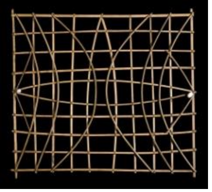
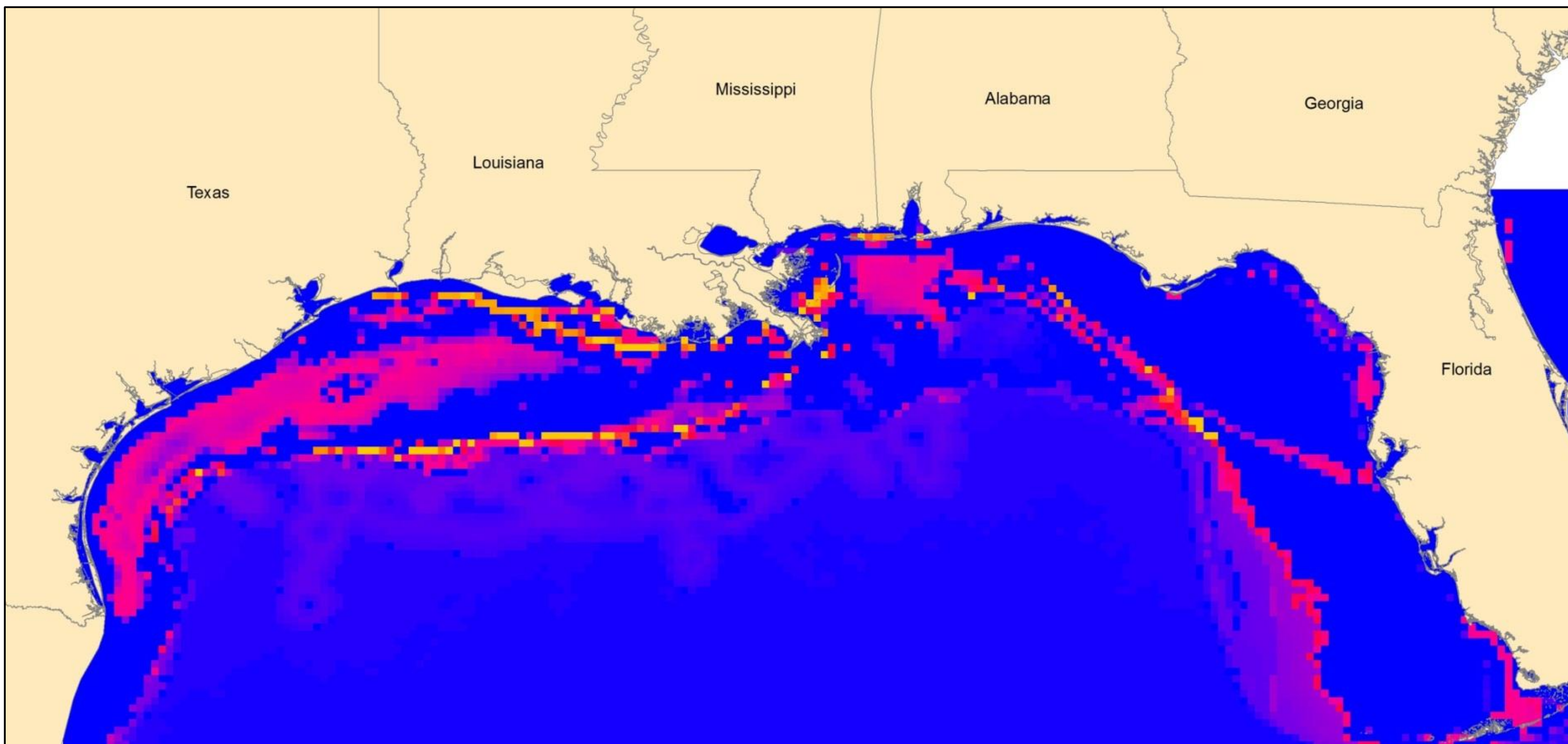


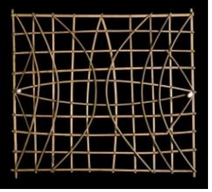



Managing Uncertainty in Habitat Suitability Models

Jim Graham and Jake Nelson

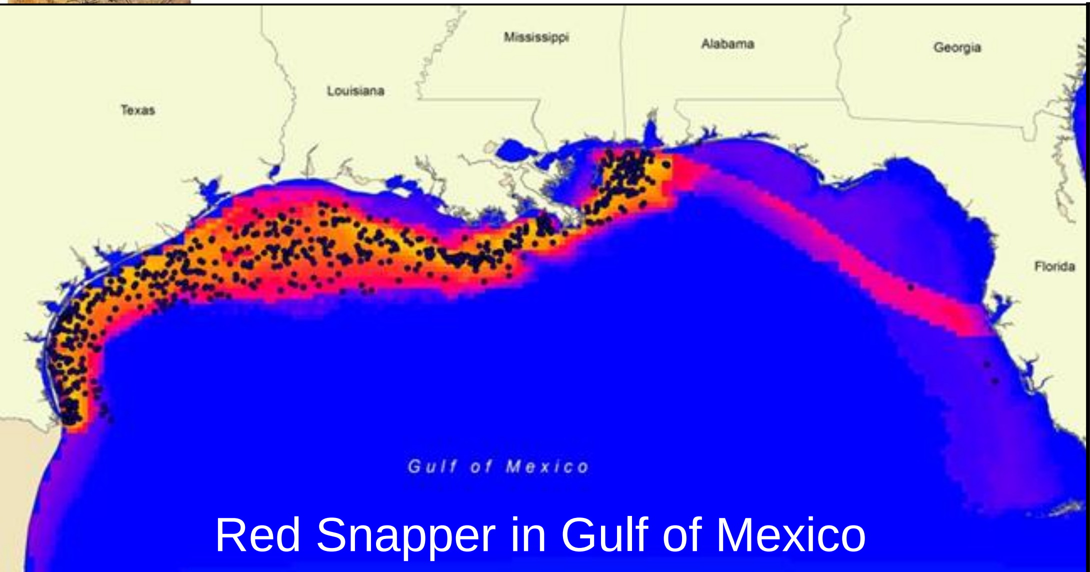
Oregon State University



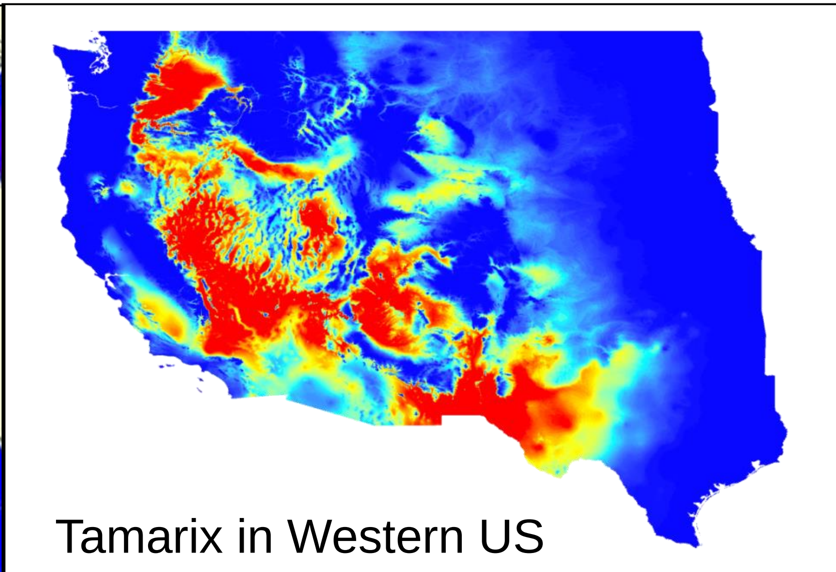


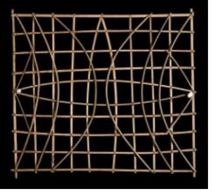

Habitat Suitability Modeling (HSM)

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



Red Snapper in Gulf of Mexico

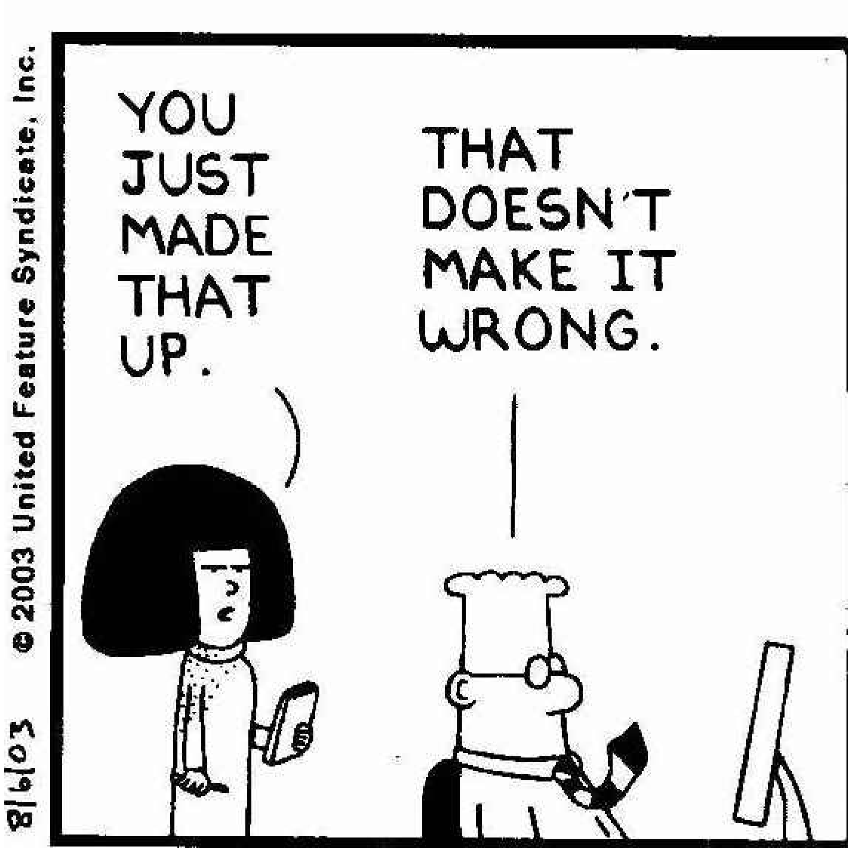

Tamarix in Western US



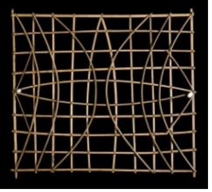


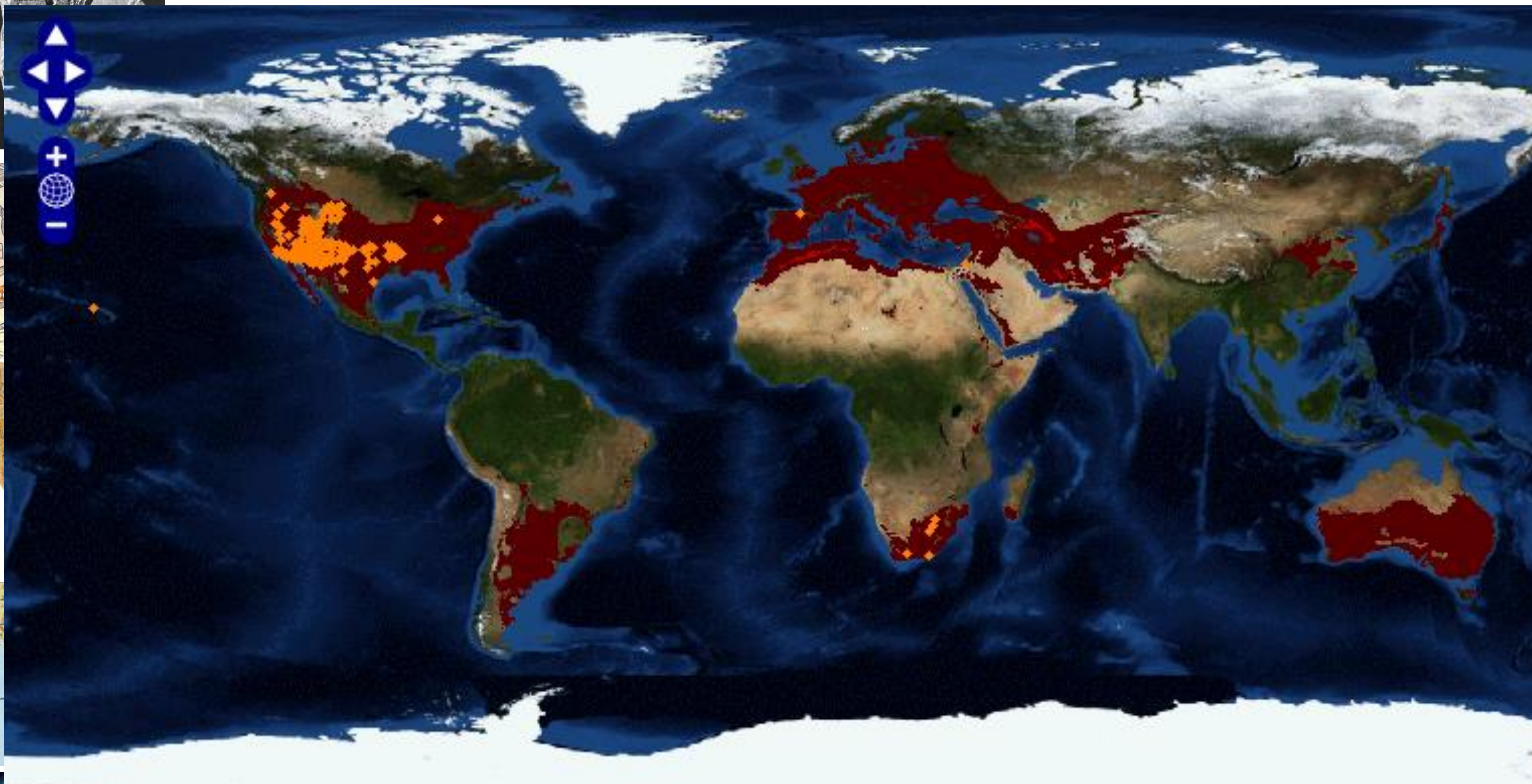

Uncertainty

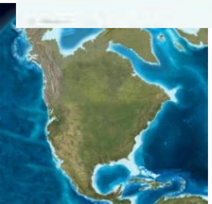
- How “certain” of the data are we?
- How much “error” does it contain?
- How well does the model match reality?
- Goal:
  - Understand and document uncertainties from data collection to publication



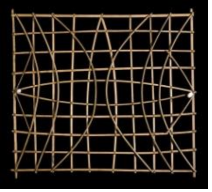
LifeMapper: *Tamarix chinensis*



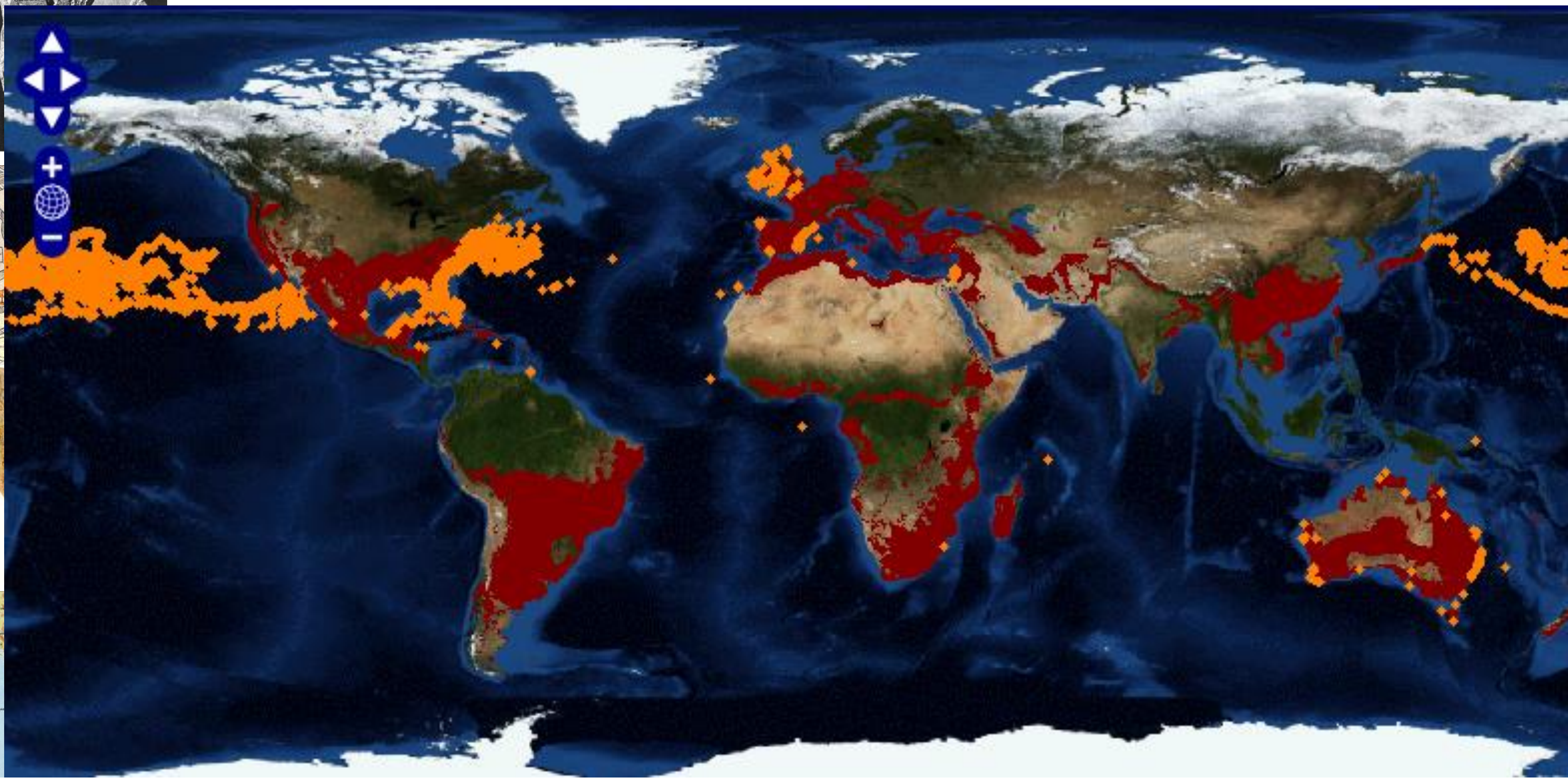




LifeMapper.org



LifeMapper: Loggerhead Turtles



LifeMapper.org

BFRO

THE BIGFOOT FIELD RESEARCHERS ORGANIZATION

Where to find Bigfoot in Redwood National Park

Introduction

Redwood National Park is located in northern California in both Del Norte and Humboldt counties. Since 1957, there have been a total of 69 reported incidents and/or encounters with Bigfoot in both counties. Thus, the park is likely to be home to a thriving Bigfoot population.

Background – What *is* a Bigfoot?

"The term "sasquatch" is an anglicized derivative of the word "Sésquac", meaning "wild man". The original word, in the Stó:lõ dialect of the Halkomelem language, is used by the Coast Salish Indians of the Fraser Valley and parts of Vancouver Island, British Columbia.

Indian tribes across North America have a total of more than sixty different terms for the sasquatch.

"Bigfoot" was a journalistic term generated in the middle of the last century during a rash of sightings in Northern California. The word has come to be recognized widely.

