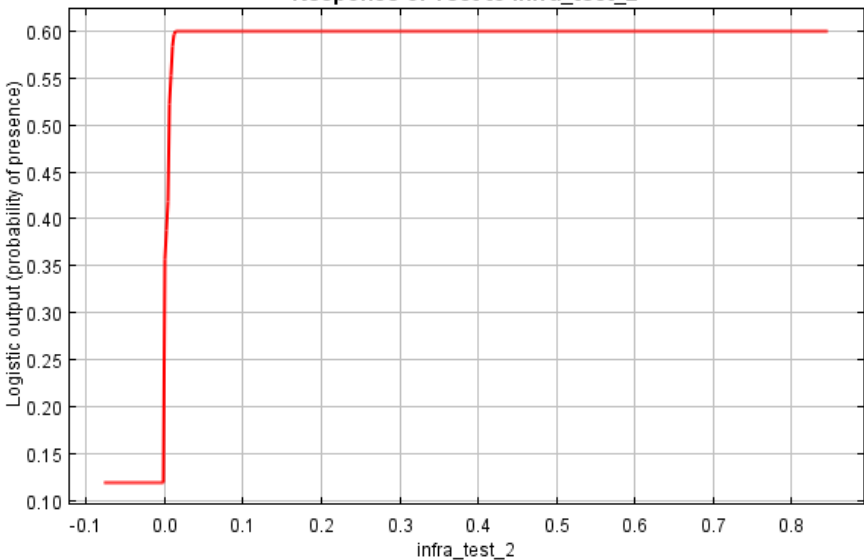
Response of Test to npp_test_1



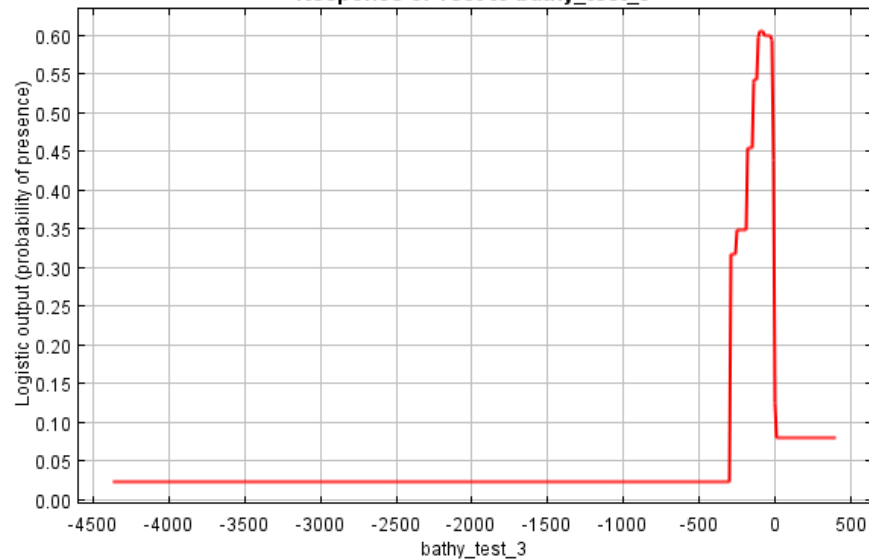


Regularization=1.2

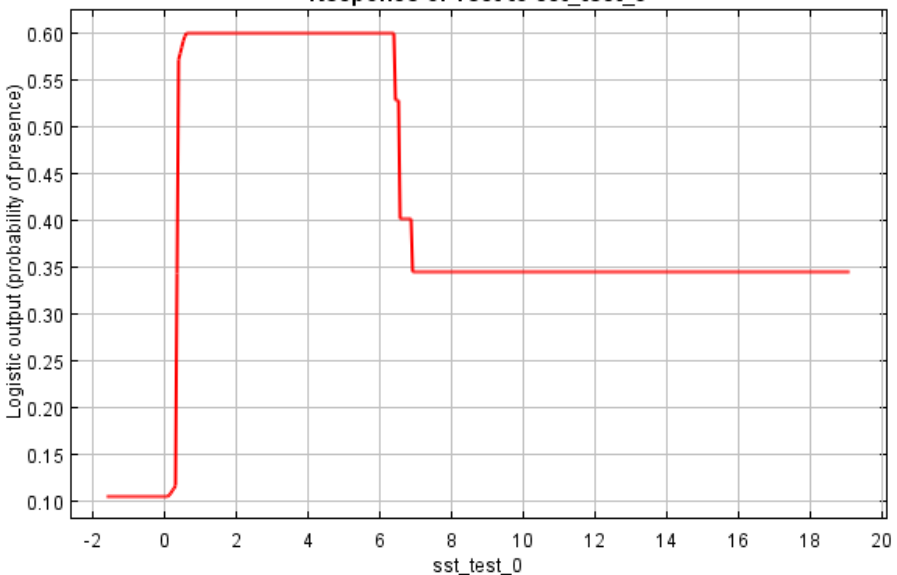
Response of Test to infra_test_2



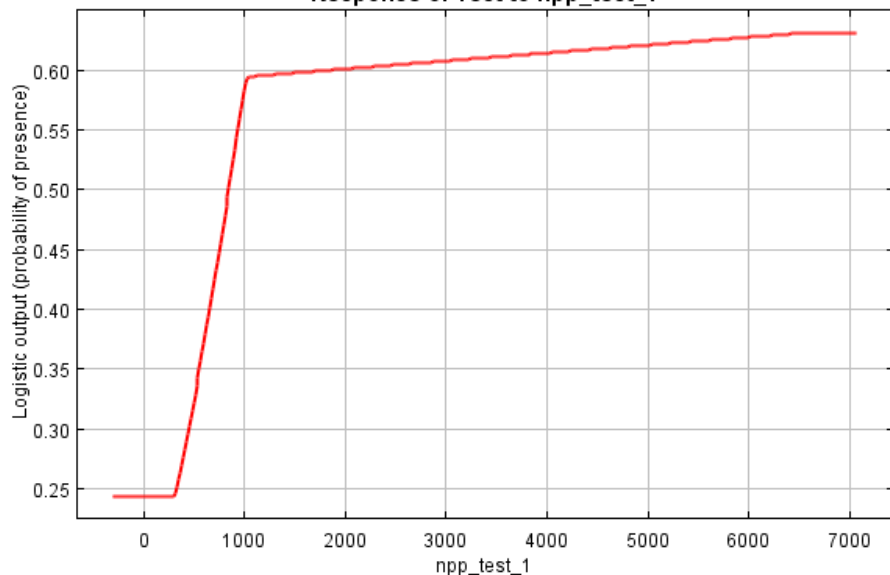
Response of Test to bathy_test_3

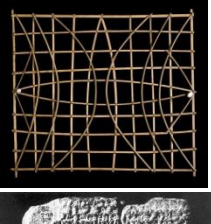


Response of Test to sst_test_0



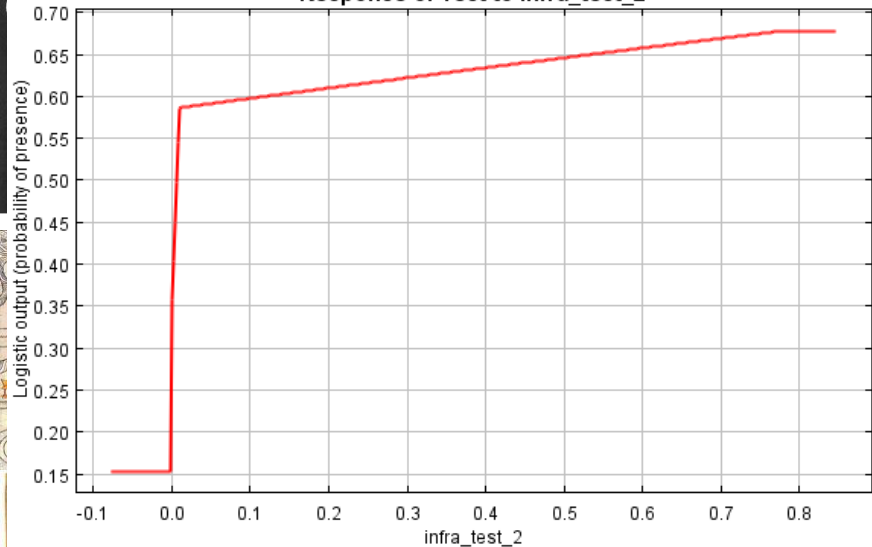
Response of Test to npp_test_1



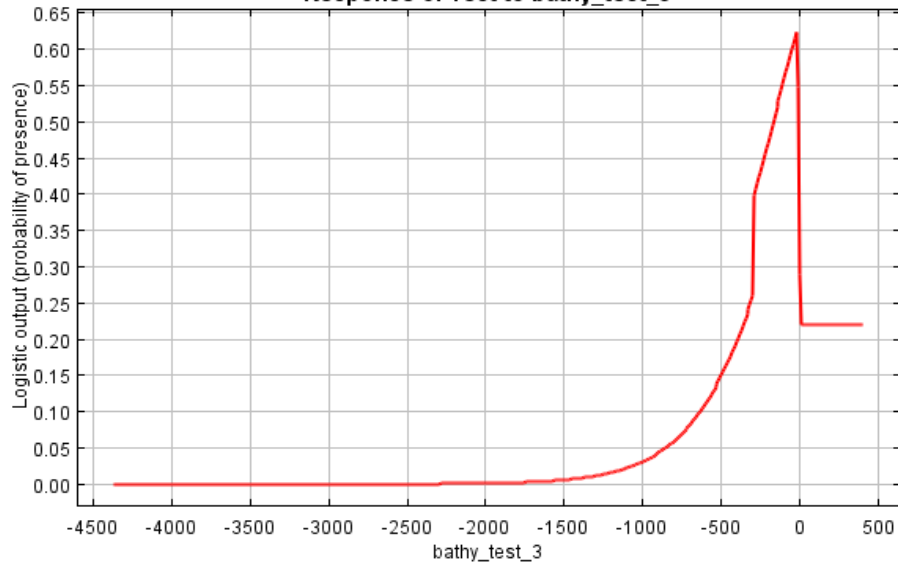


Regularization=10

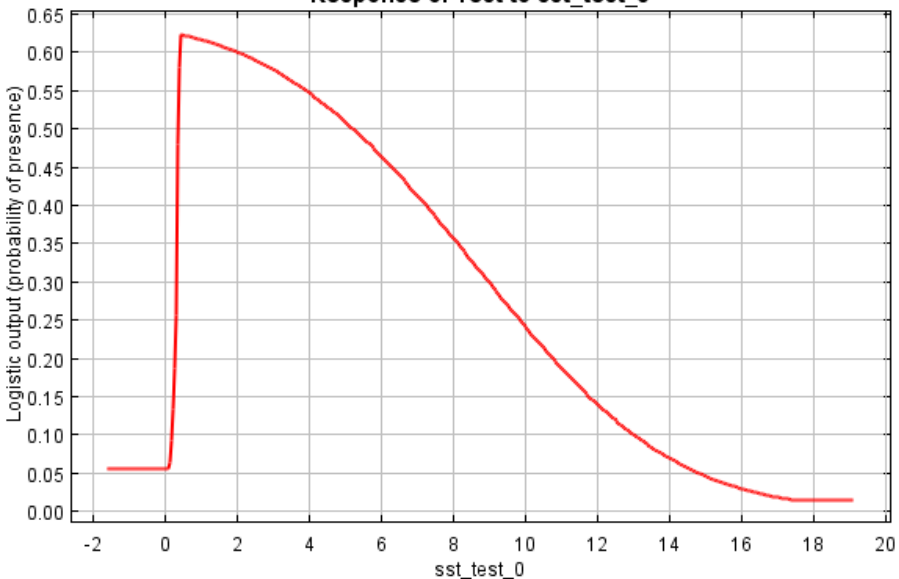
Response of Test to infra_test_2



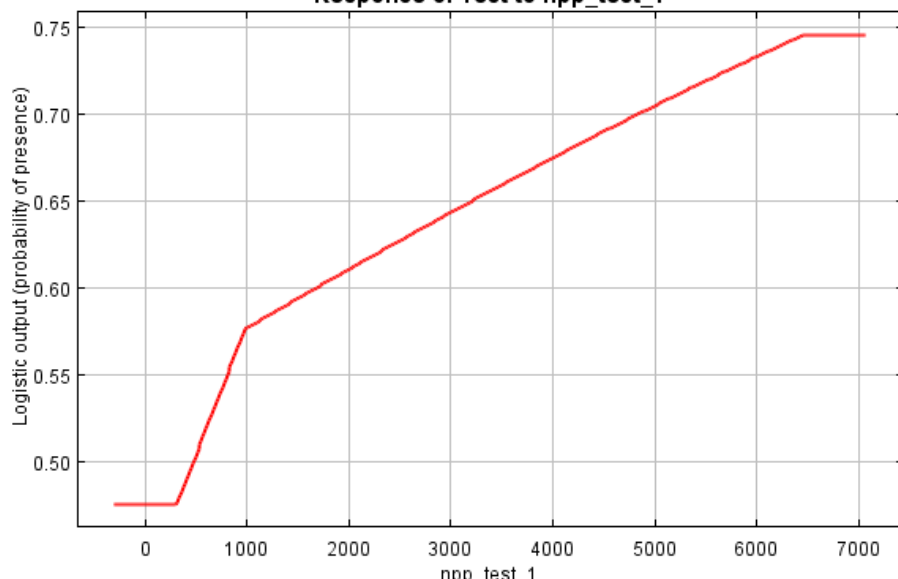