Many different terms have been used by pioneers and later non-native inhabitants of North America, including "skookums" and "mountain devils"."

From www.bfro.net, "What is *a Bigfoot, or Sasquatch?"*

Objective

To determine where the most likely places to see Bigfoot are in order to aid The Bigfoot Field Researchers Organization and Bigfoot enthusiasts in their quests to find Bigfoot.

Assumptions

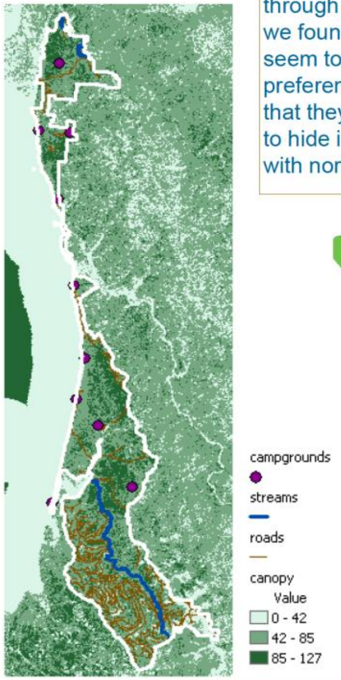
Bigfoot prefer dense canopy cover, nearby sources of water and shady regions. They don't like people, so they avoid campgrounds and roads, though there have been multiple records where citing occur near and on campgrounds and roads.

Limitations

- No detailed layer of stream data
- Hand-plotted campsites subject to human error
- Lack of metadata on collected data layers

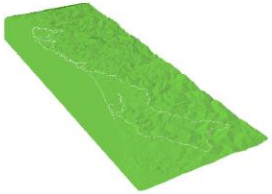
Factors

Campgrounds
Streams
Roads
Canopy
Aspect



3D Analysis

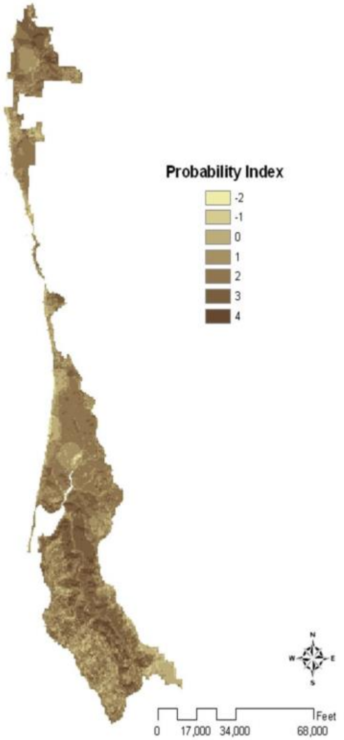
We built a TIN from a NED elevation model and considered the types of slopes and aspects Bigfoot would feel most comfortable in. After going through a number of accounts, we found that slope does not seem to affect Bigfoots' preferences. Instead, we decided that they preferred shadier areas to hide in and so we found areas with north-facing aspects.



Suitability Analysis

We created buffers around streams, campgrounds, and roads and assigned opportunity and constraint values to each feature. We then assigned opportunity and constraint values to canopy density and aspect maps. These values were used to create a map of Bigfoot habitat suitability.

Suitability Criteria	
	Values
Opportunities	
Dense canopy layer	
80+	+2
60-80	+1
Less than 60	0
1 mile buffer from major streams	+1
North-facing aspect	+1
Constraints	
200 feet buffer from roads	-1
1 mile radius from campsites	-1



"The lady giving me a ride into town told me to look out my window to the right... I saw a very large hairy creature walking across a wide, dried up creekbed… My friend told me she sees them all the time out there near her home, so it was just a normal thing for her"
- Bigfoot sighting, Del Norte County, January1979

Conclusion

The most suitable Bigfoot habitats are along Redwood Creek ~~and near the northern border of the park. Areas shaded in dark~~ brown are where you should go to find Bigfoot. However, these probabilities are not very strong and we suggest further analysis on other parks to better find Bigfoot.

Acknowledgments

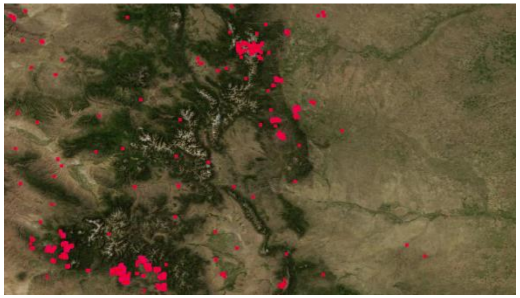
Poster & analysis by Amanda Carlson & Katharine Guan, University of California, Berkeley

Data Sources: USGS Seamless Server, atlas.ca.gov, www.bfro.net

Projection: NAD 1983 California State Plane I (feet)

HSM Process

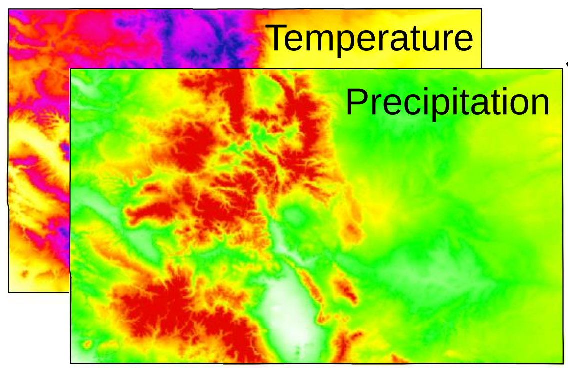
Occurrences (Sample Data)



Spreadsheets

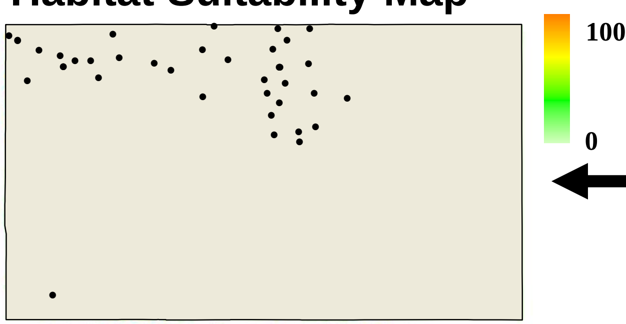
Lat	Lon	Temp	Precip
-105.504	40.35819	5.32	58.4
-107.472	40.498	6.31	47.6

Environmental Layers (Predictors)



Modeling Method

Habitat Suitability Map

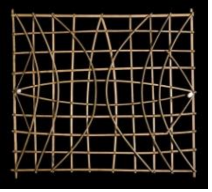


Map Generation

Model Parameters

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

Model of Purple Loosestrife in Colorado, Catherine Jarnevich, PhD

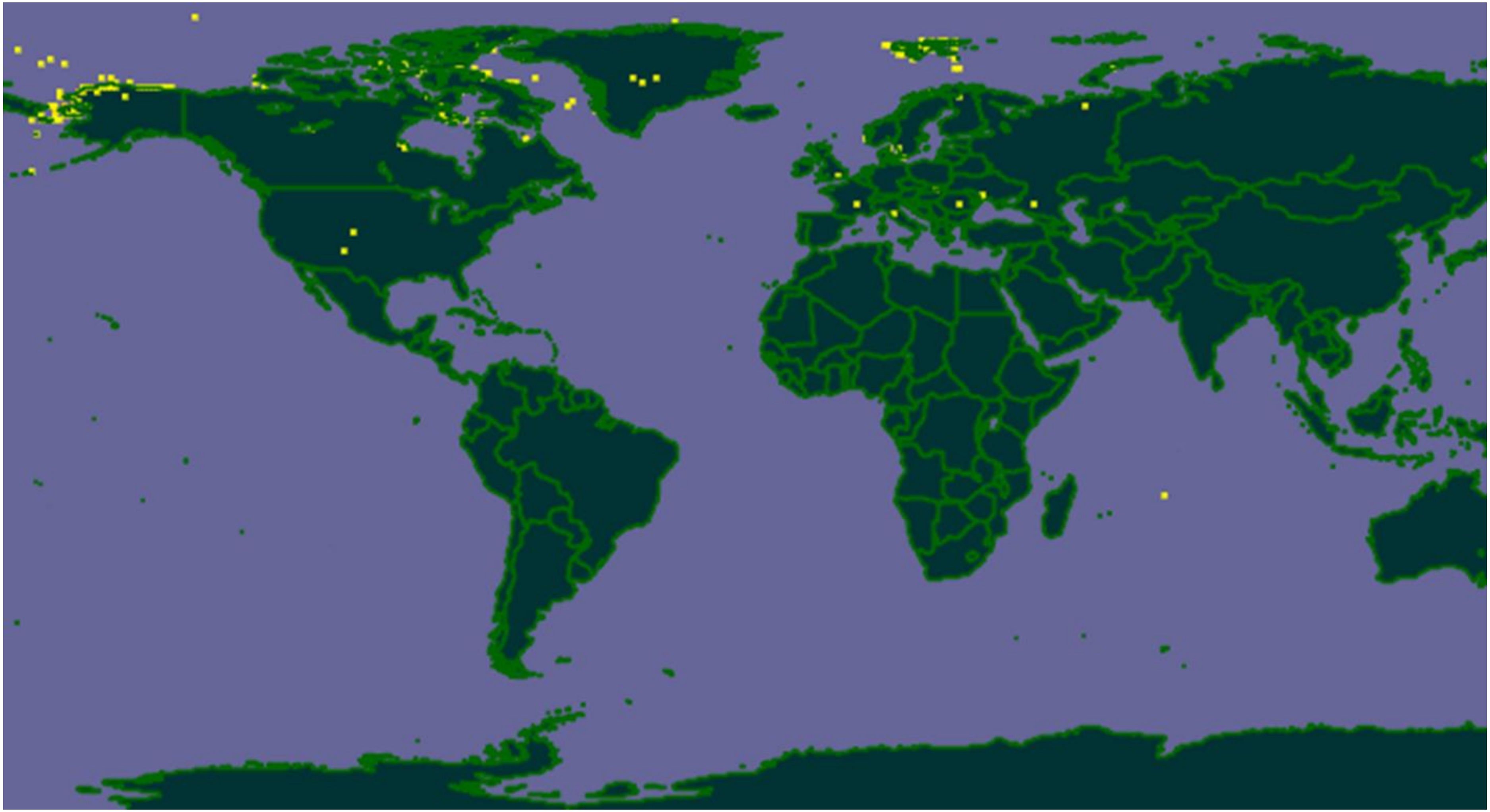





Input Data

- 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



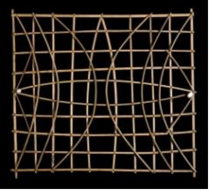
From The Global Biodiversity Information Facility (www.gbif.org, 2011)

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

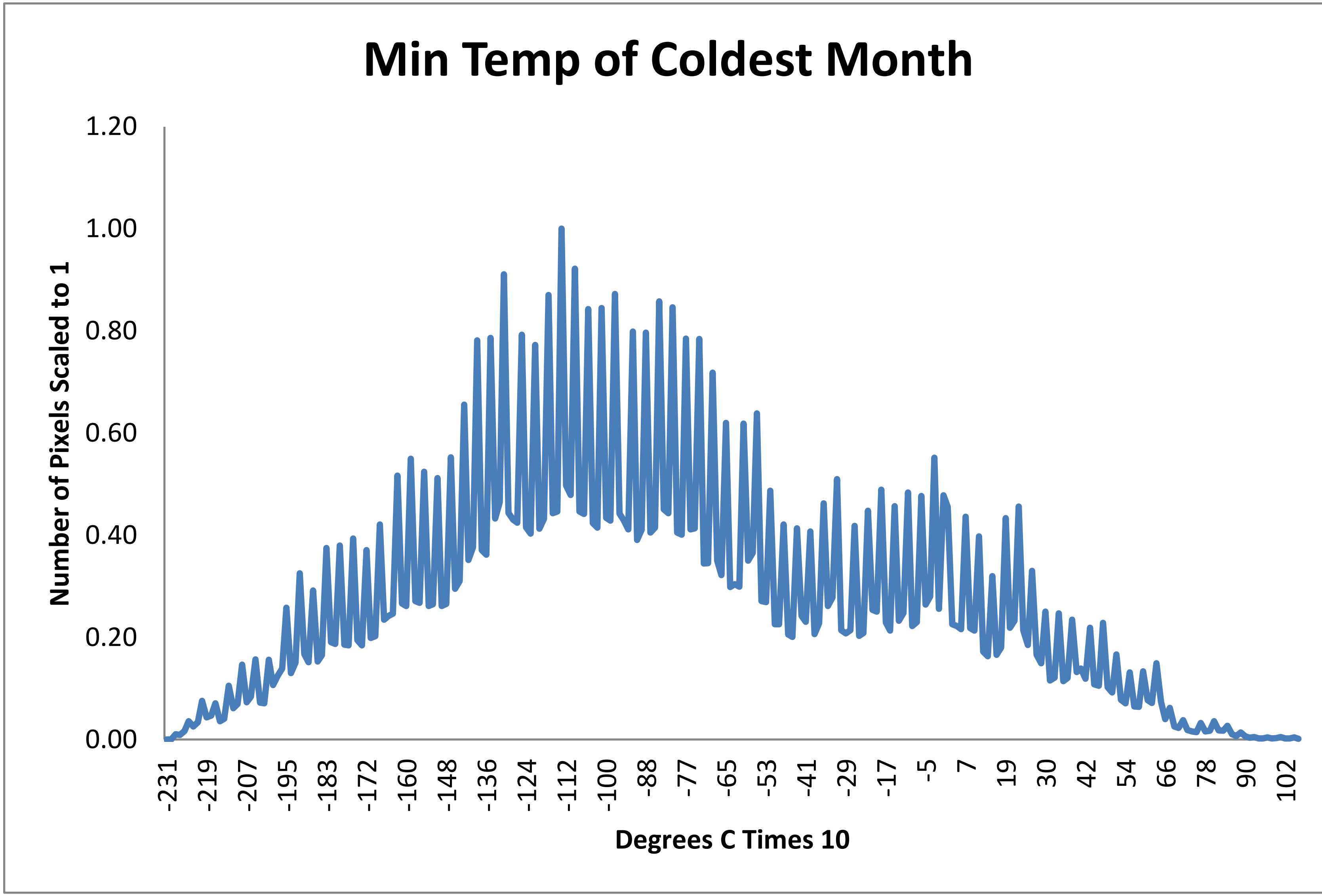
Predictor Layers

- 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

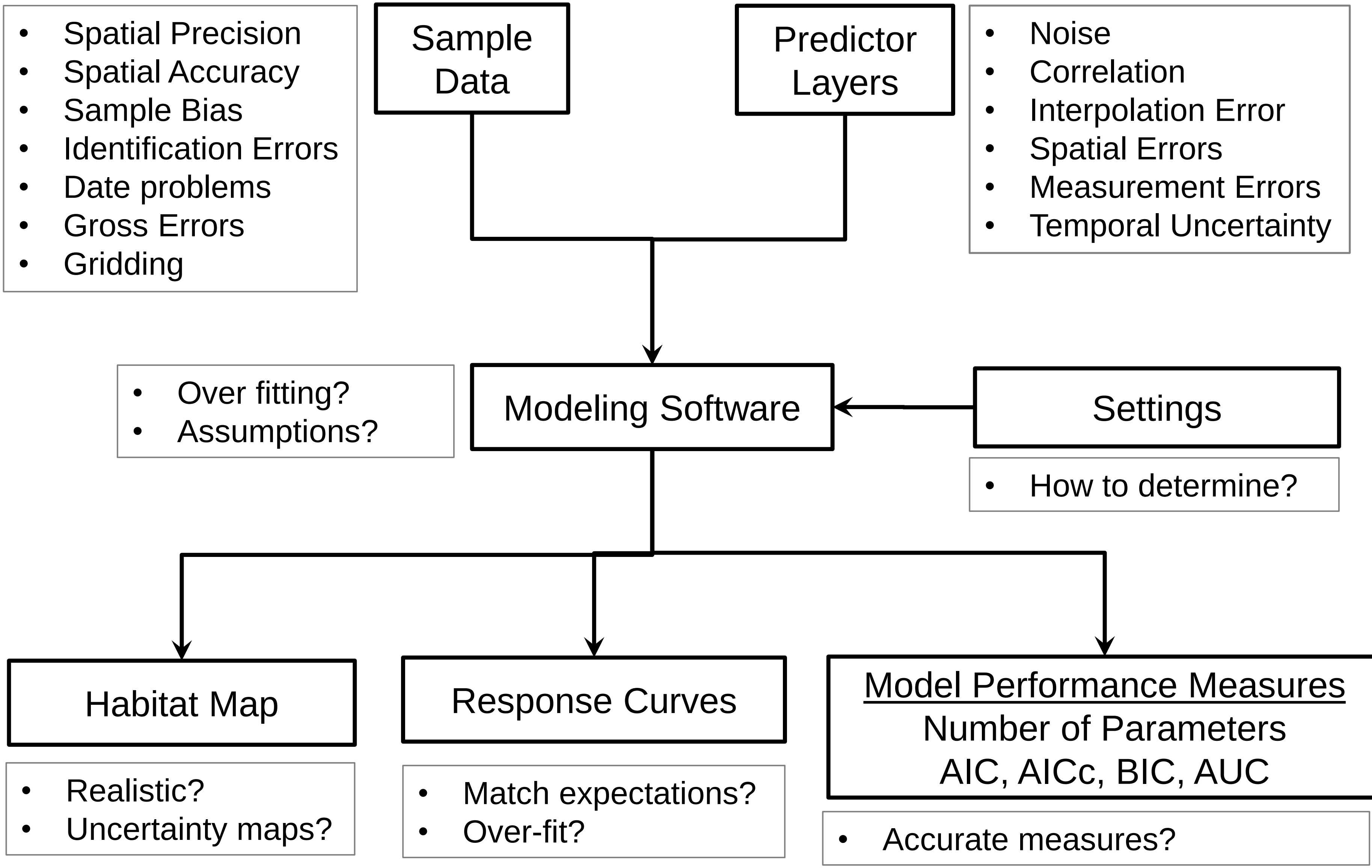





BioClim/WorldClim Data

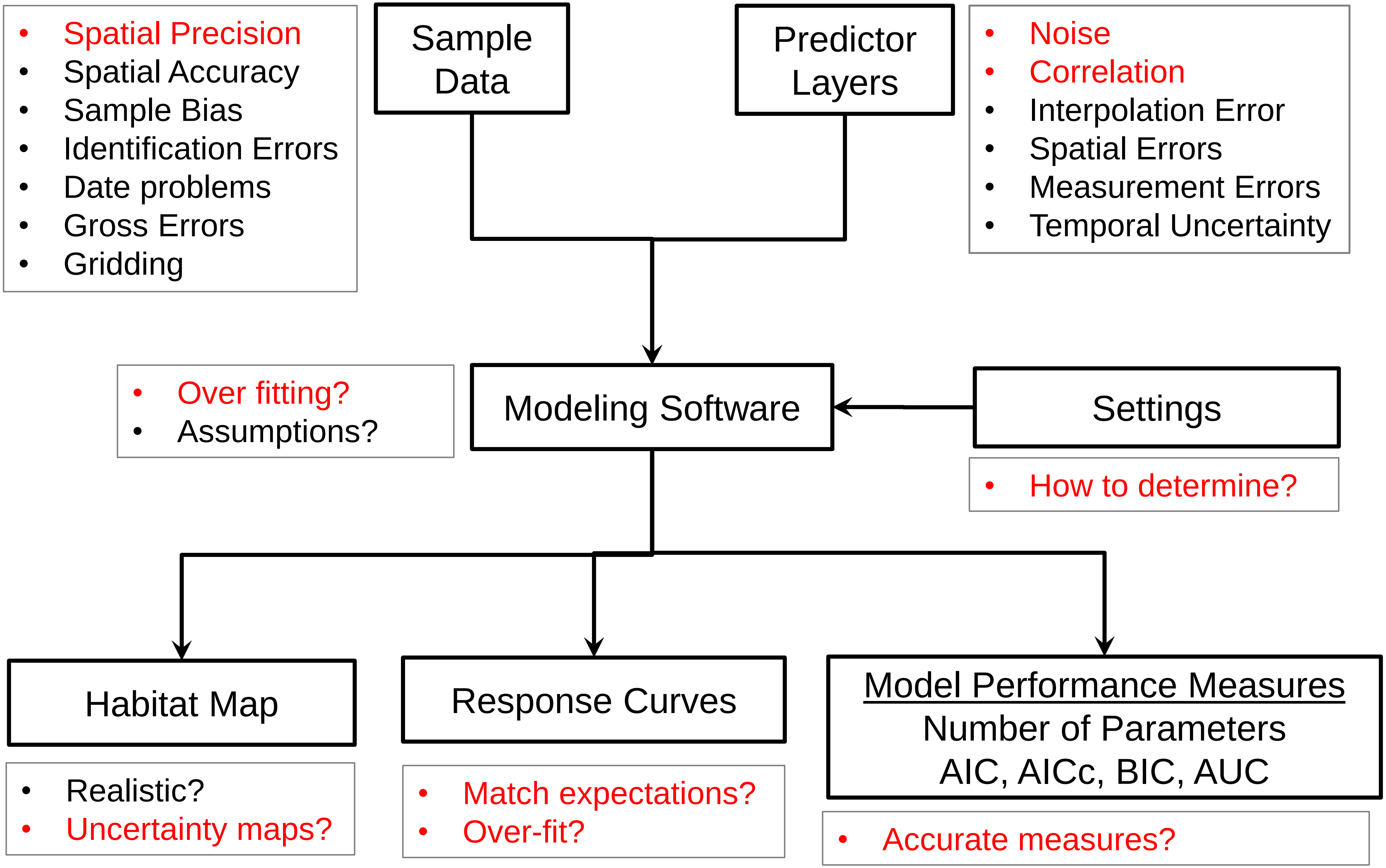


Road Map of Uncertainty

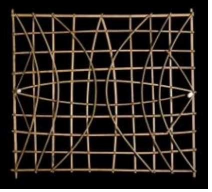


What is the best model?