Response of Test to bathy_test_3

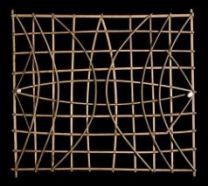


Response of Test to sst_test_0



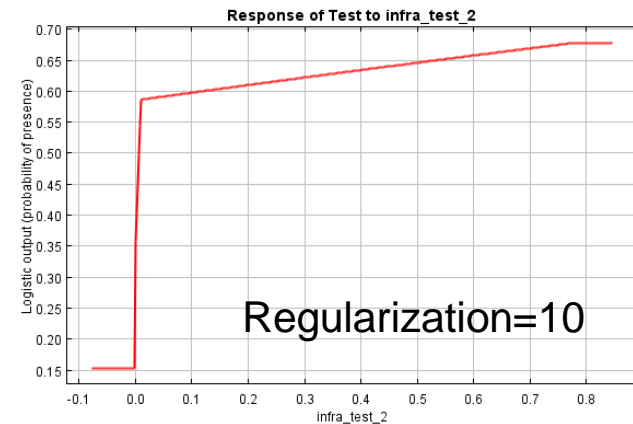
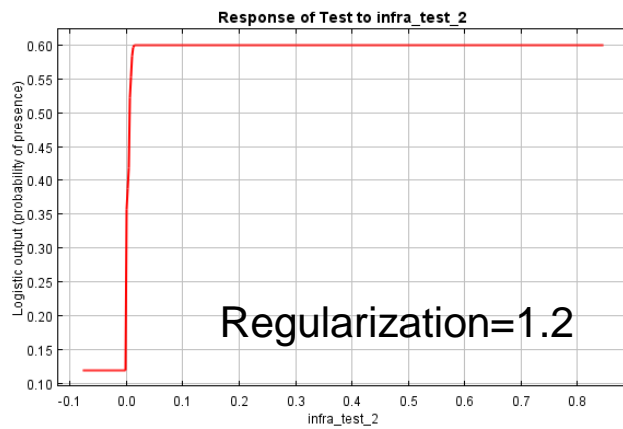
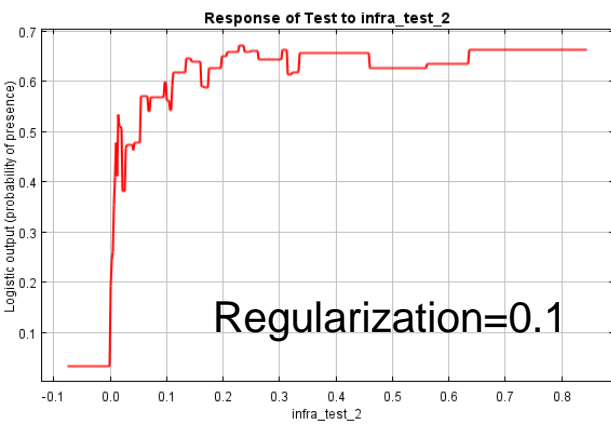
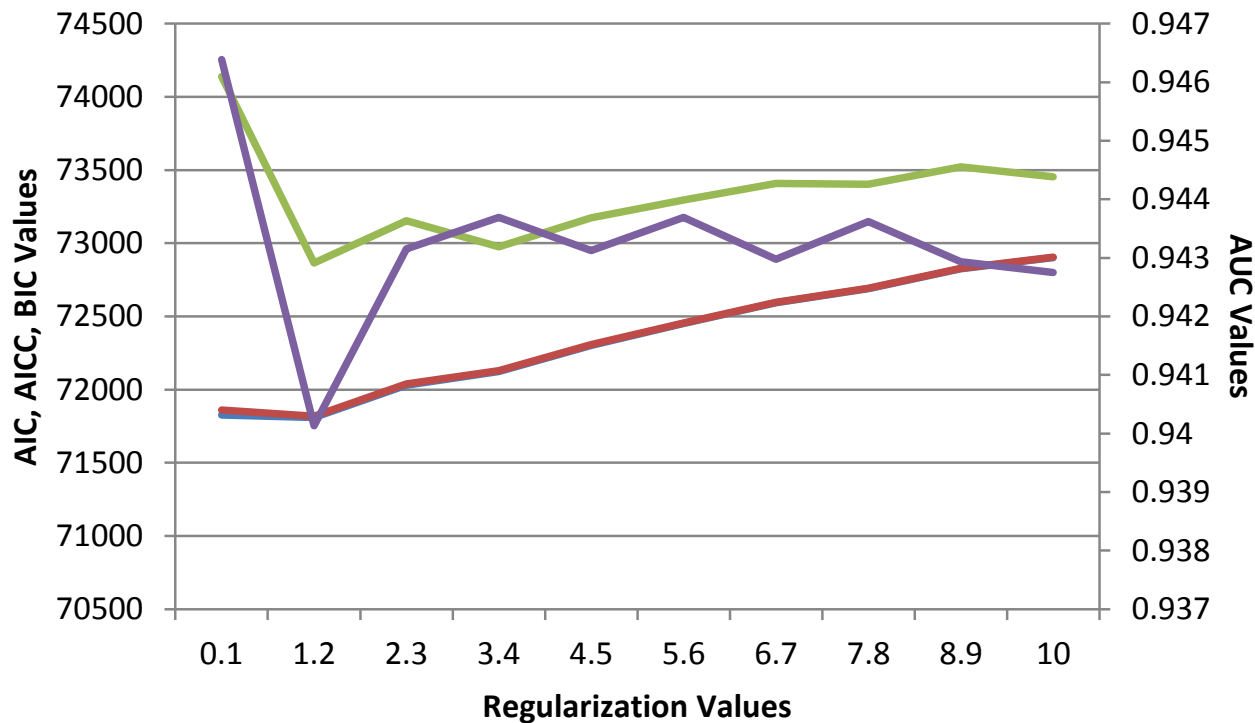
Response of Test to npp_test_1