Road Map of Uncertainty

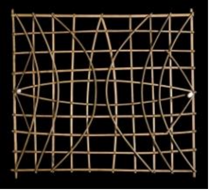


What is the best model?




SEAMAP

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



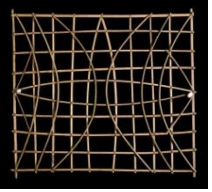
Red Snapper

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

www.safmc.net

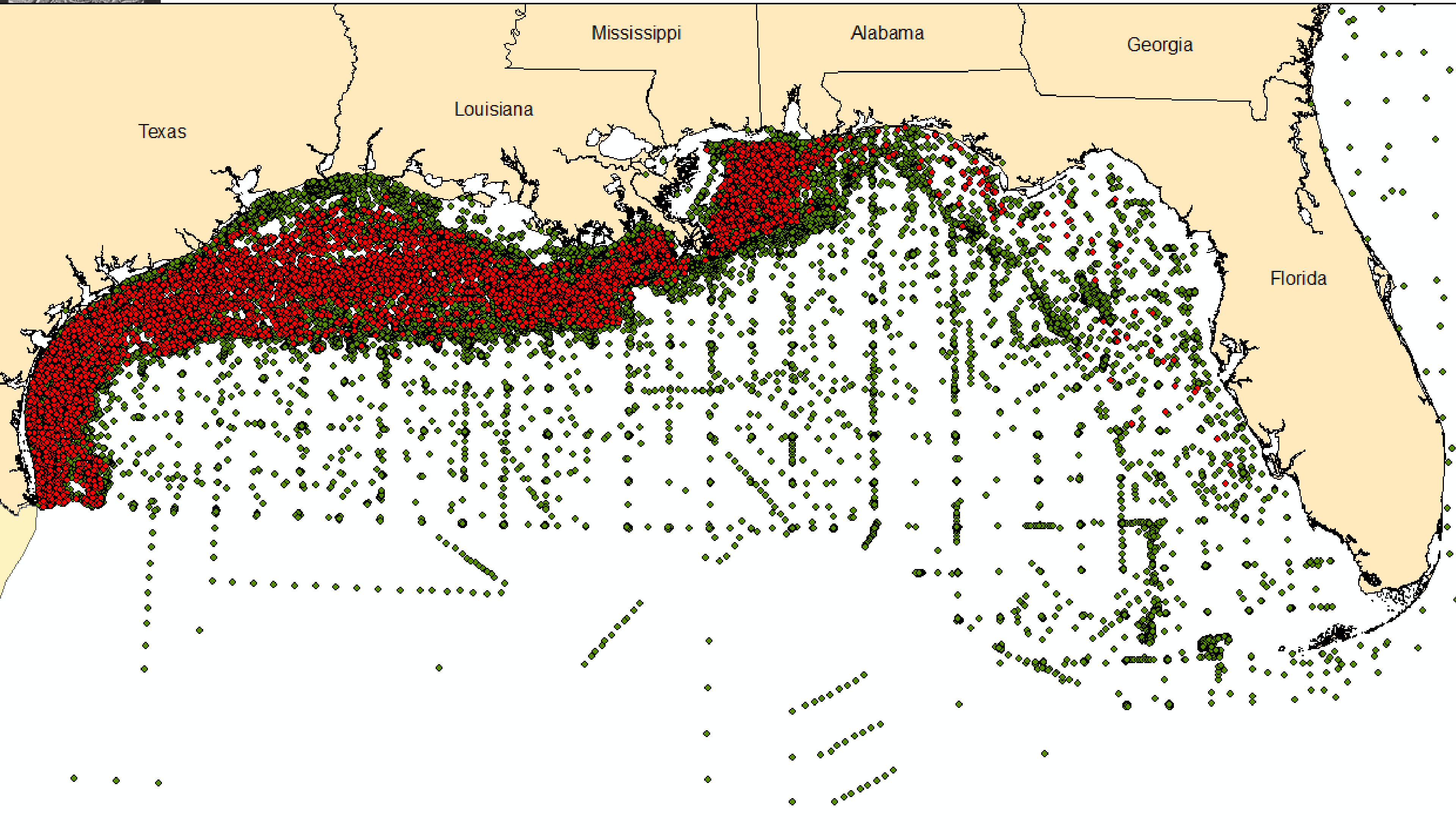
lickyourownbowl.wordpress.com

outdooralabama.com



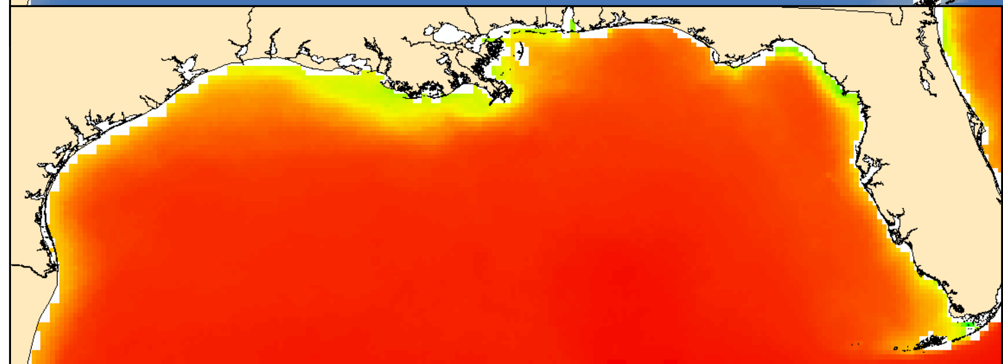
SEAMAP Trawls (>47,000 records)

Red Snapper Occurrences (>6,000 records)

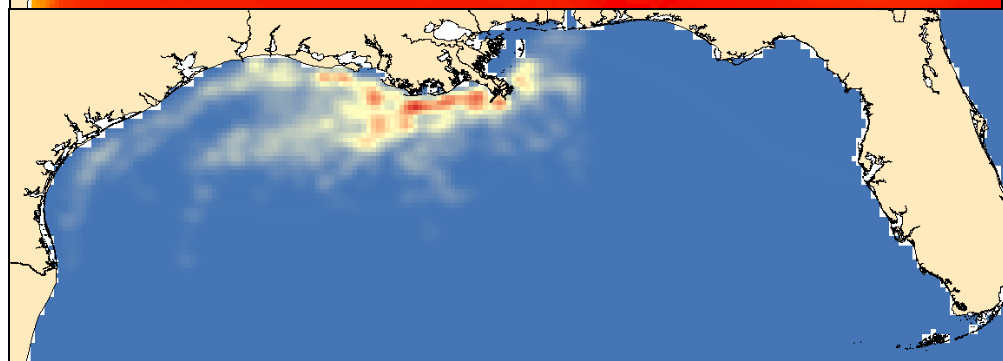


Sea Surface Temperature (SST)
NOAA
AVHRR Pathfinder Satellite
Spatial: 9km
Measured: <0.4 K

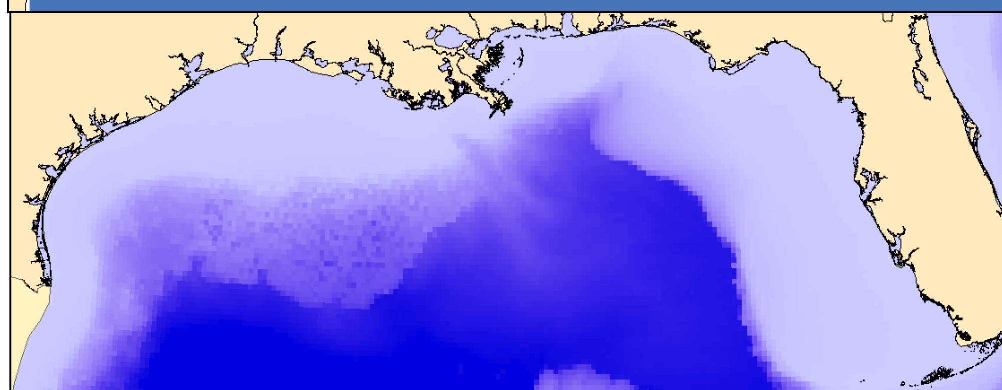
Net Primary Production (NPP)
Milligrams of Carbon per
Meter Squared per Day
OSU Ocean Productivity
Uncertainty?



Density of Platforms and Pipelines
Created from Bureau of Ocean Energy
Management (BOEM) Point Data Set
Uncertainty?

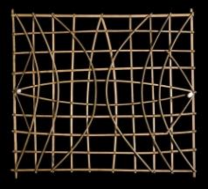


Bathymetry
Resampled from 90m
NOAA Coastal Inundation Dataset
And others
Uncertainty < 9km



Predictor Resolution

- Should model at the lowest resolution of the predictor layers or lower
- Lower resolution:
  - – Can reduce problems from sample data uncertainty
  - – In same cases can combine sample data into measures such as CPUE, density, abundance
  - – Reduces detail of final model
- For these tests: 9km

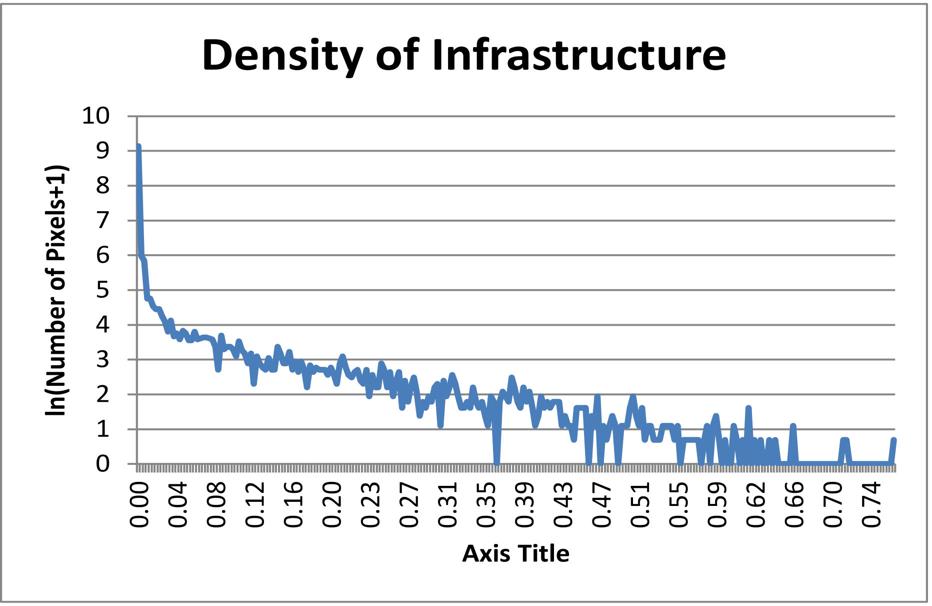


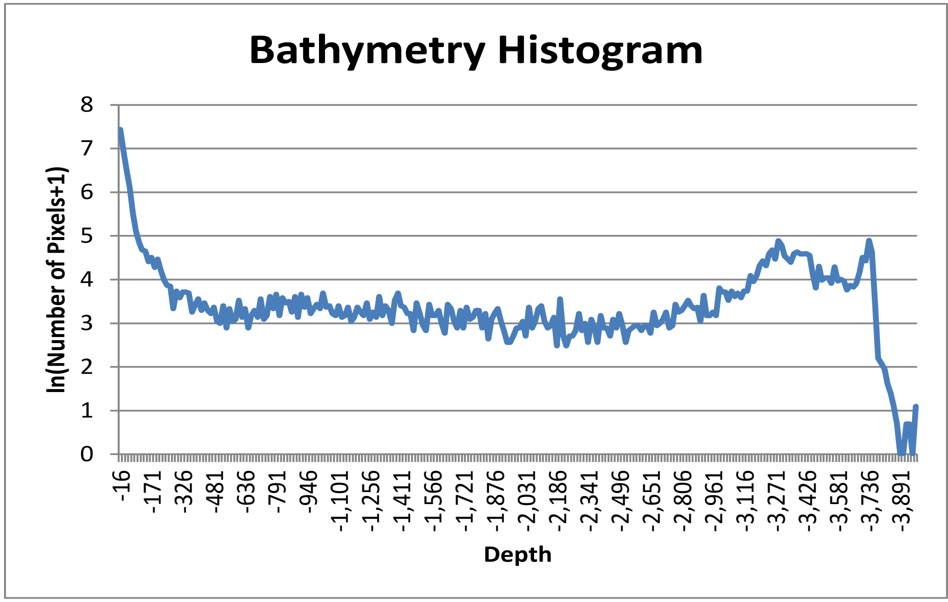


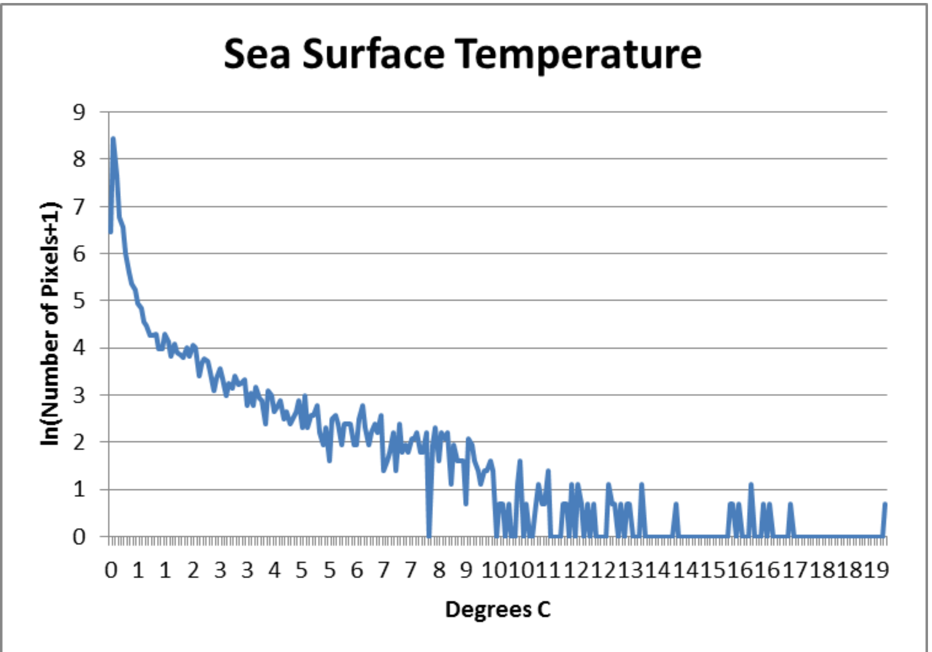

Sample Data

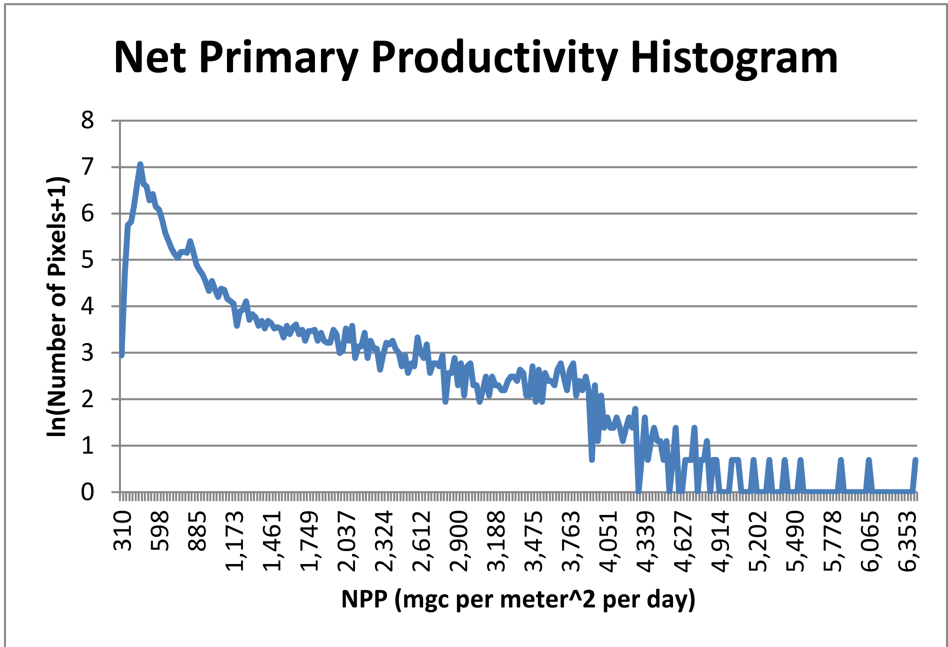
- Spatial Precision
  - Standard Deviation of about 2 km
- Spatial Accuracy:
  - < 1km
- Sampling Bias:
  - Some areas more heavily sampled
- Identification Errors:
  - Unknown but red snapper are pretty easy to identify
- Date problems:
  - Not a temporal model
- Gridding:
  - Not at 9 kilometers