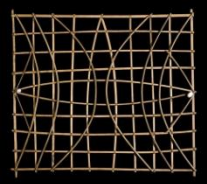




Regularization Results

— AIC — AICC — BIC — AUC





10%
Sample
Points



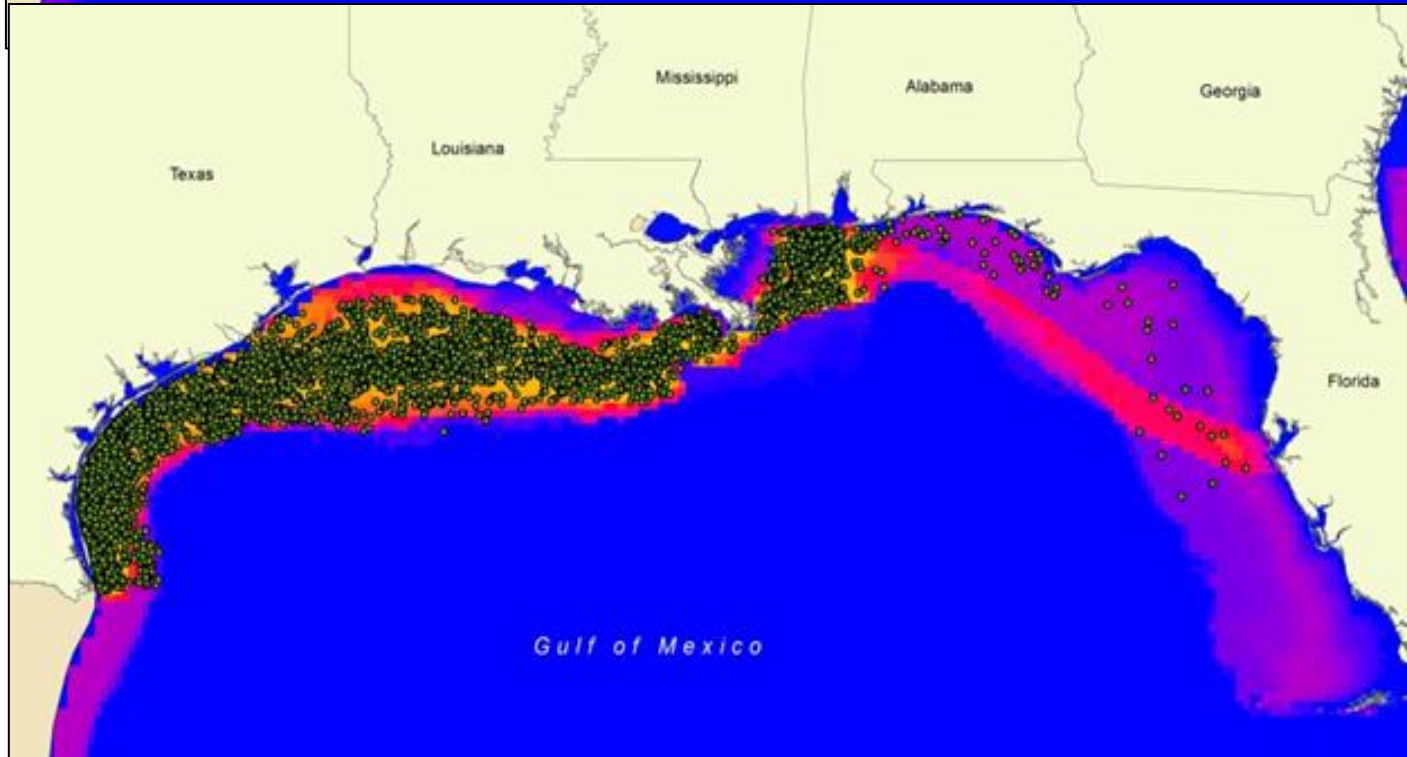
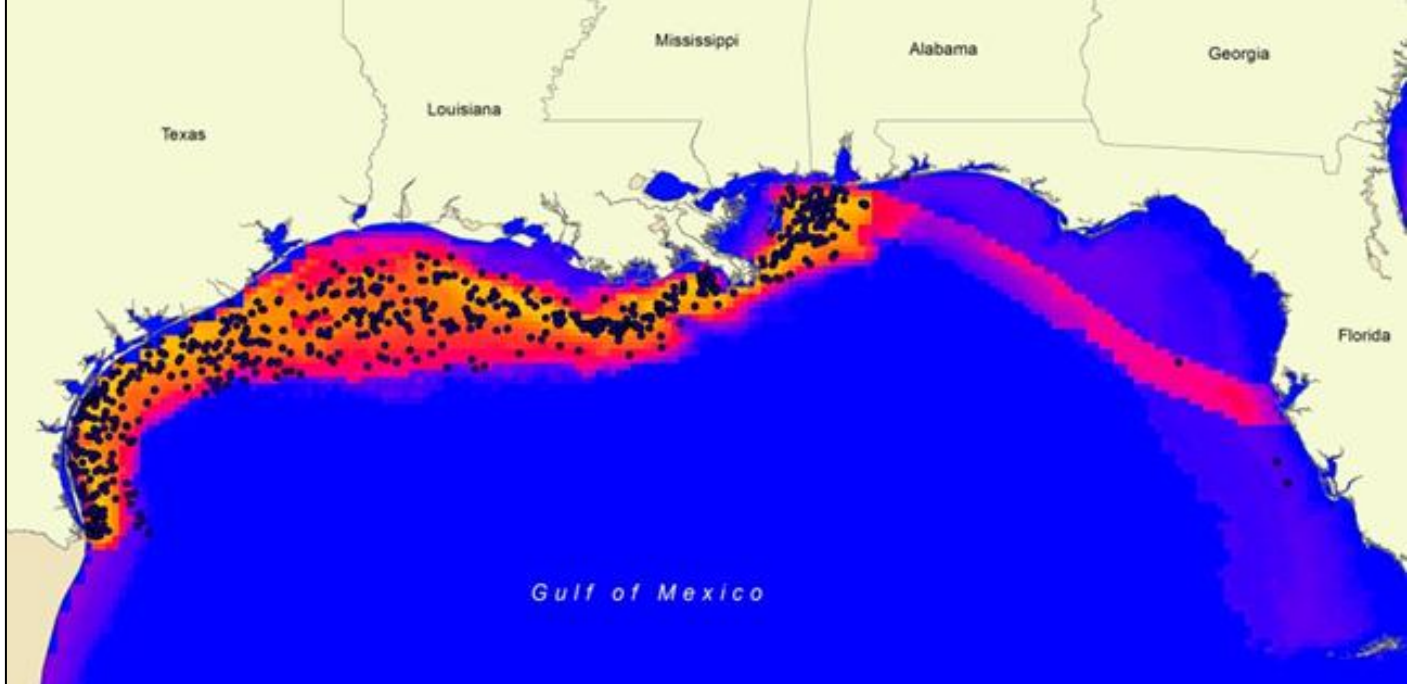
Optimal
Habitat

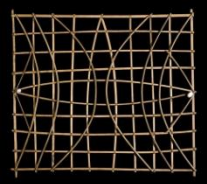


Poor
Habitat



100%
Sample
Points





10%
Sample
Points



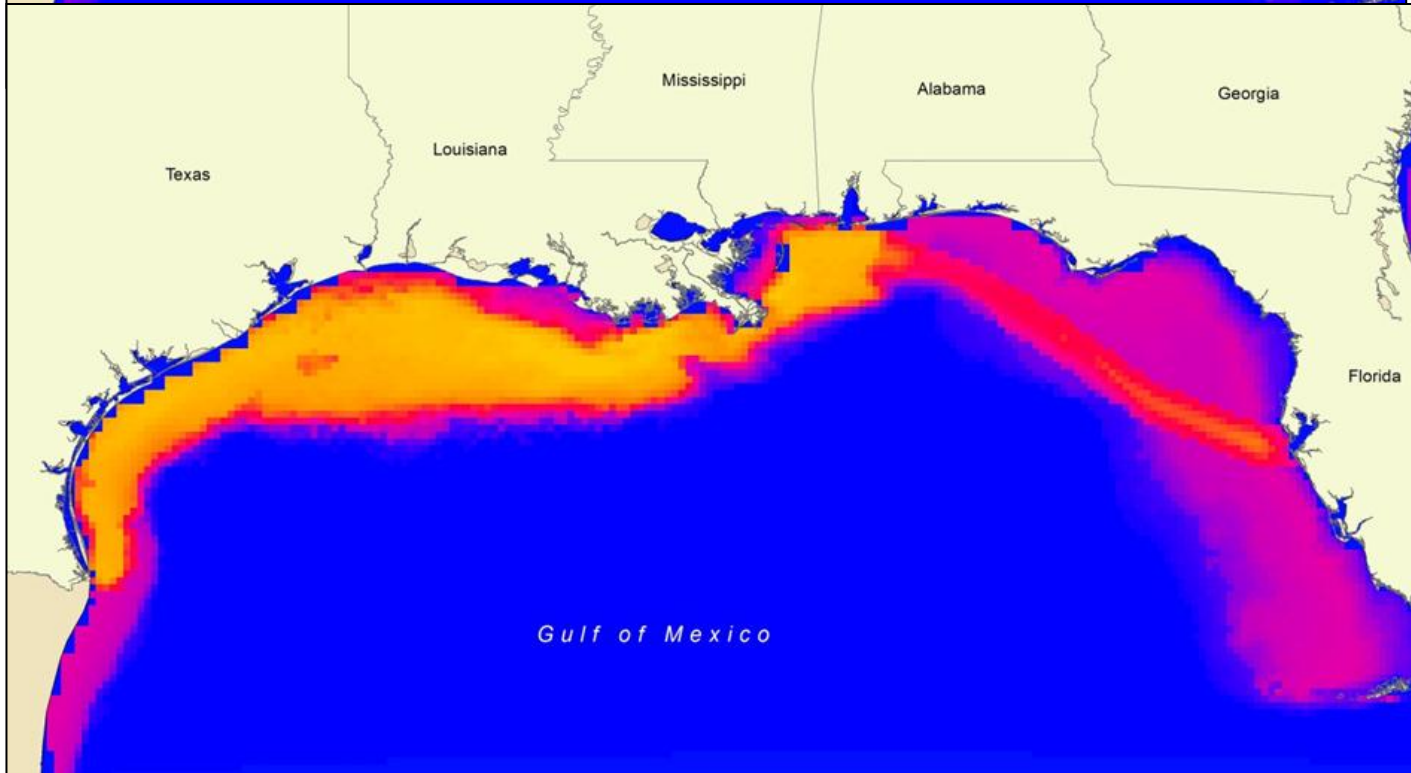
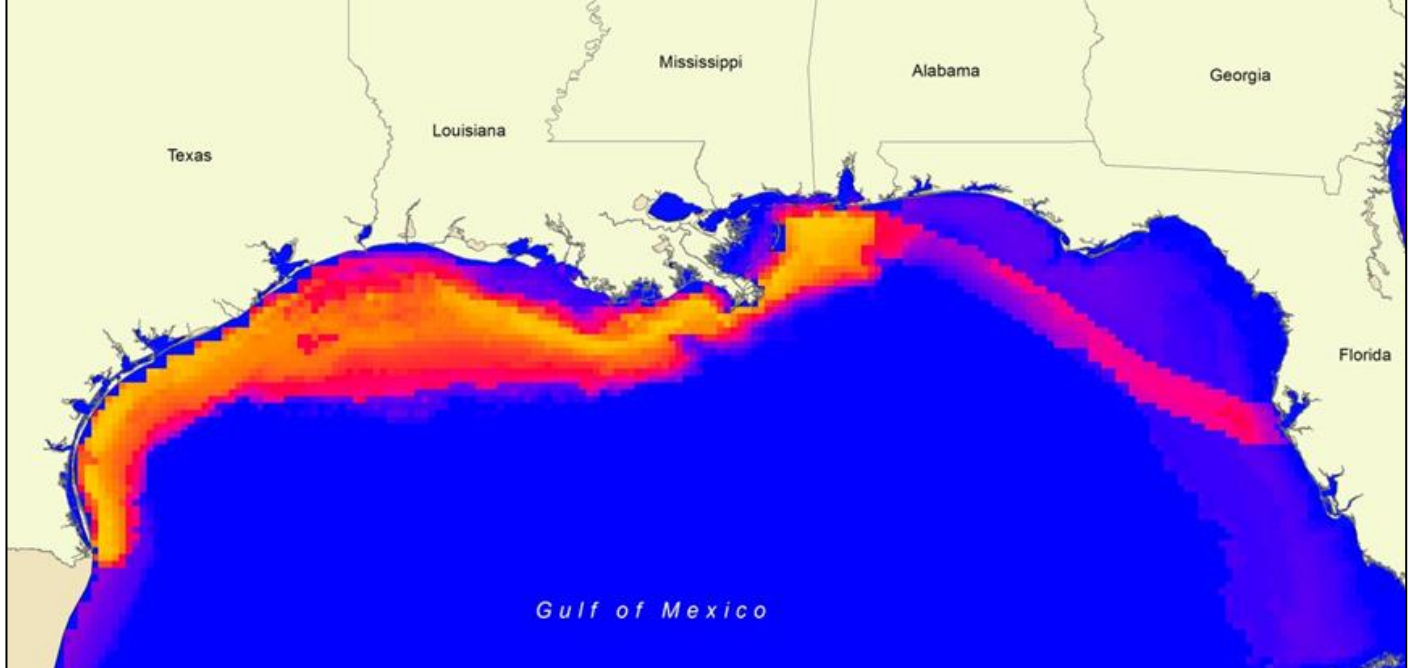
Optimal
Habitat