Histograms of Predictor Layers



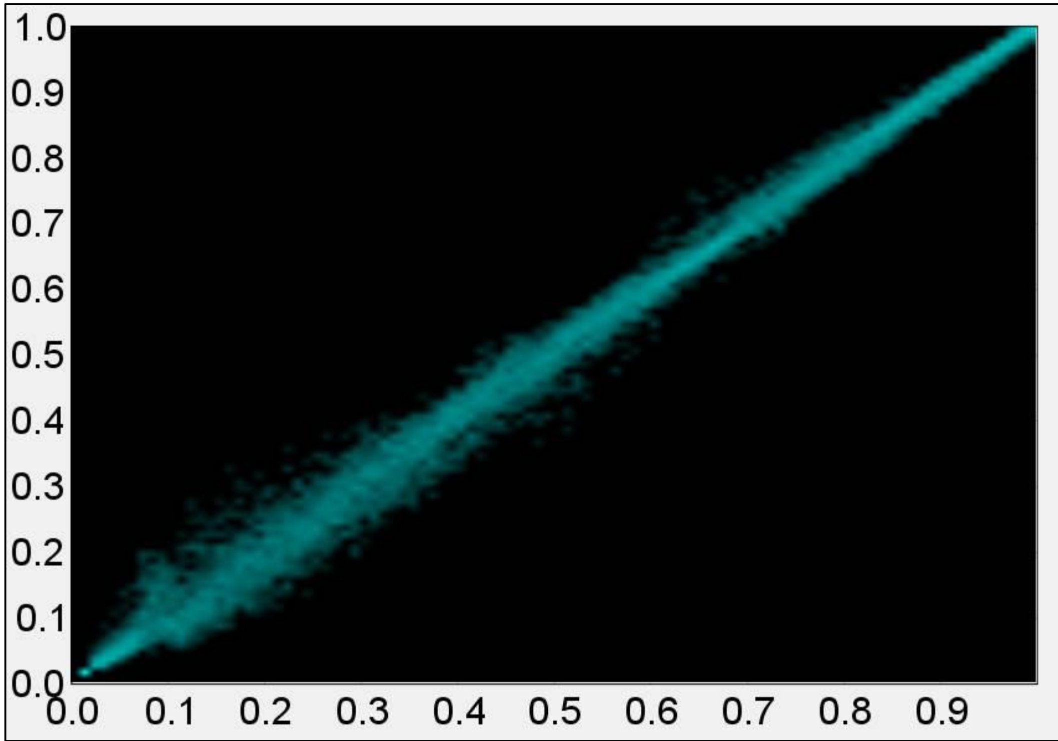






Predictor Layer Uncertainty

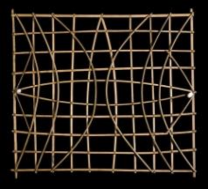
- Noise: Minimal
- Correlation: Eliminate NPP
- Spatial Errors: Unknown, < 9km?
- Measurement errors: Minimal to Unknown



Spearman's Coefficient
NPP to SST
Correlation: 0.99

Modeling Software: Maxent

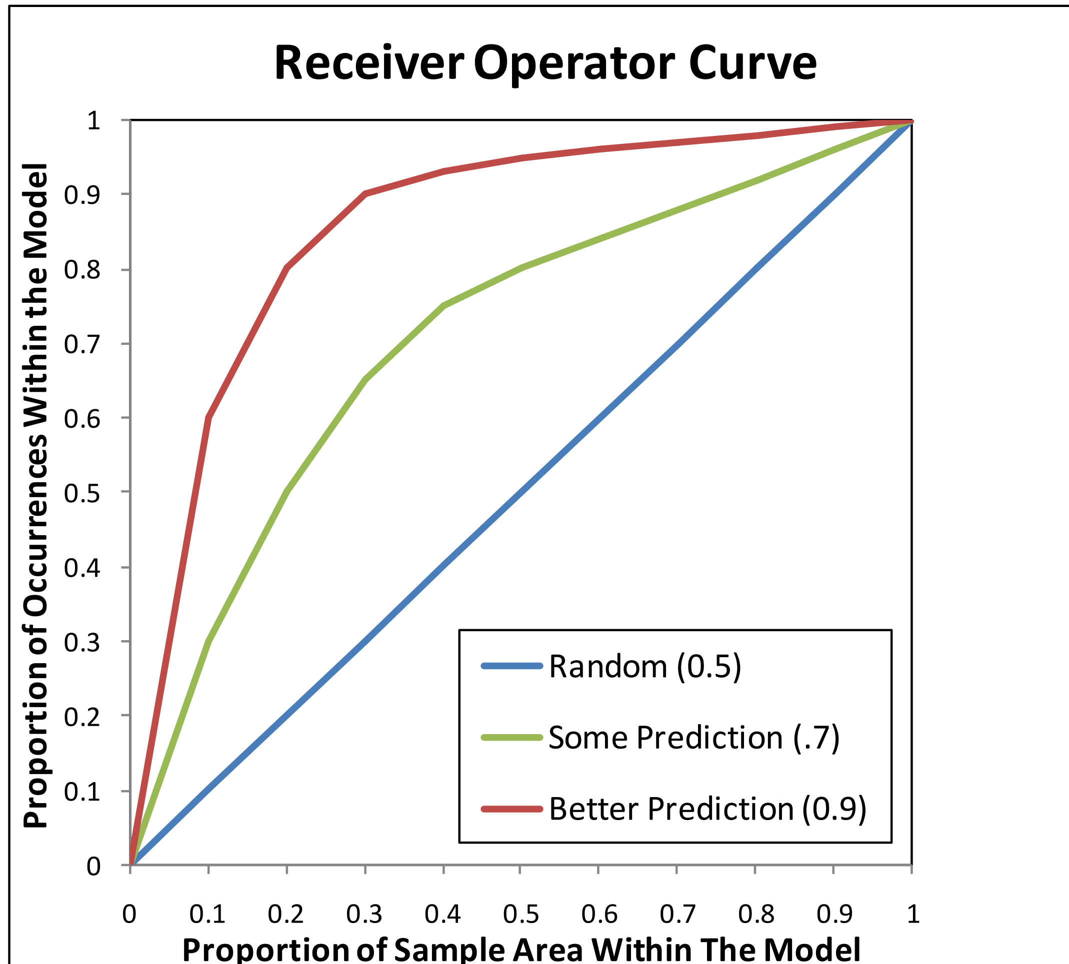
- Popular, relatively easy to use
- Only requires presence points
- Known for over-fitting
- Performs "piece-wise" regression
- Assumes:
  - Predictors are error free
  - Independence of errors in response
  - Lack of correlation in predictors
  - Constant variance
  - Random data collection over sample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

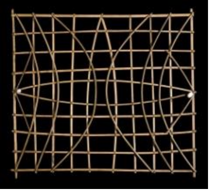





Area Under the Curve (AUC)

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



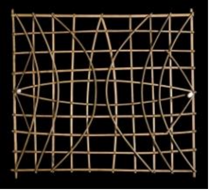




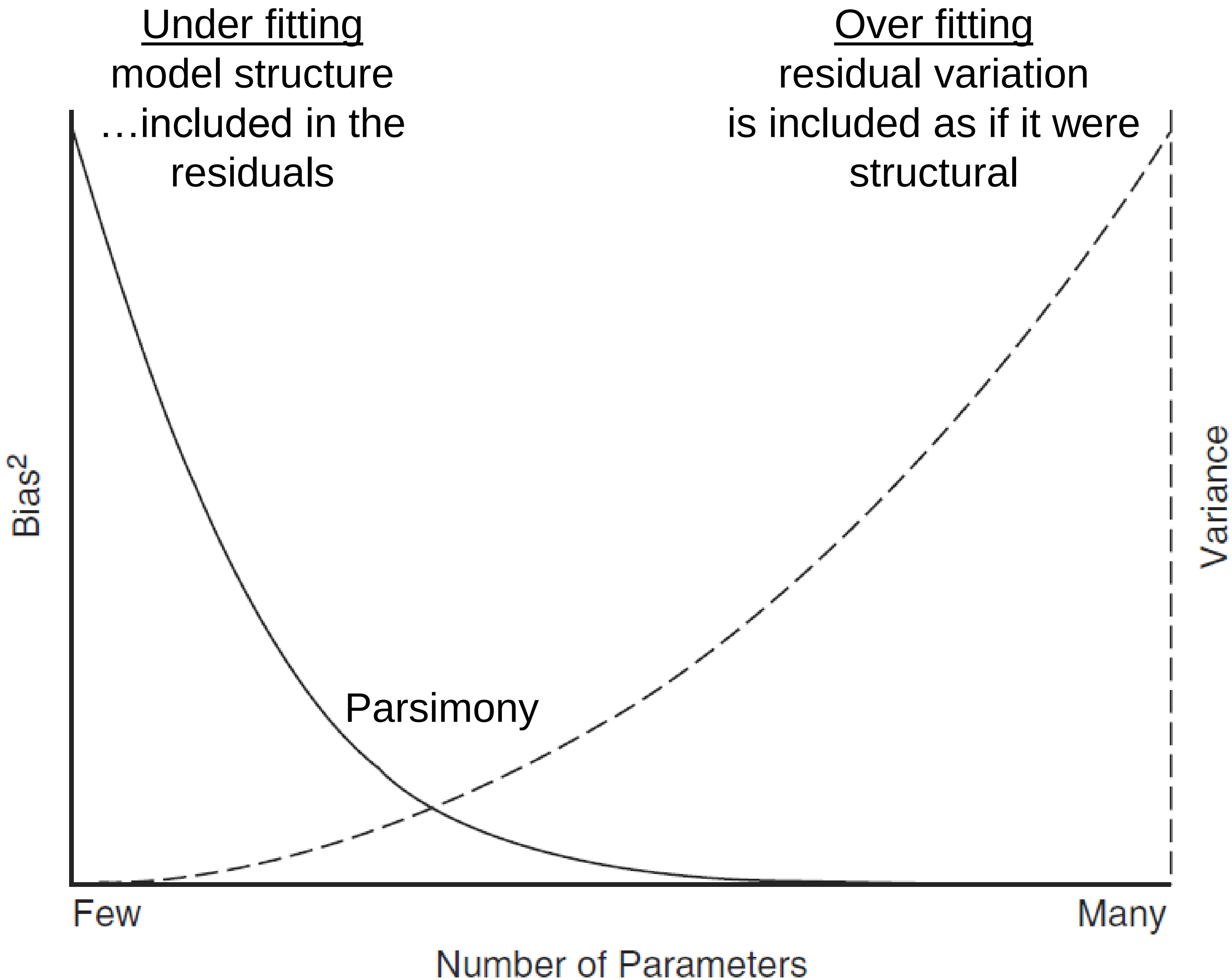

Akaike Information Criterion (AIC)

- 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

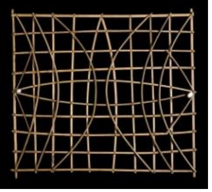
Parsimony





Anderson



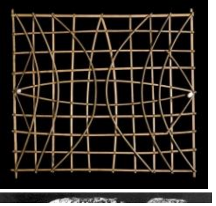



Experiments

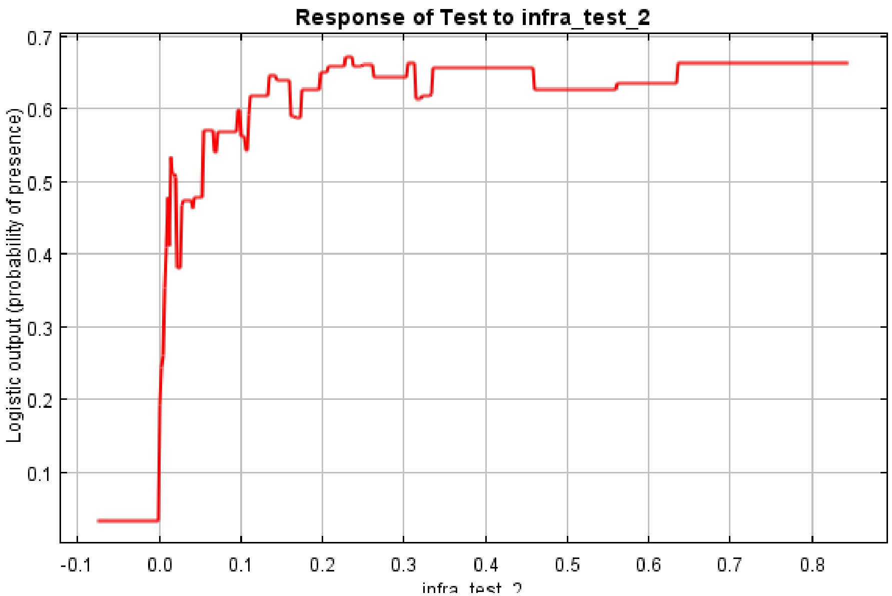
- Different regularization multipliers
  - Select regularization for other tests
- Try different numbers of samples
  - Determine the number of samples to use
- “Jiggle” the sample data spatially
  - Introduce different amounts of spatial error

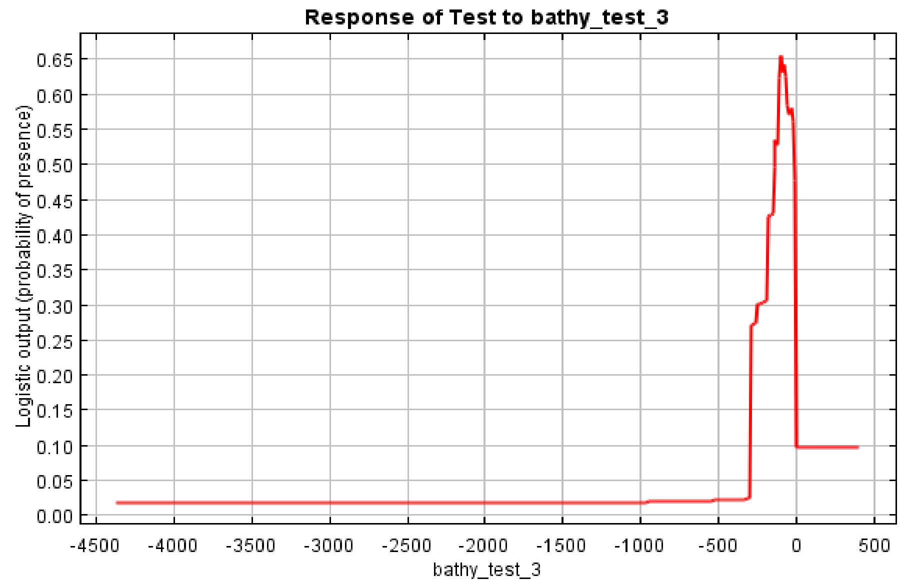
BlueSpray

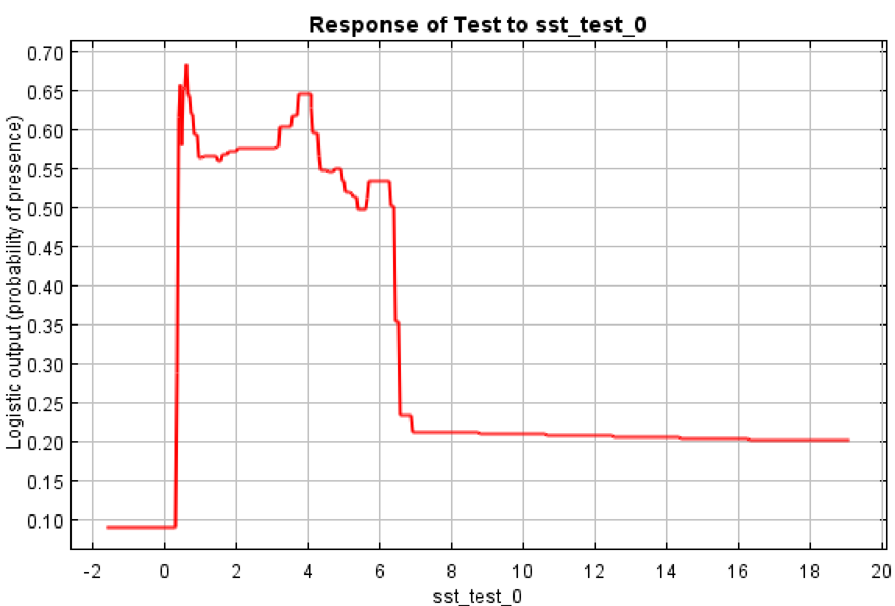
- GIS Application for Natural Resources and Environmental Research
  - Specifically HSM
- Available at www.schoonerturtles.com
- Free for beta testers in environmental research and conservation
- Contact Jim at: jimg@schoonerturtles.com for the passcode

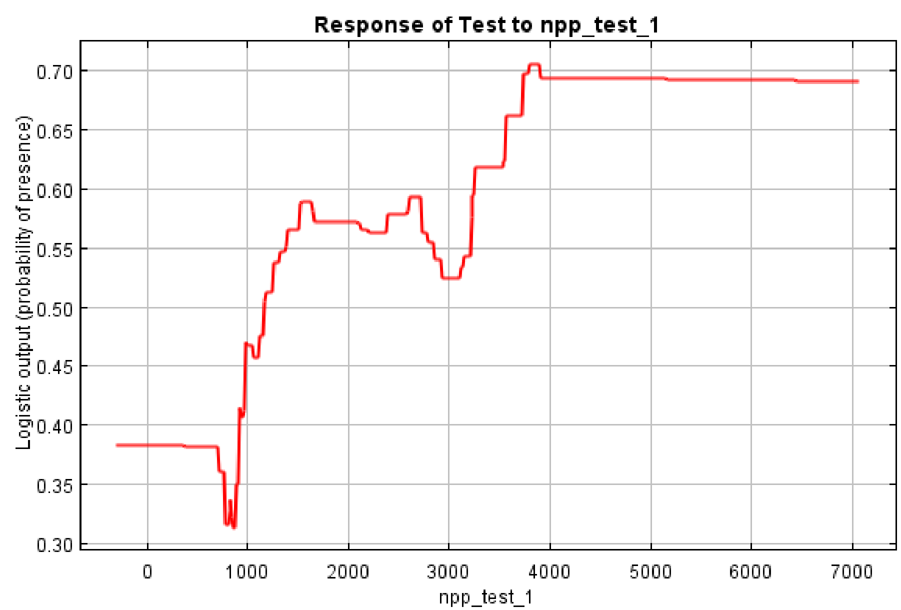


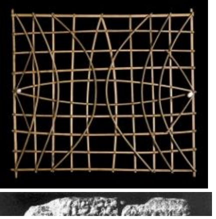
Regularization=0.1



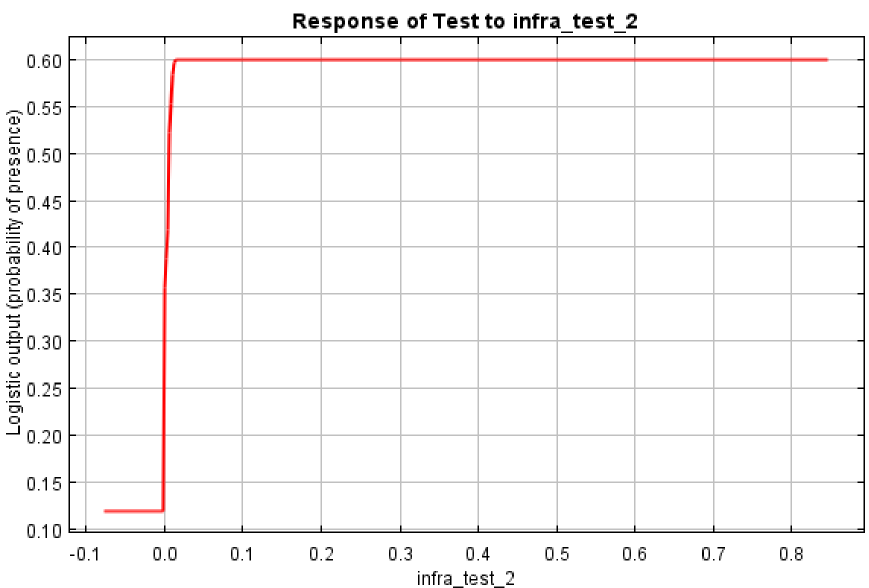


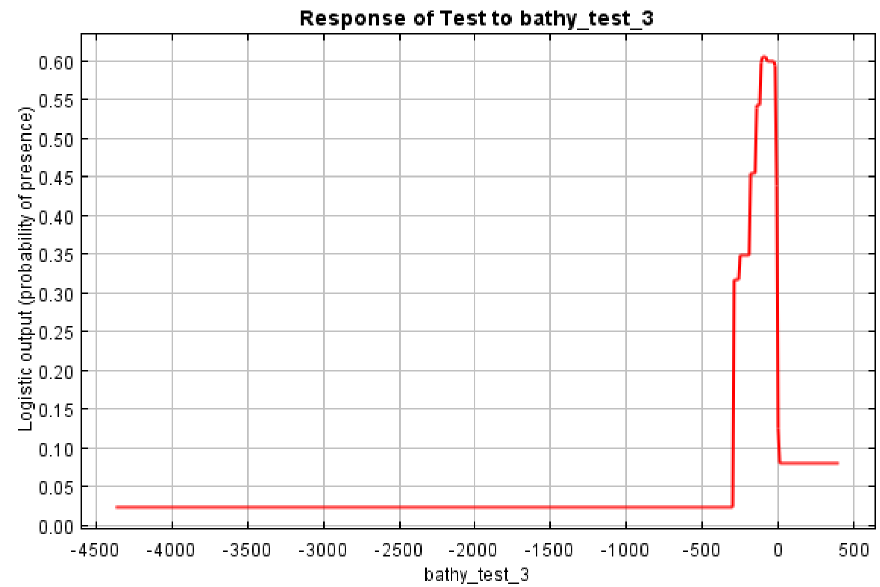


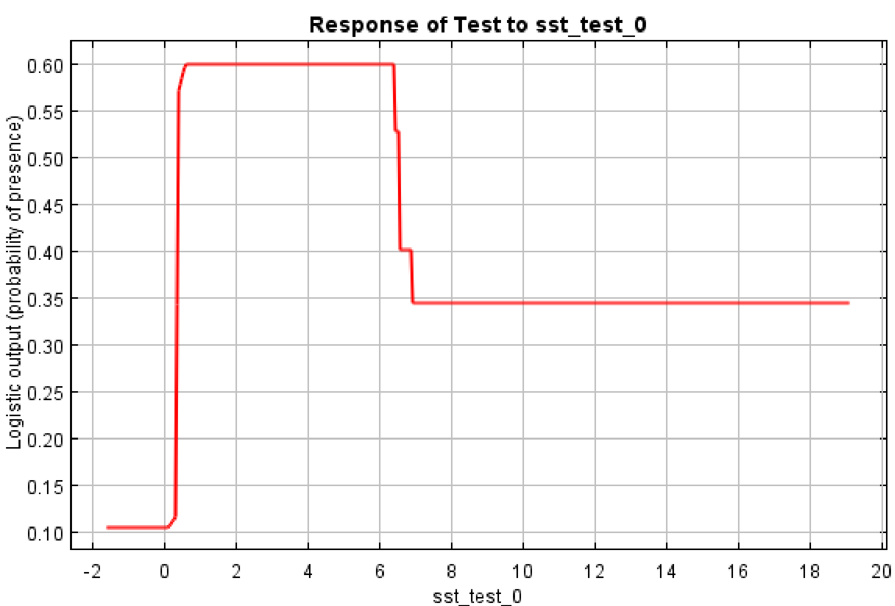


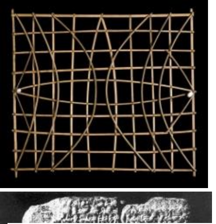


Regularization=1.2

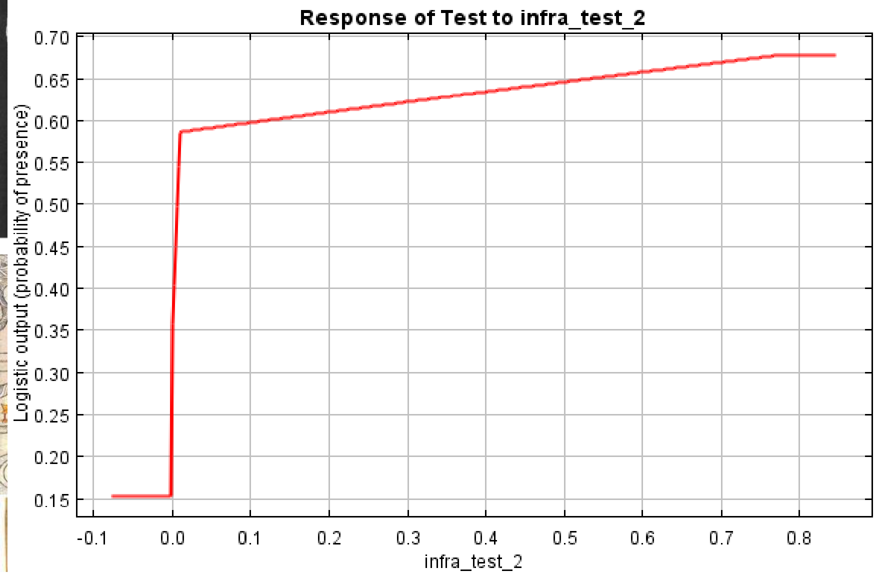


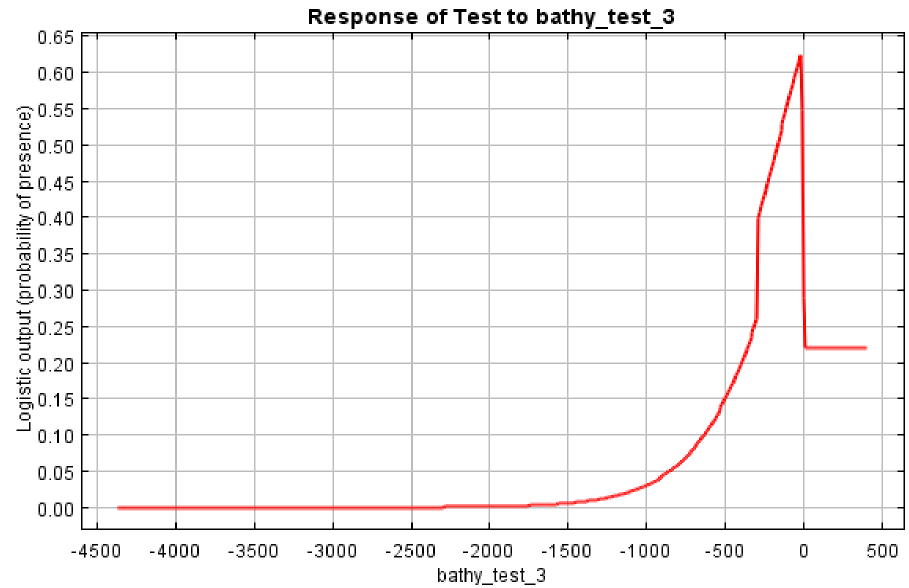


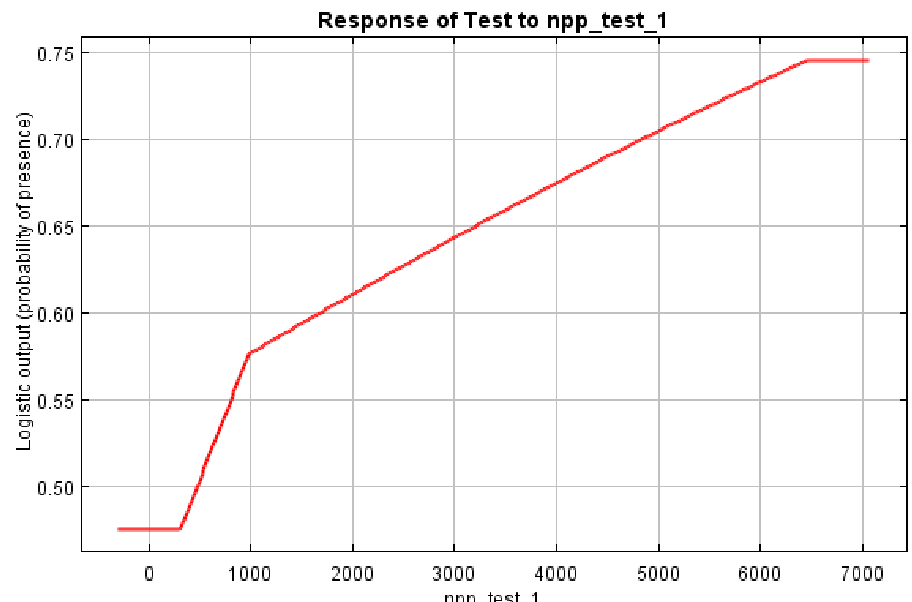


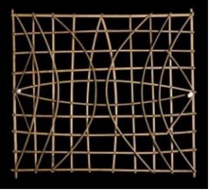


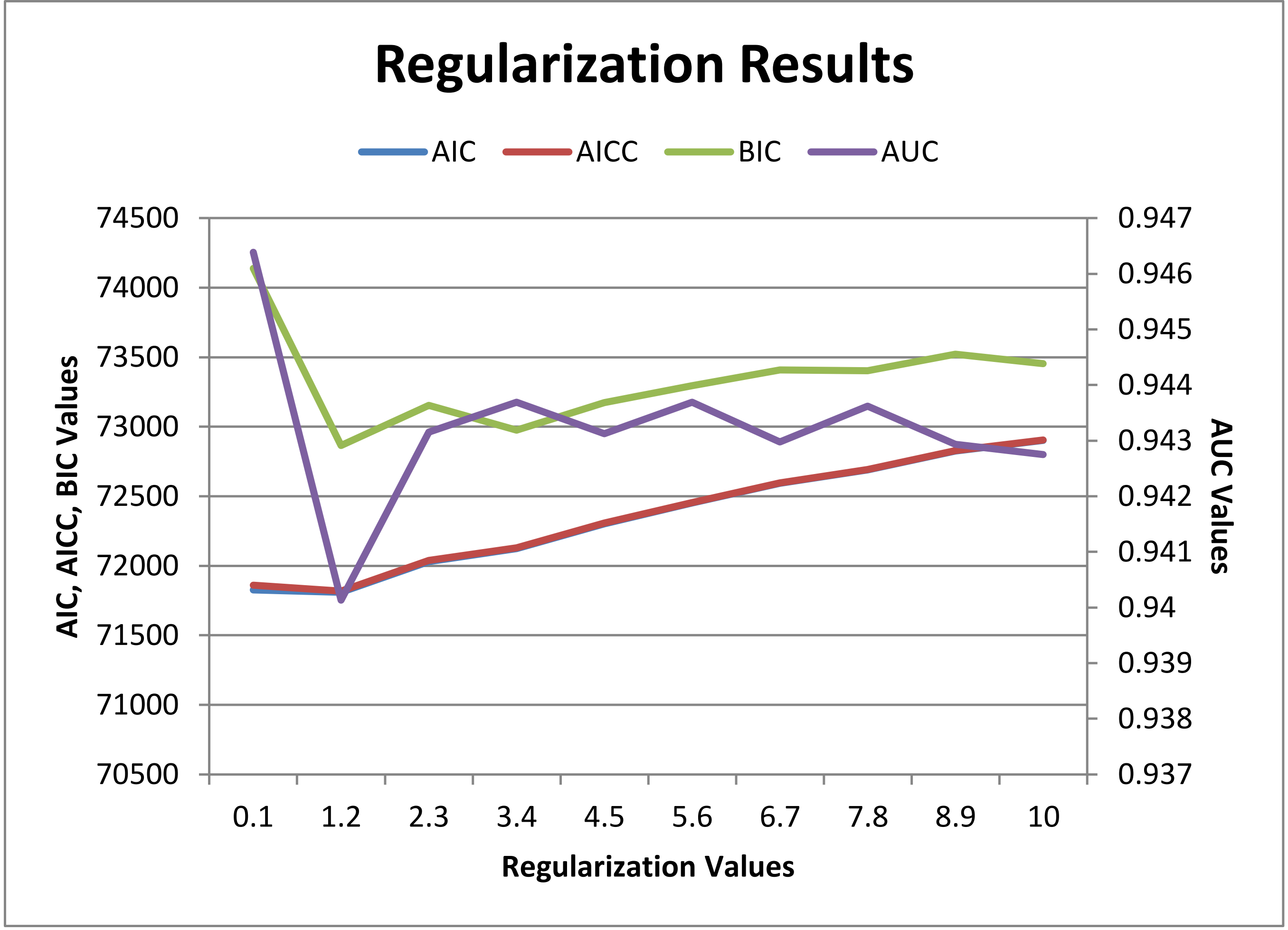
Regularization=10

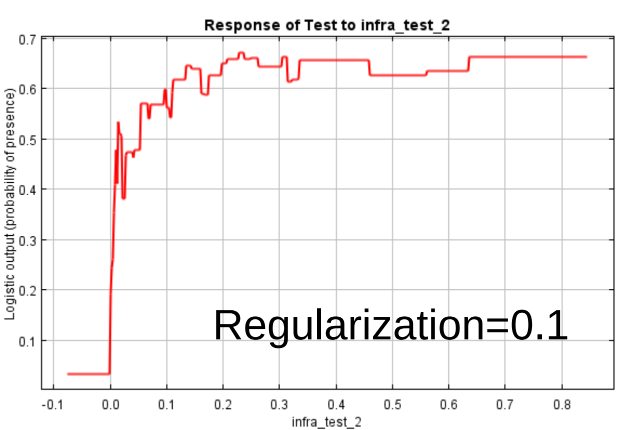


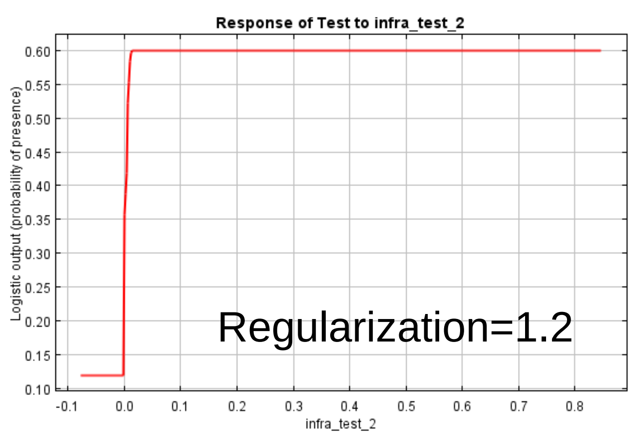


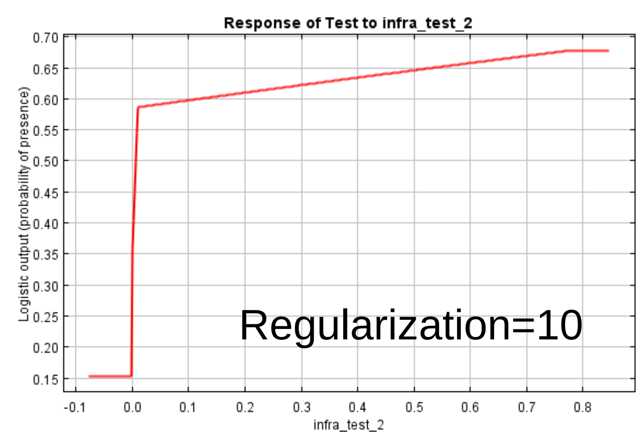


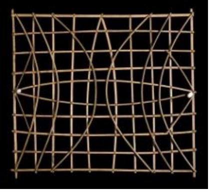



Regularization Results
AIC
AICC
BIC
AUC
74500
74000
73500
73000
72500
72000
71500
71000
70500
AIC, AICC, BIC Values
0.947
0.946
0.945
0.944
0.943
0.942
0.941
0.94
0.939
0.938
0.937
AUC Values
0.1
1.2
2.3
3.4
4.5
5.6
6.7
7.8
8.9
10
Regularization Values


Response of Test to infra_test_2
Logistic output (probability of presence)
infra_test_2
Regularization=0.1

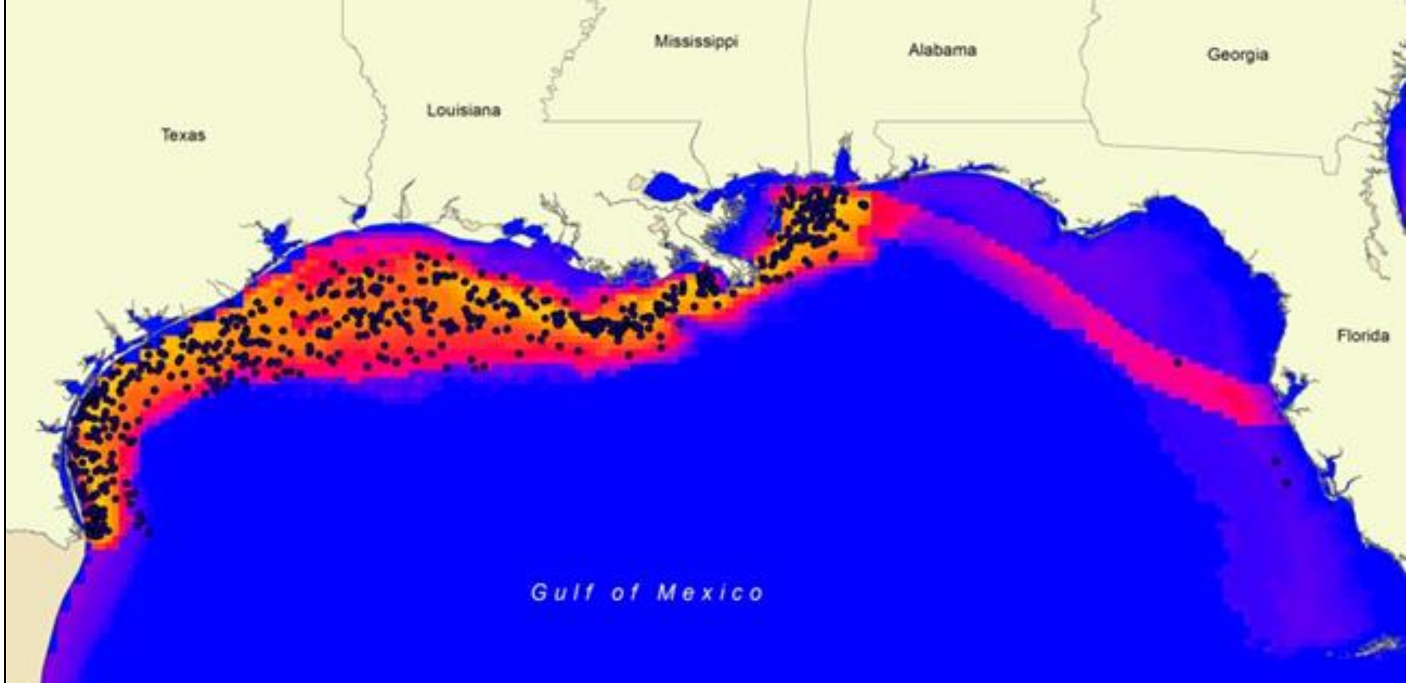

Response of Test to infra_test_2
Logistic output (probability of presence)
infra_test_2
Regularization=1.2