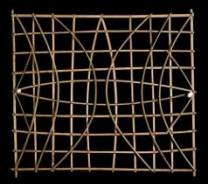


Poor
Habitat

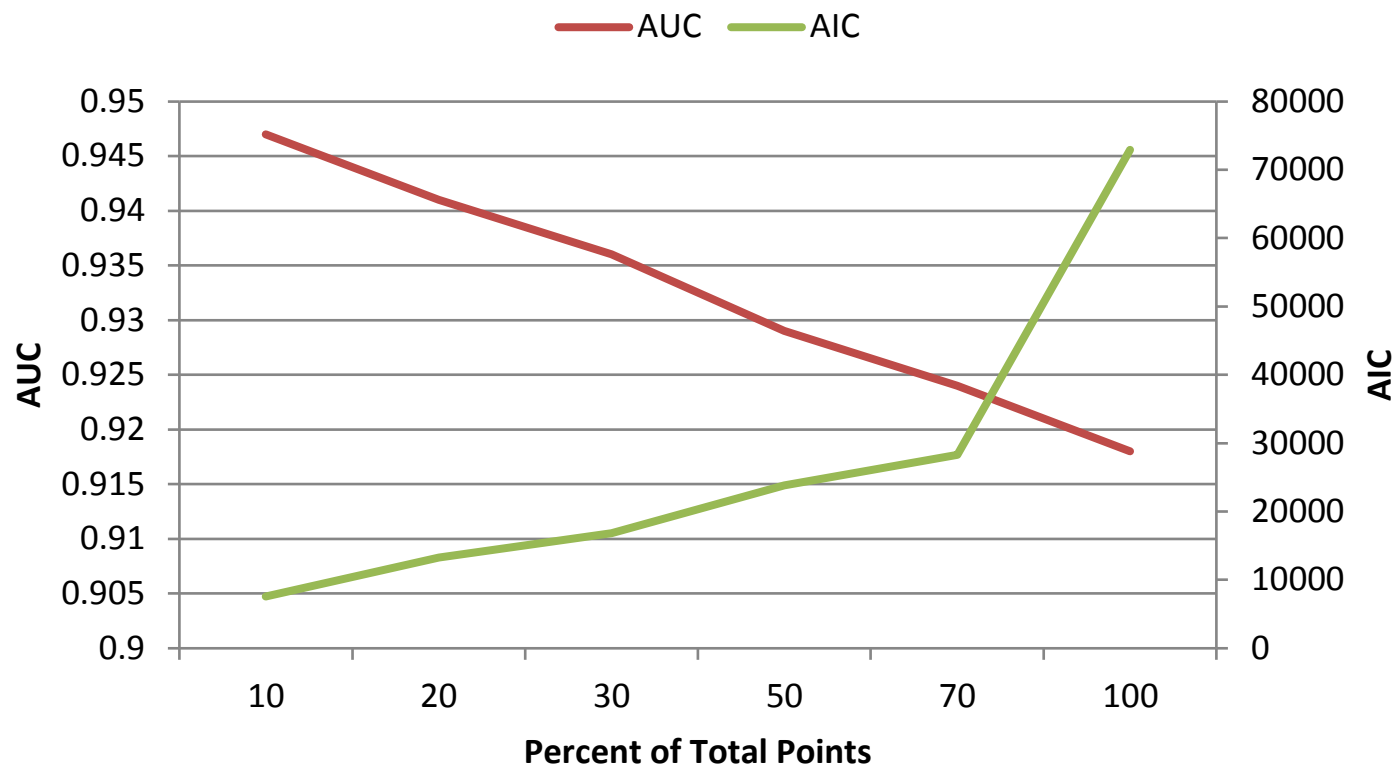


100%
Sample
Points





Impact of Sample Points on Performance Measures

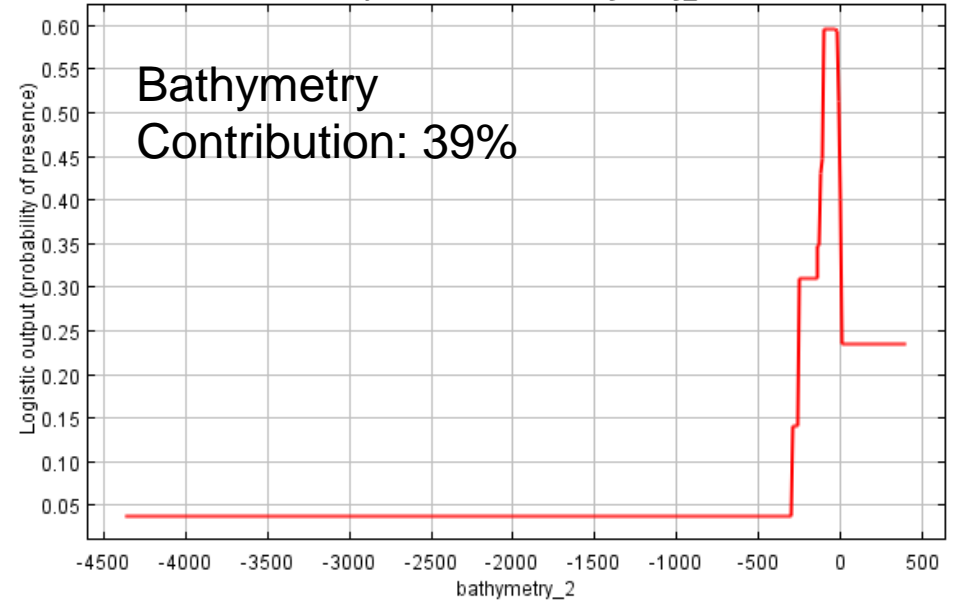


Percent	Number of Samples	Number of Parameters	AIC	AUC
10%	665	136	7549	0.947
20%	1330	166	13,258	0.941
30%	1995	120	16,850	0.936
50%	3326	159	23,839	0.929
70%	4656	162	28,293	0.924
100%	6651	81	72,901	0.918

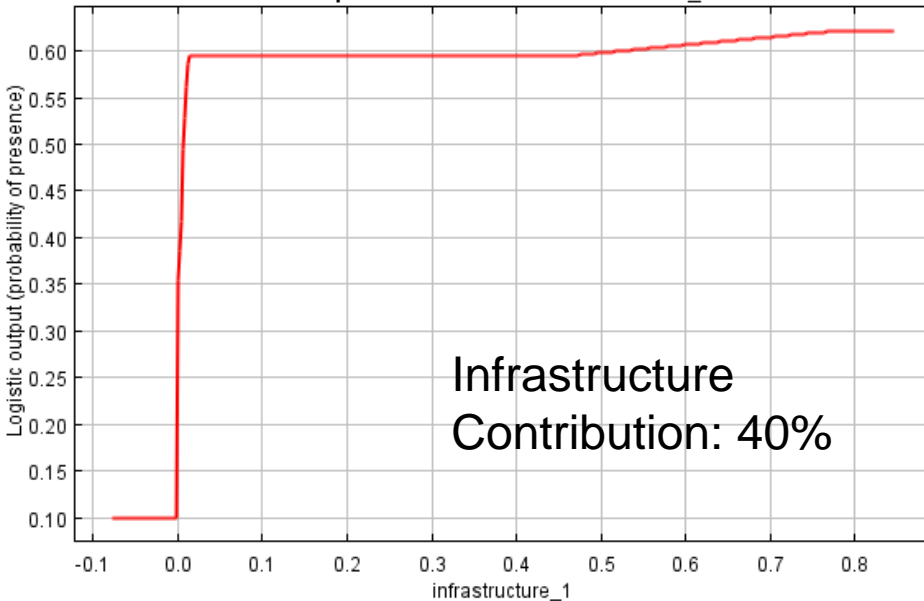
Results

- AUC: 0.92

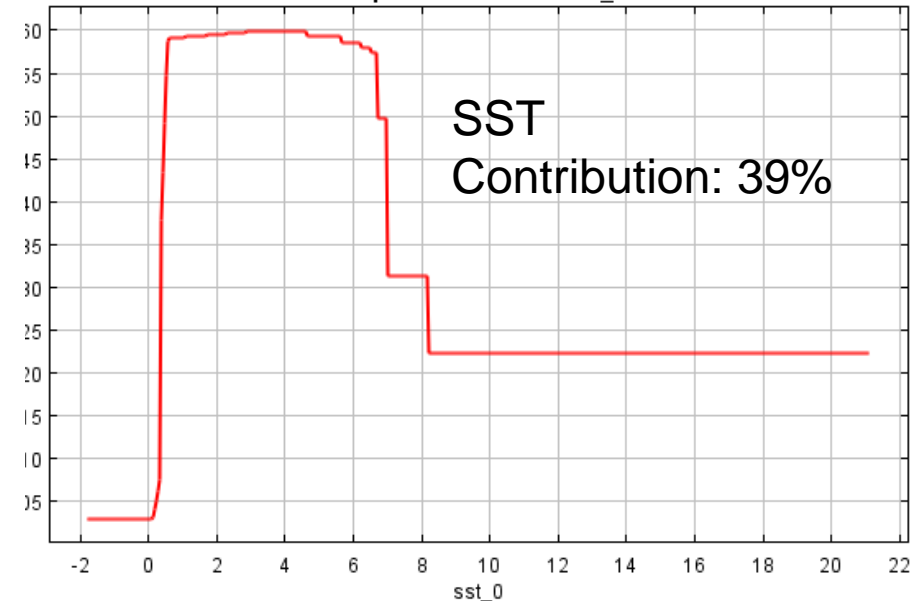
Response of Test to bathymetry_2

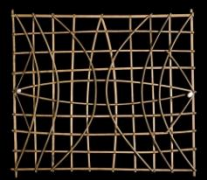


Response of Test to infrastructure_1



Response of Test to sst_0

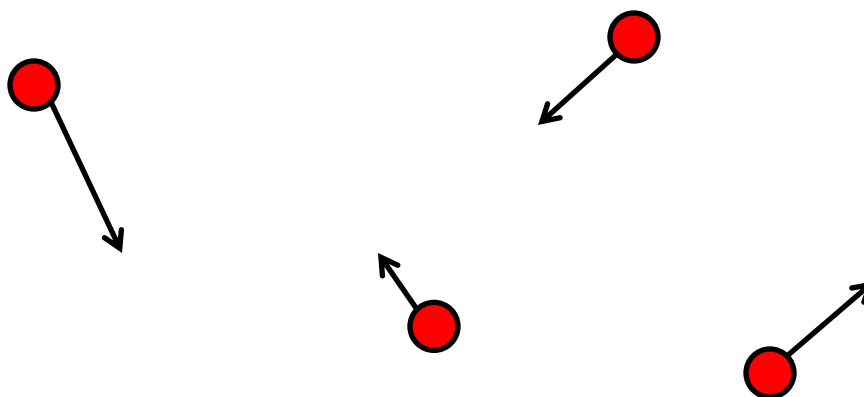


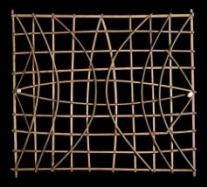


Jiggling The Samples



- Randomly shifting the position of the points based on a given standard deviation based on sample uncertainty
- Running the model repeatedly to see the potential effect of the uncertainty