Response of Test to infra_test_2
Logistic output (probability of presence)
infra_test_2
Regularization=10

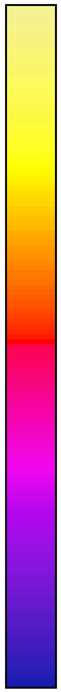




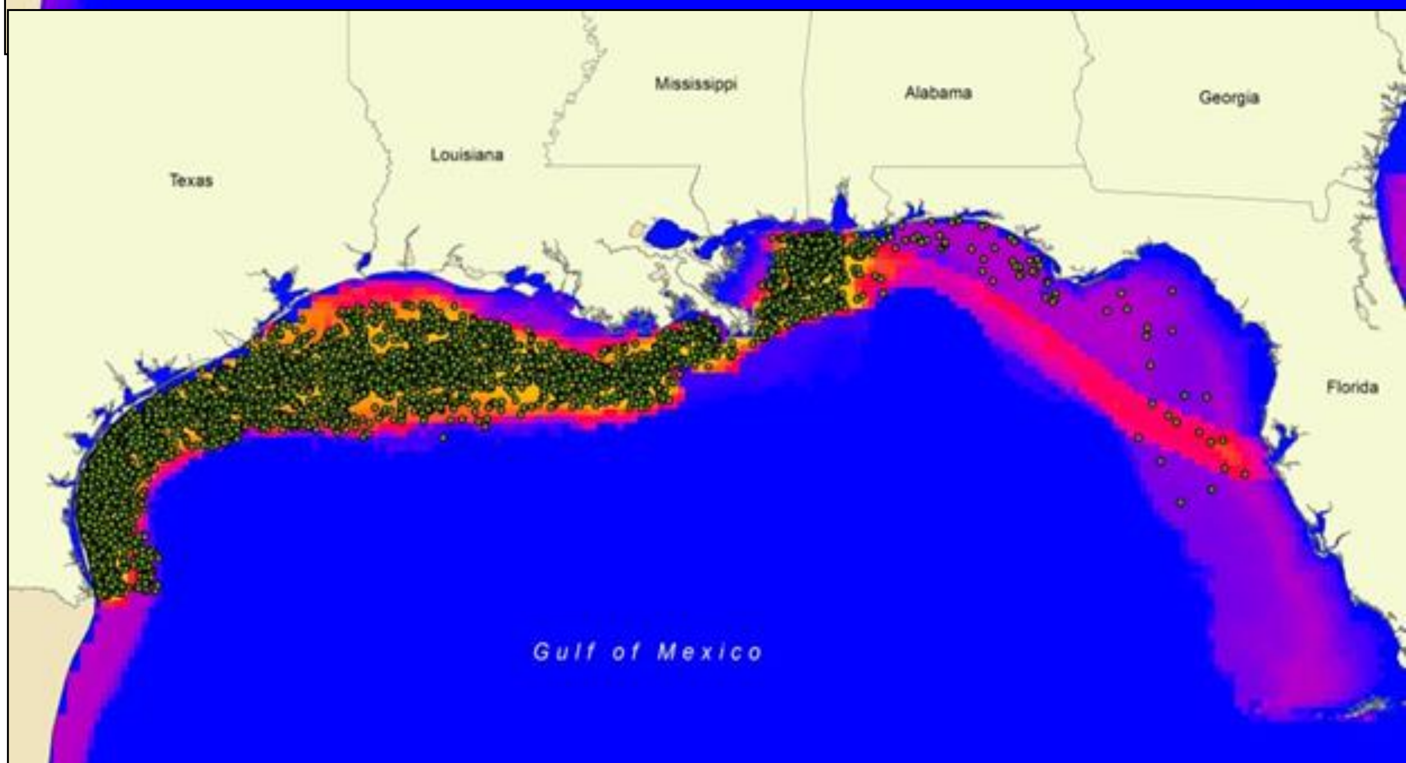

10% Sample Points



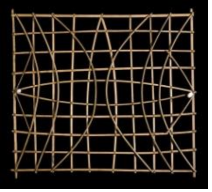
Optimal Habitat



Poor Habitat



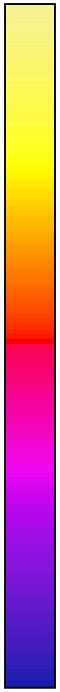
100% Sample Points





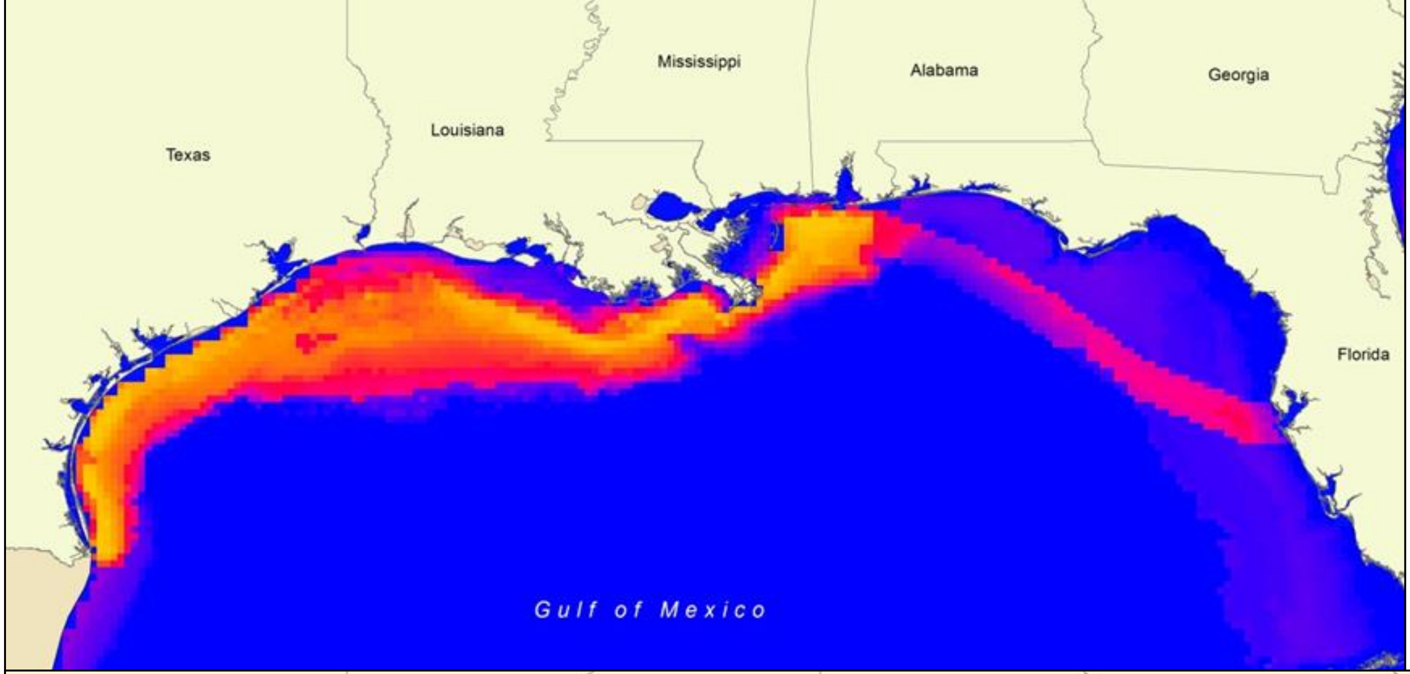

10% Sample Points

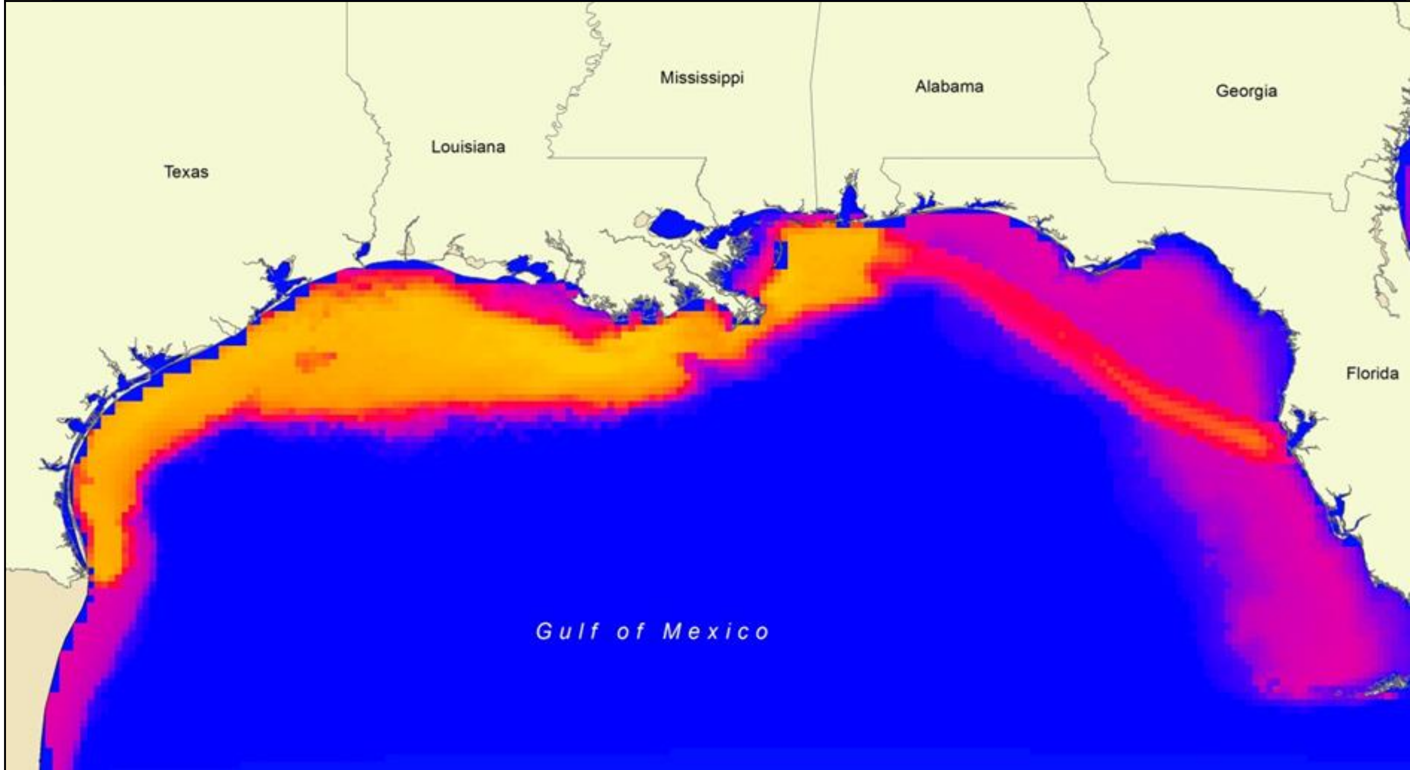
Optimal Habitat

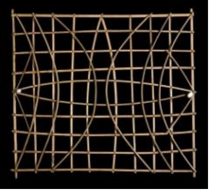


Poor Habitat

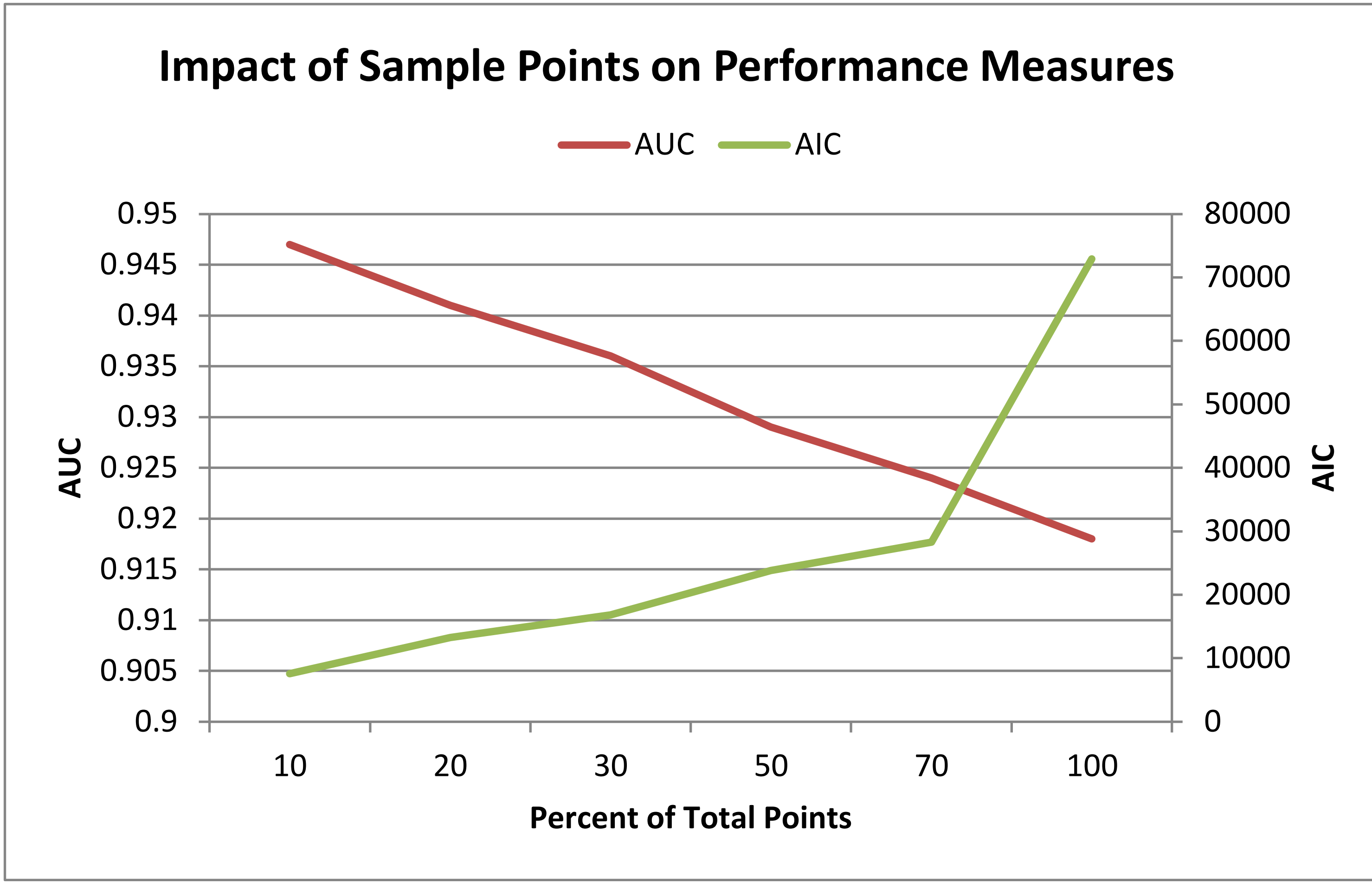
100% Sample Points







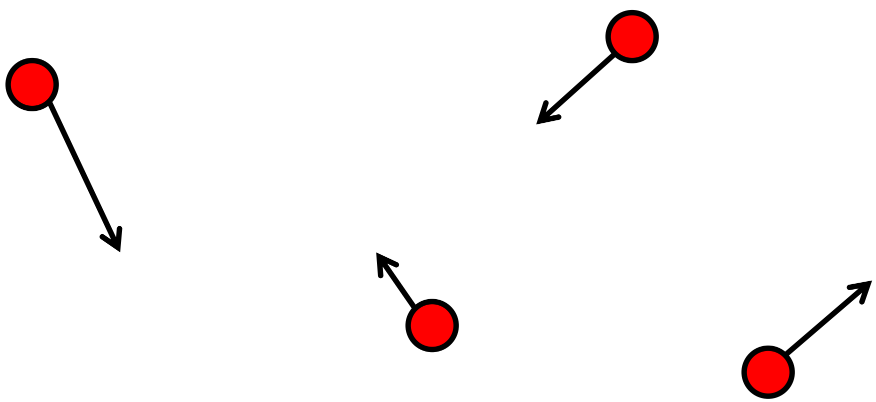


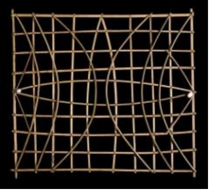



Percent	Number of Samples	Number of Parameters	AIC	AUC
10%	665	136	7549	0.947
20%	1330	166	13,258	0.941
30%	1995	120	16,850	0.936
50%	3326	159	23,839	0.929
70%	4656	162	28,293	0.924
100%	6651	81	72,901	0.918

Jiggling The Samples

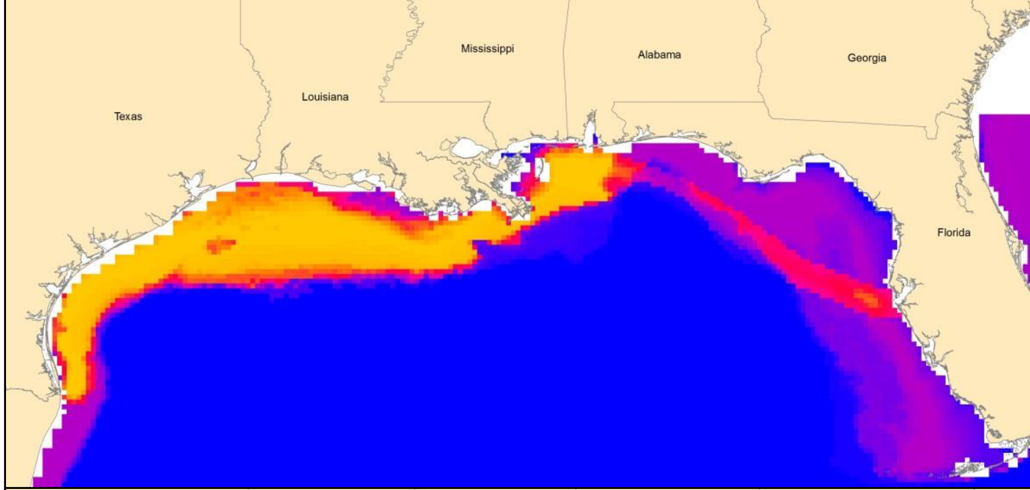
- Randomly shifting the position of the points based on a given standard deviation based on sample uncertainty
- Running the model repeatedly to see the potential effect of the uncertainty