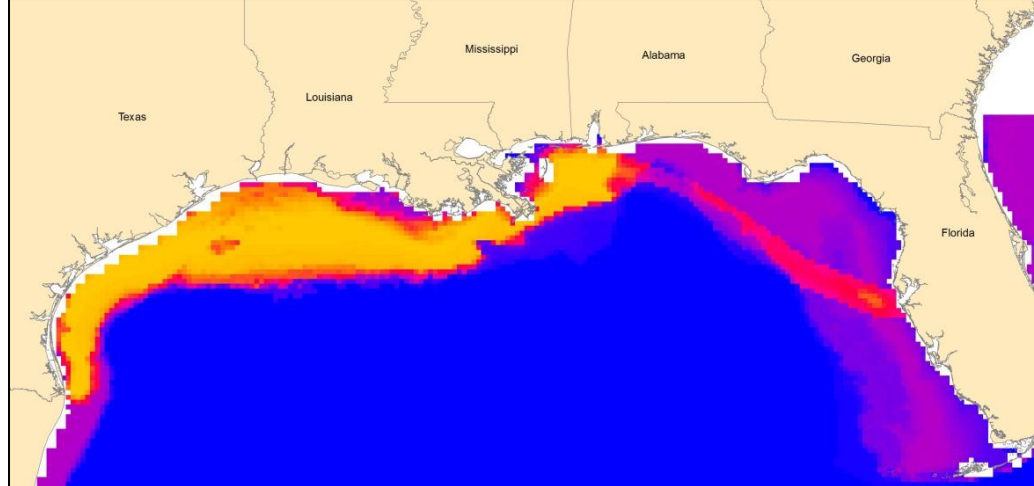




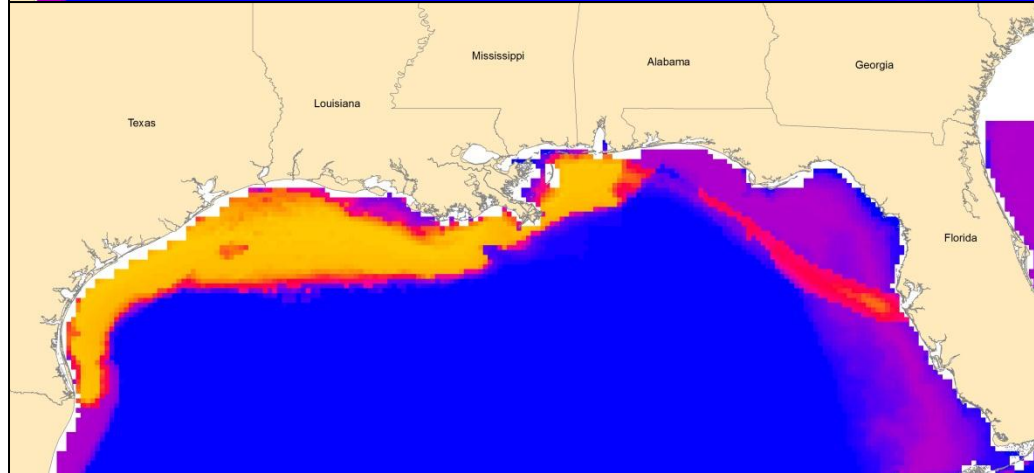
Jiggling



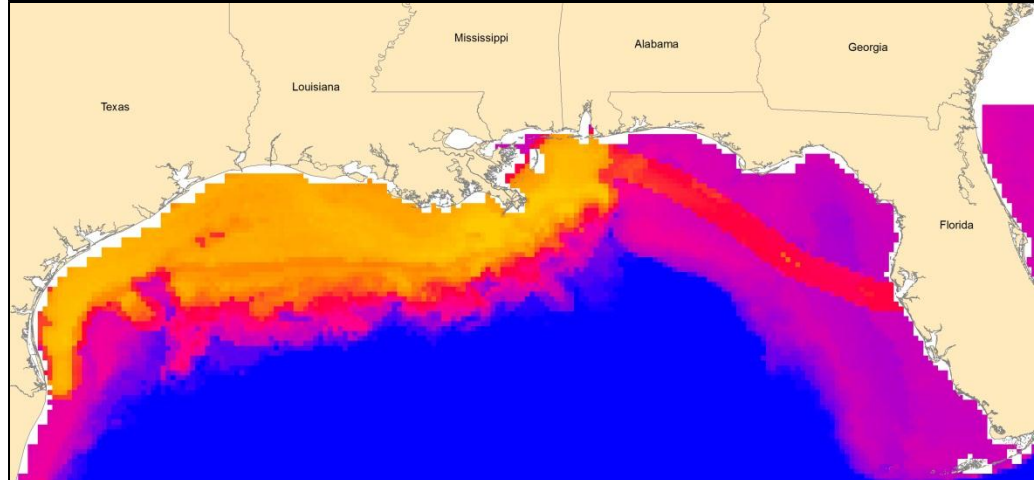
No Jiggling

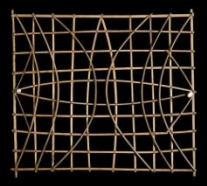


Std Dev=4.4km



Std Dev=55km





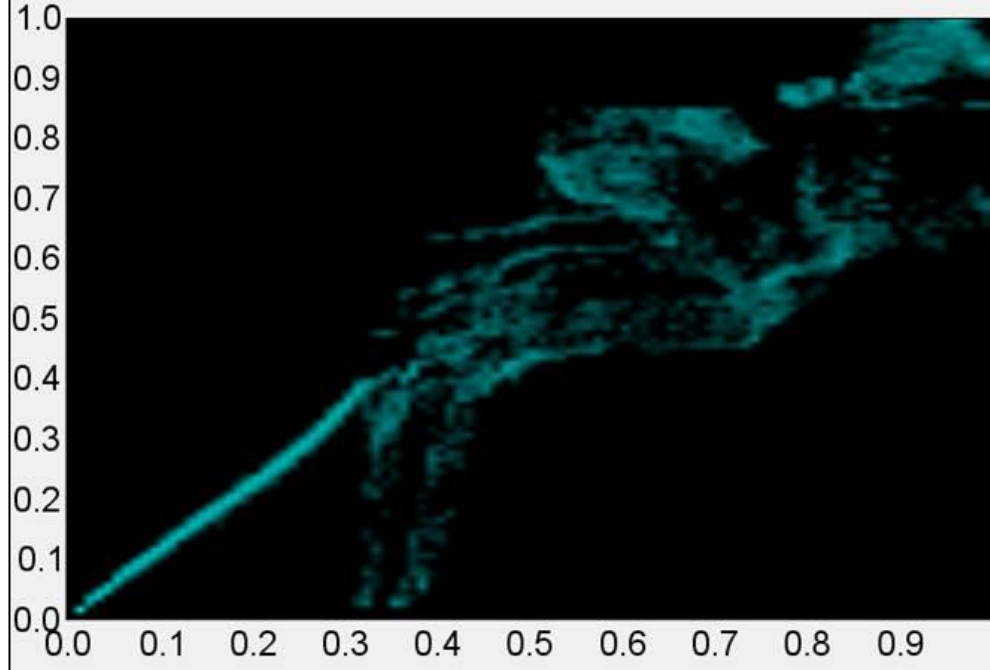
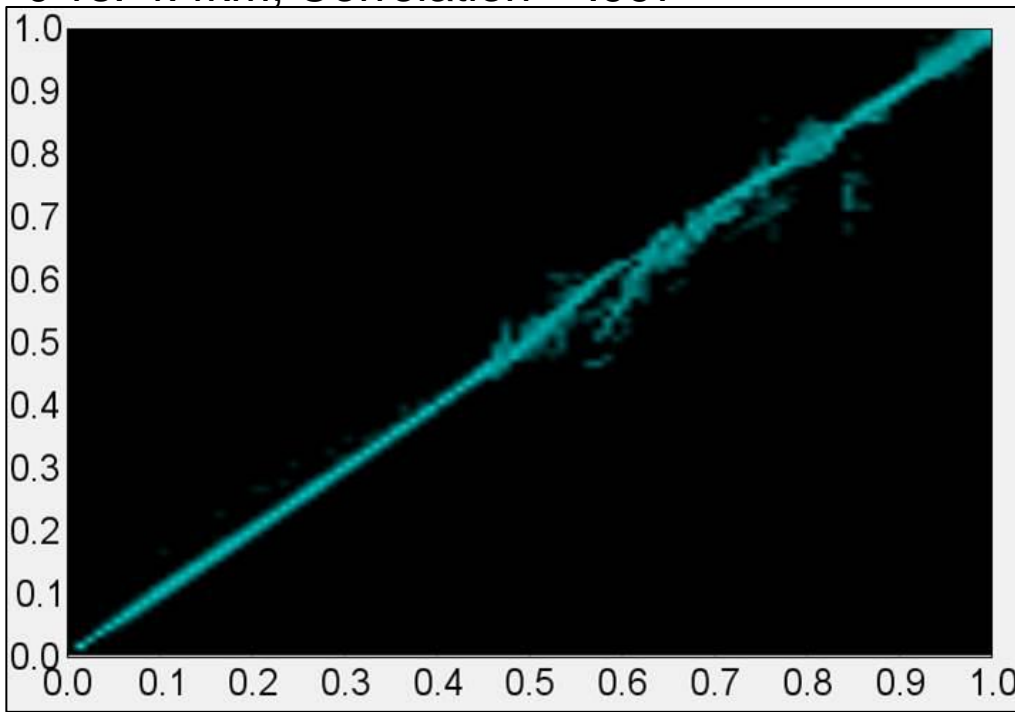
Jiggling