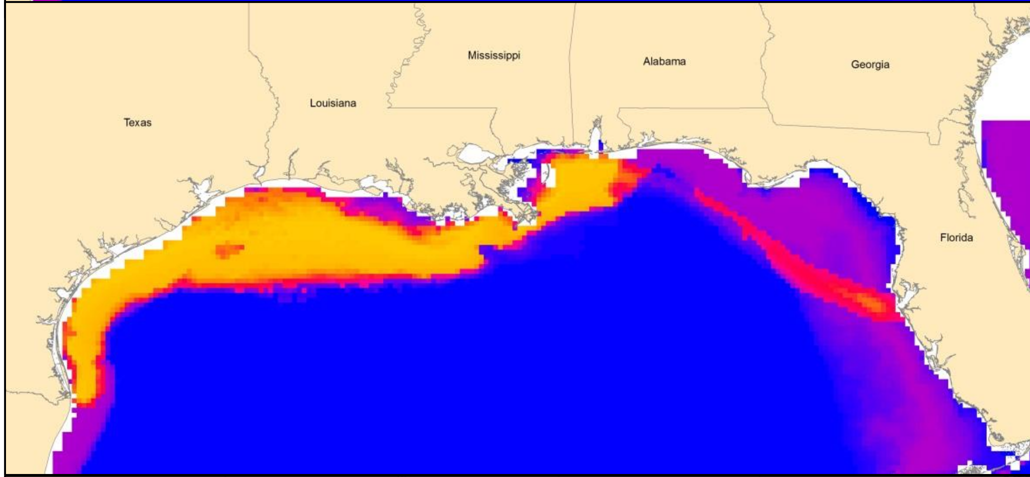


Jiggling

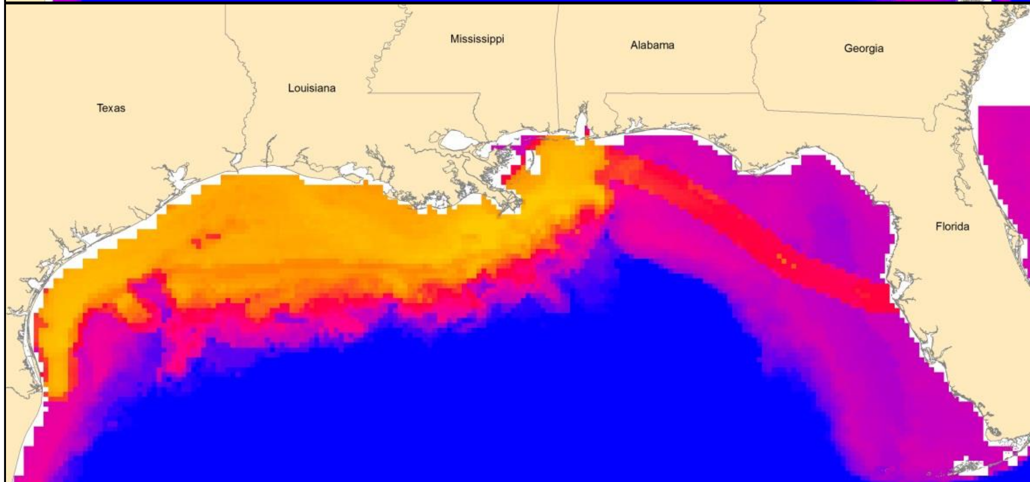
No Jiggling



Std Dev=4.4km

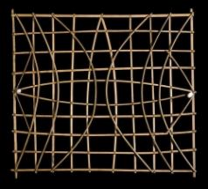


Std Dev=55km

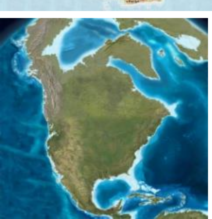


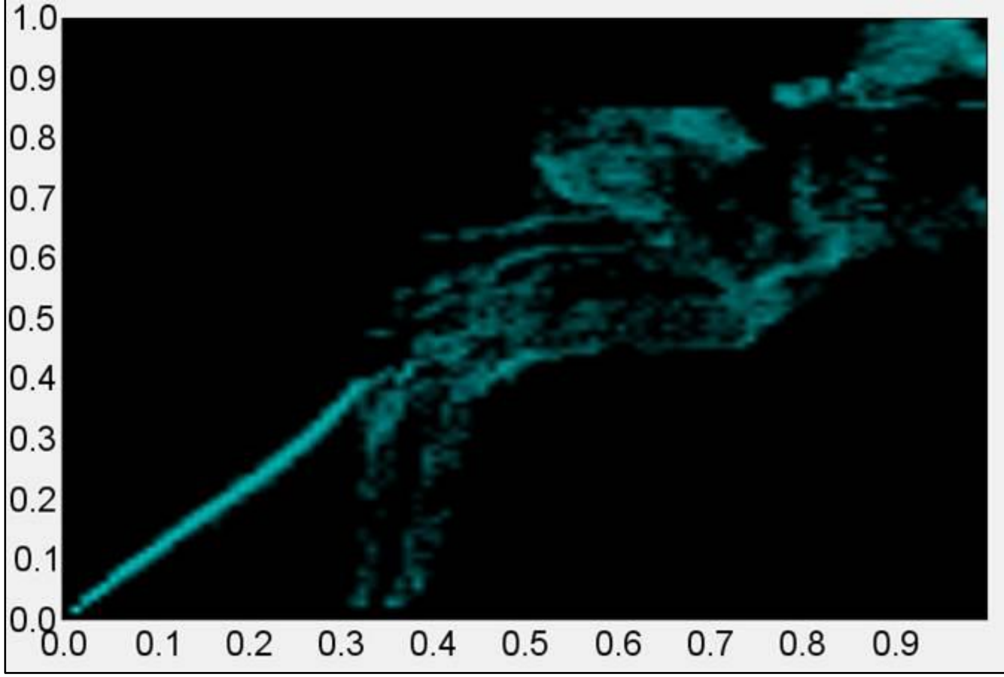
Jiggling

- Spearman's Correlation



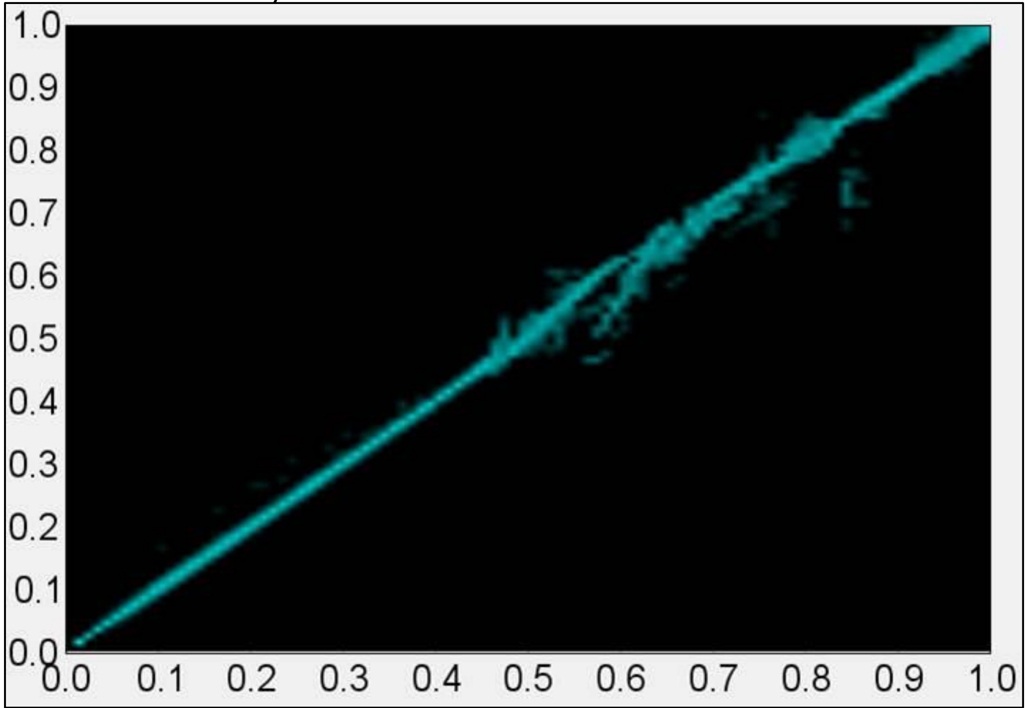


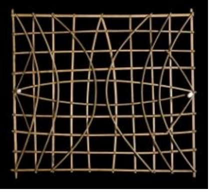





0 vs. 55km, Correlation = .919

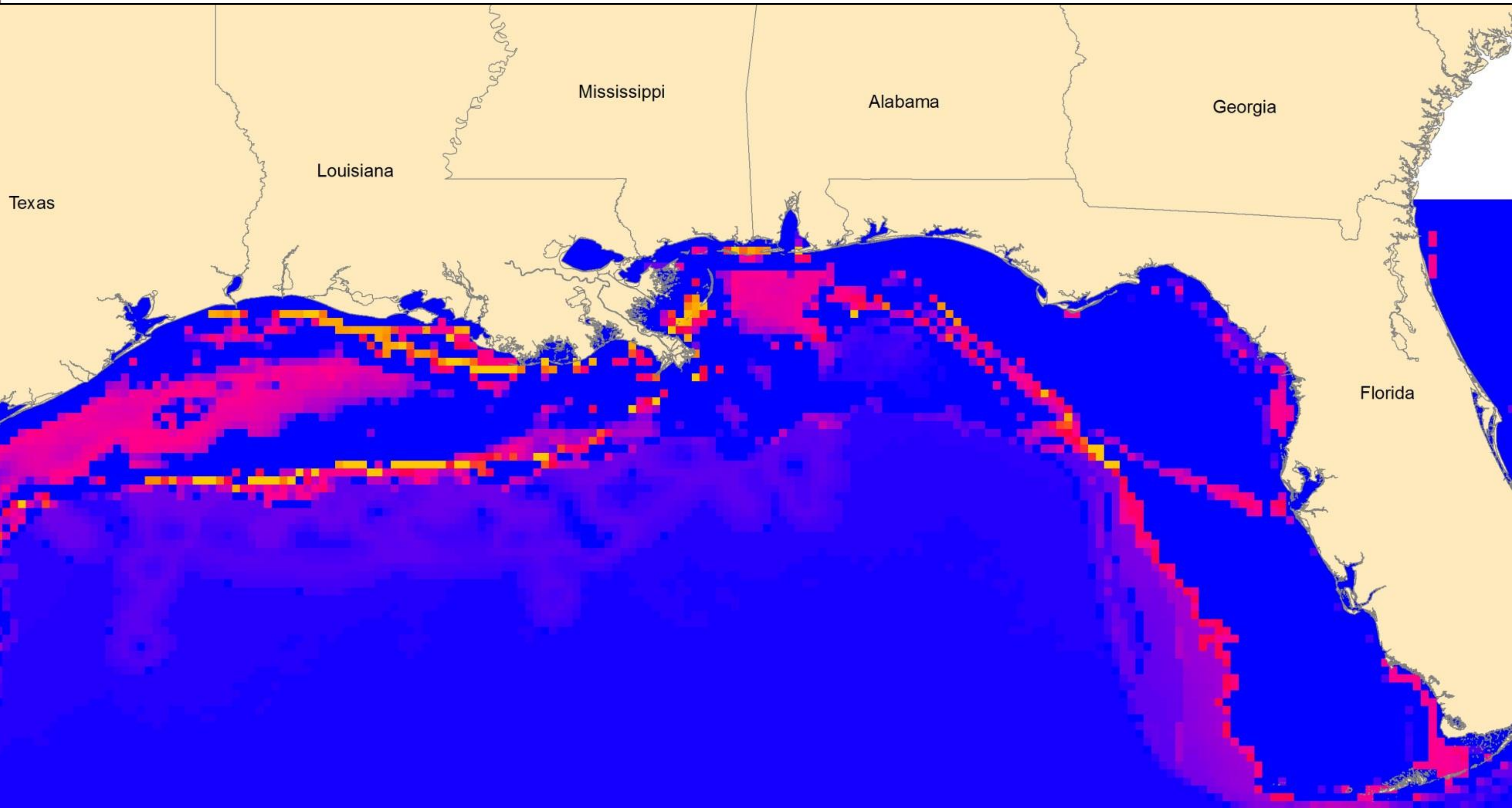
0 vs. 4.4km, Correlation = .997

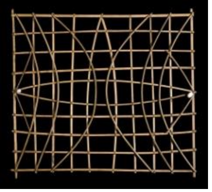




Uncertainty Maps

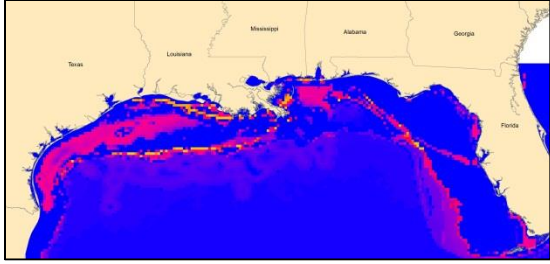
- Standard Deviation of Jiggling Points by 4.4km



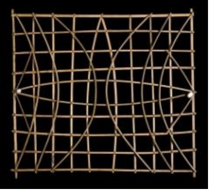





Conclusion



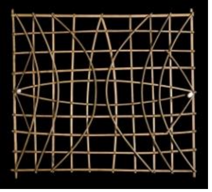
- We can improve HSMs over the default settings in Maxent
- We can determine uncertainty in some cases
- Next Steps
  - Support uncertainty for each data point
  - Improve uncertainty analysis and visualization for predictor layers
  - Begin adding uncertainty analysis and maps to products






Acknowledgements

- The following individuals helped in the development of these ideas:
  - Greg Newman, Catherine Jarnevich, Thomas Stohlgren, Paul Evangelista
- Colorado State University, Oregon State University, SeaMap, and SchoonerTurtles provided resources for the study

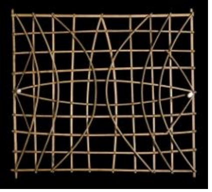




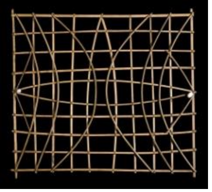

References

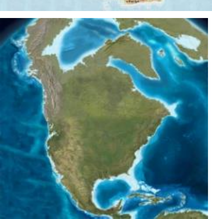
- Anderson, Model Based Inference in the Life Sciences: A Primer on Evidence
- AVHRR Technical Specifications: http://data.nodc.noaa.gov/pathfinder/Version5.2/GDS_TechSpecs_v2.0.pdf
- Chil's JP, Delfiner P , Geostatistics: Modeling Spatial Uncertainty
- Franklin J, Mapping Species Distributions: Spatial 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

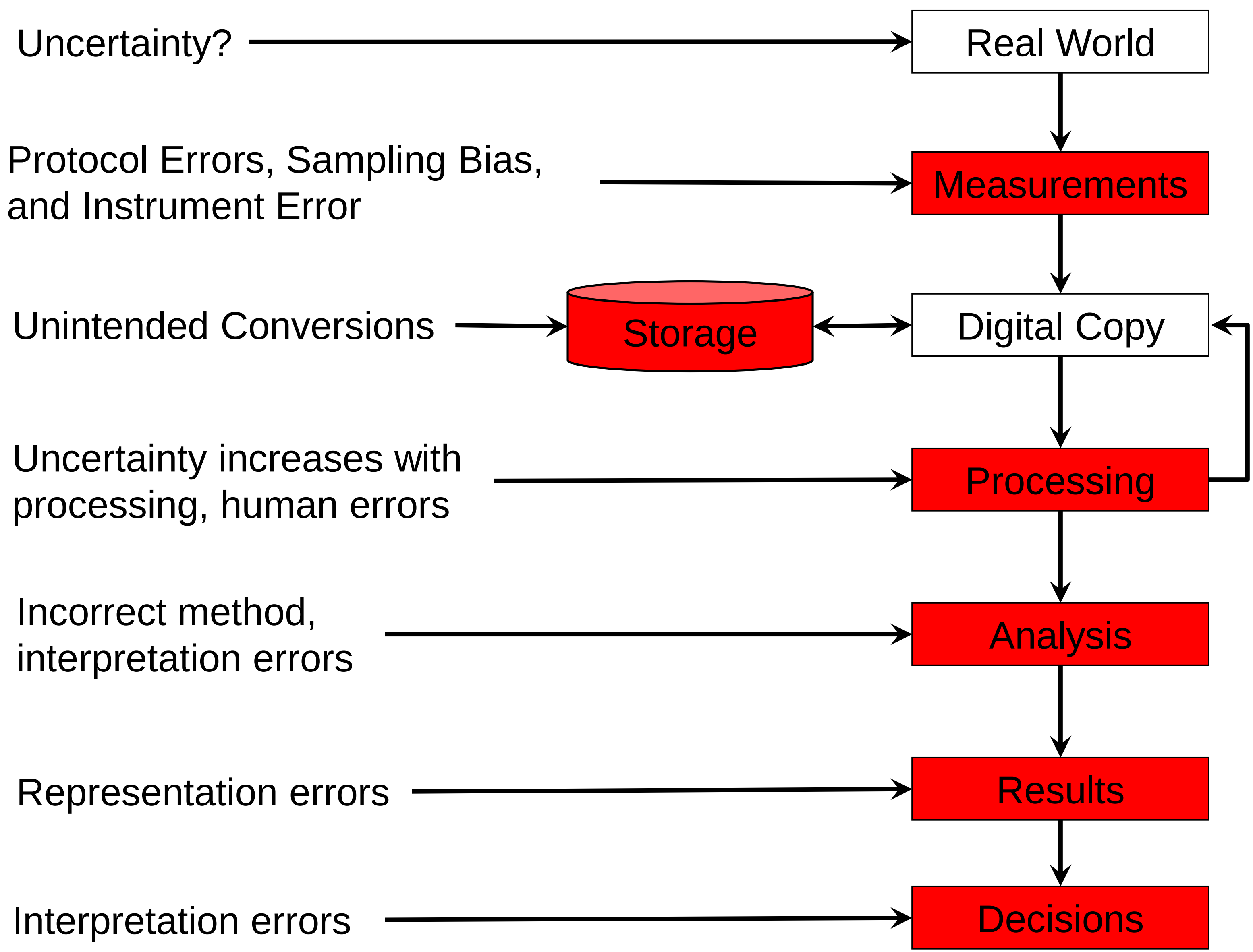
Additional Slides



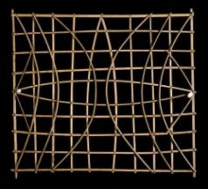
Sources of Uncertainty



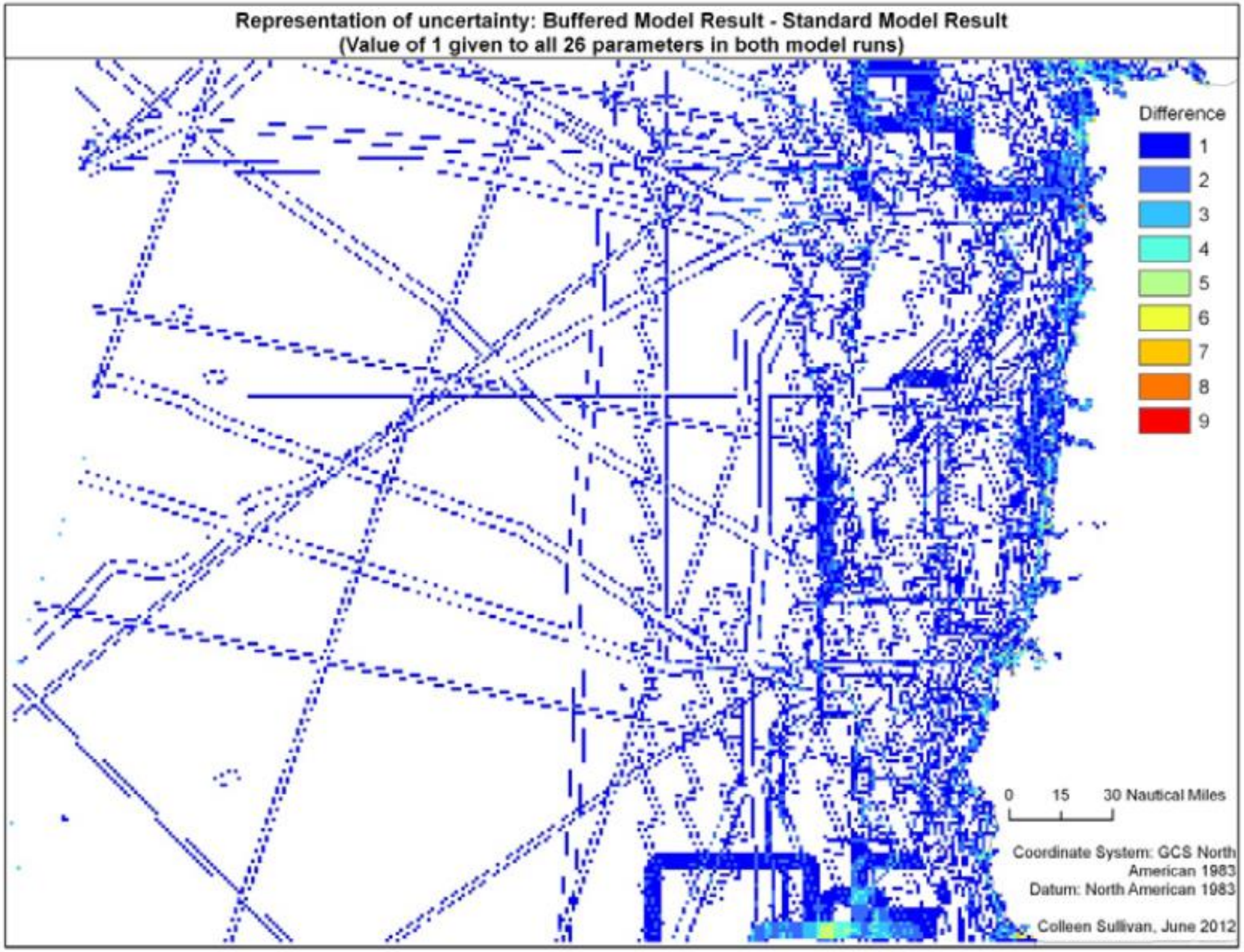




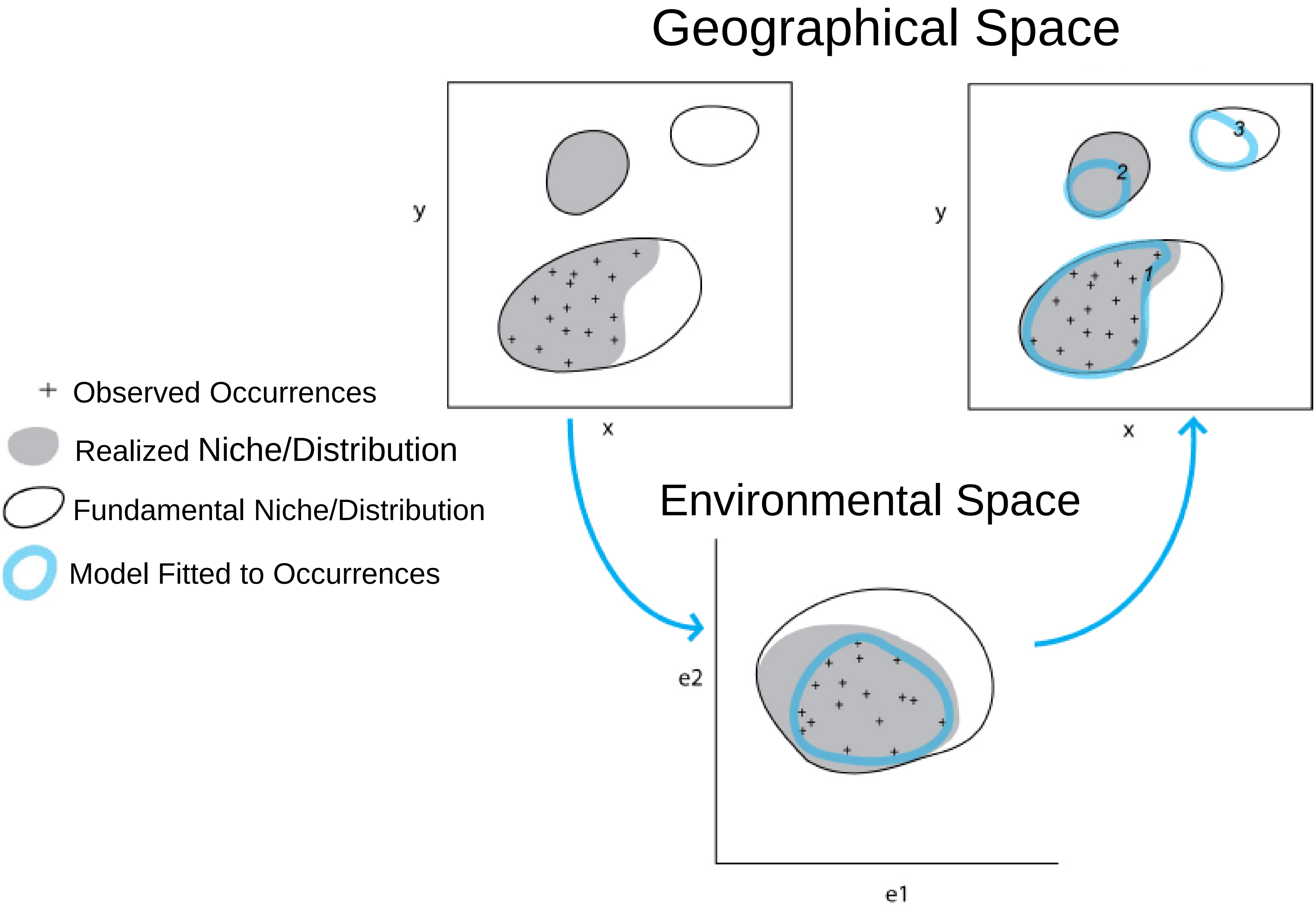
Communicating Uncertainy





Colleen Sullivan, 2012



Adapted from Richard Pearson, Center for Biodiversity and Conservation at the American Museum of Natural History

NIISS - Map

www.niiss.org/cwis438/Browse/TiledMap/Scene_Basic.php?WebSiteID=1

state plane

Welcome guest Login | My Profile | January 8th 2

niiss The National Institute of Invasive Species Science

- Home
- About Us

Gather Data
- Field Methods
- Field Tools

Browse Data
- By Location
- By Species
- By Project
- By Map

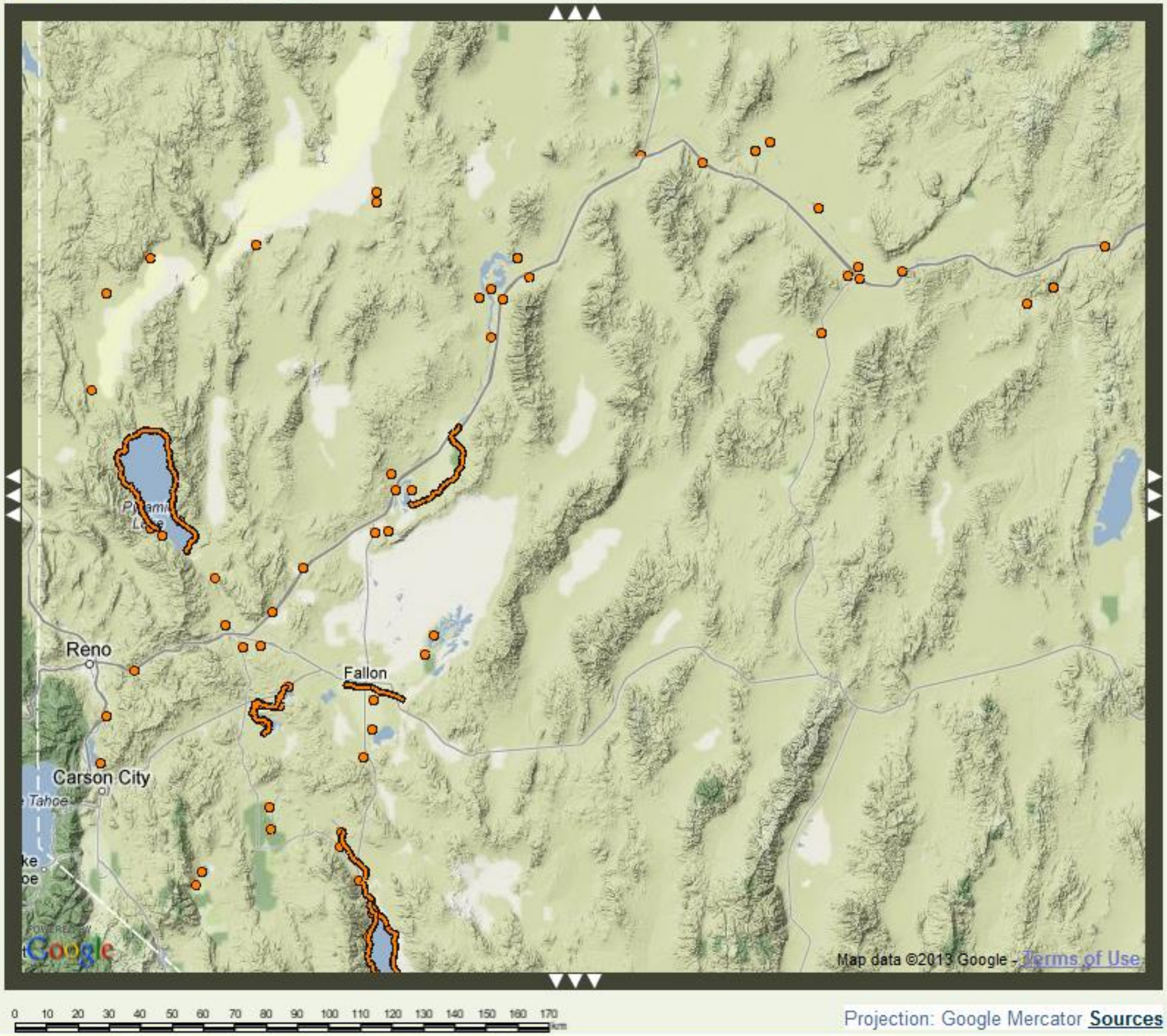
Contribute Data
- Survey Data
- New Sightings
- Data Standards
- GeoRasters

Download Data
- Downloads
- GeoRasters

- Contact Us
- Help
- Register
- Login

Navigation

Current Project: **NIISS - General User** **Help**



Projection: Google Mercator **Sources**

Location

Legend

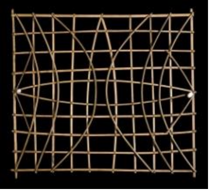
Plants
- [x] Tamarisk

Backgrounds
- (•) Google: Terrain
- ( ) Google: Map
- ( ) Google: Satellite
- ( ) Google: Hybrid

An IBIS website

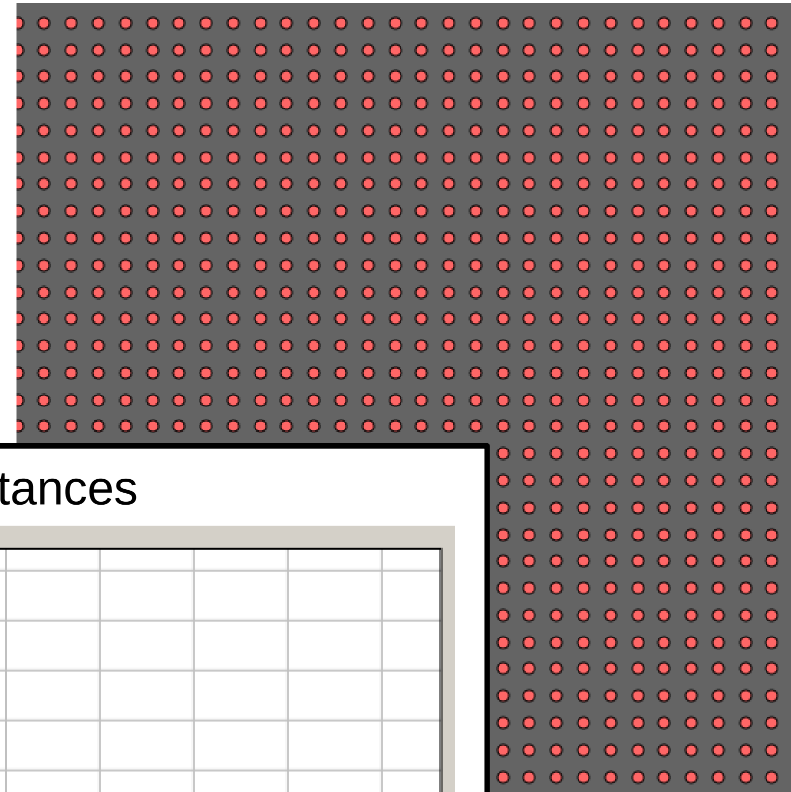
Updated 12/31/20

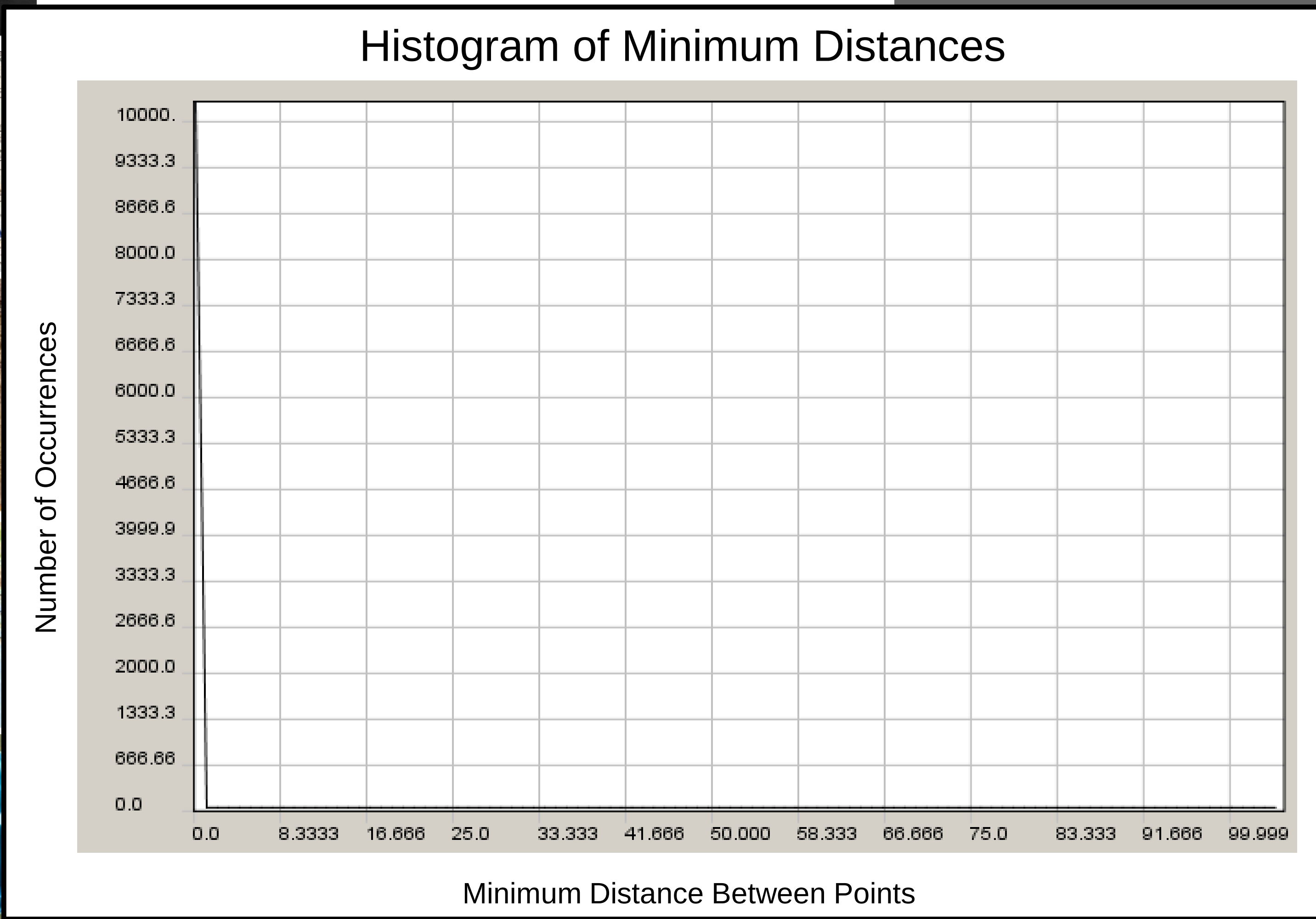
Uniform Data



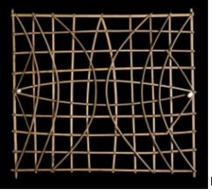


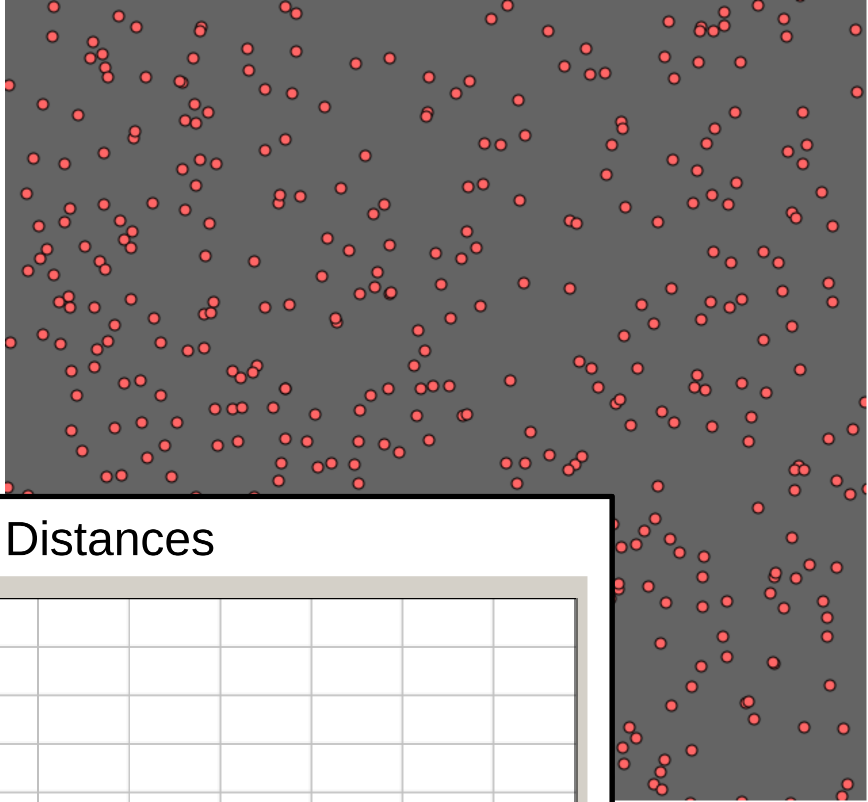