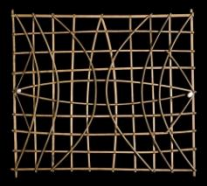
- Spearman's Correlation



0 vs. 4.4km, Correlation = .997



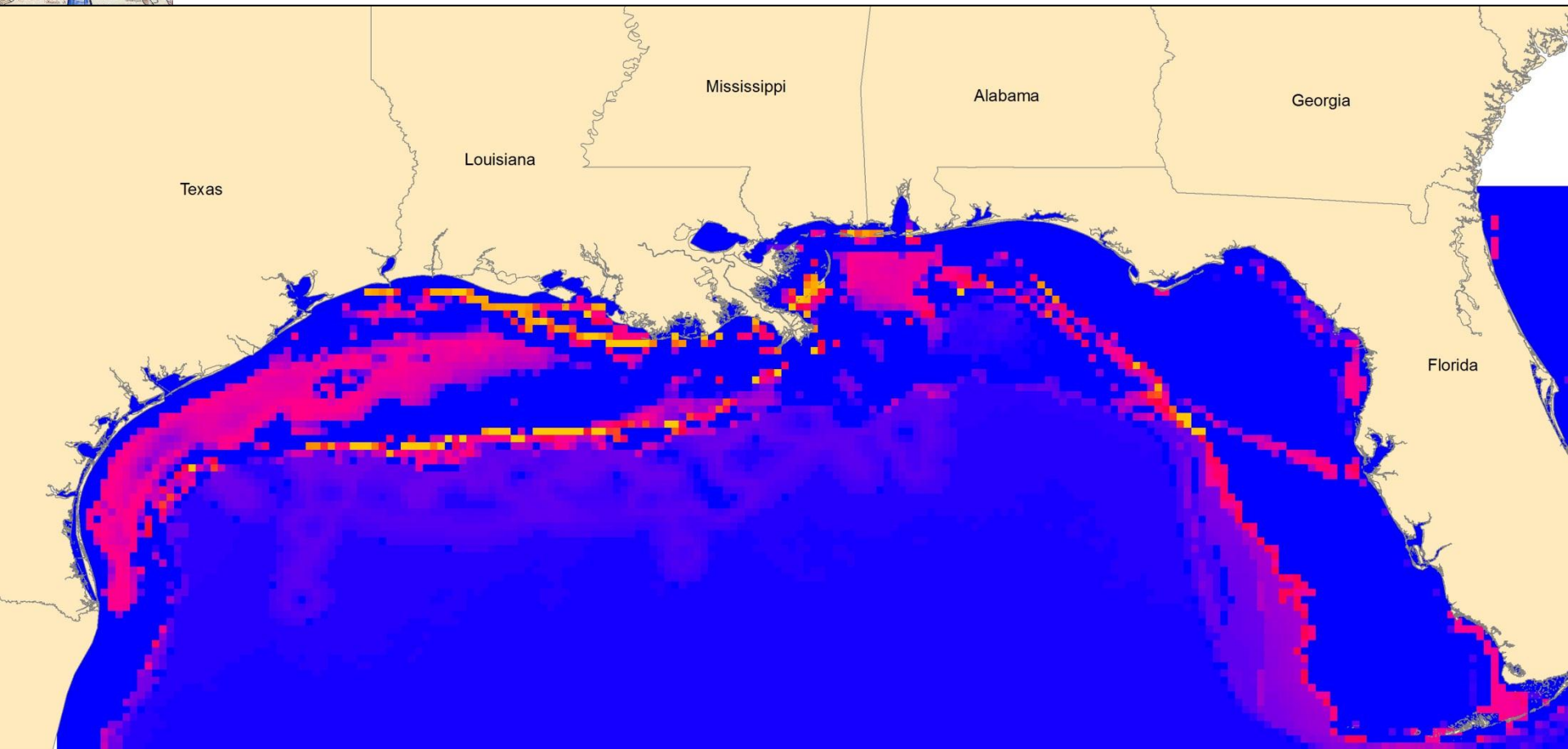
0 vs. 55km, Correlation = .919



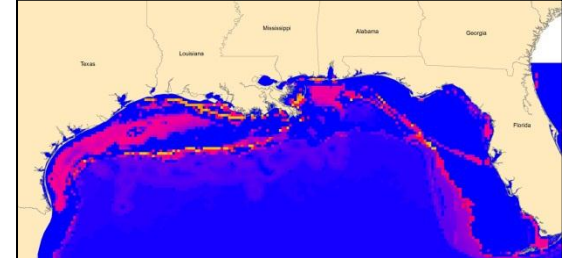
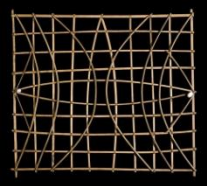
Uncertainty Maps



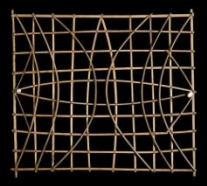
- Standard Deviation of Jiggling Points by 4.4km



Conclusion



- We can improve HSMs over the default settings in Maxent
- We can determine uncertainty in some cases
- Next Steps
 - Support uncertainty for each data point
 - Improve uncertainty analysis and visualization for predictor layers
 - Begin adding uncertainty analysis and maps to products



References



- Anderson, Model Based Inference in the Life Sciences: A Primer on Evidence
- AVHRR Technical Specifications:
http://data.nodc.noaa.gov/pathfinder/Version5.2/GDS_TechSpecs_v2.0.pdf
- Chil's JP, Delfiner P , Geostatistics: Modeling Spatial Uncertainty
- Franklin J, Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge
- Hunsaker CT et. al., Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications
- Maxent: <http://www.cs.princeton.edu/~schapire/maxent/>
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190 (3-4):231-259.
- Warren DL, Seifert SN (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications 21 (2):335-342.
- Wenzhong S, Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses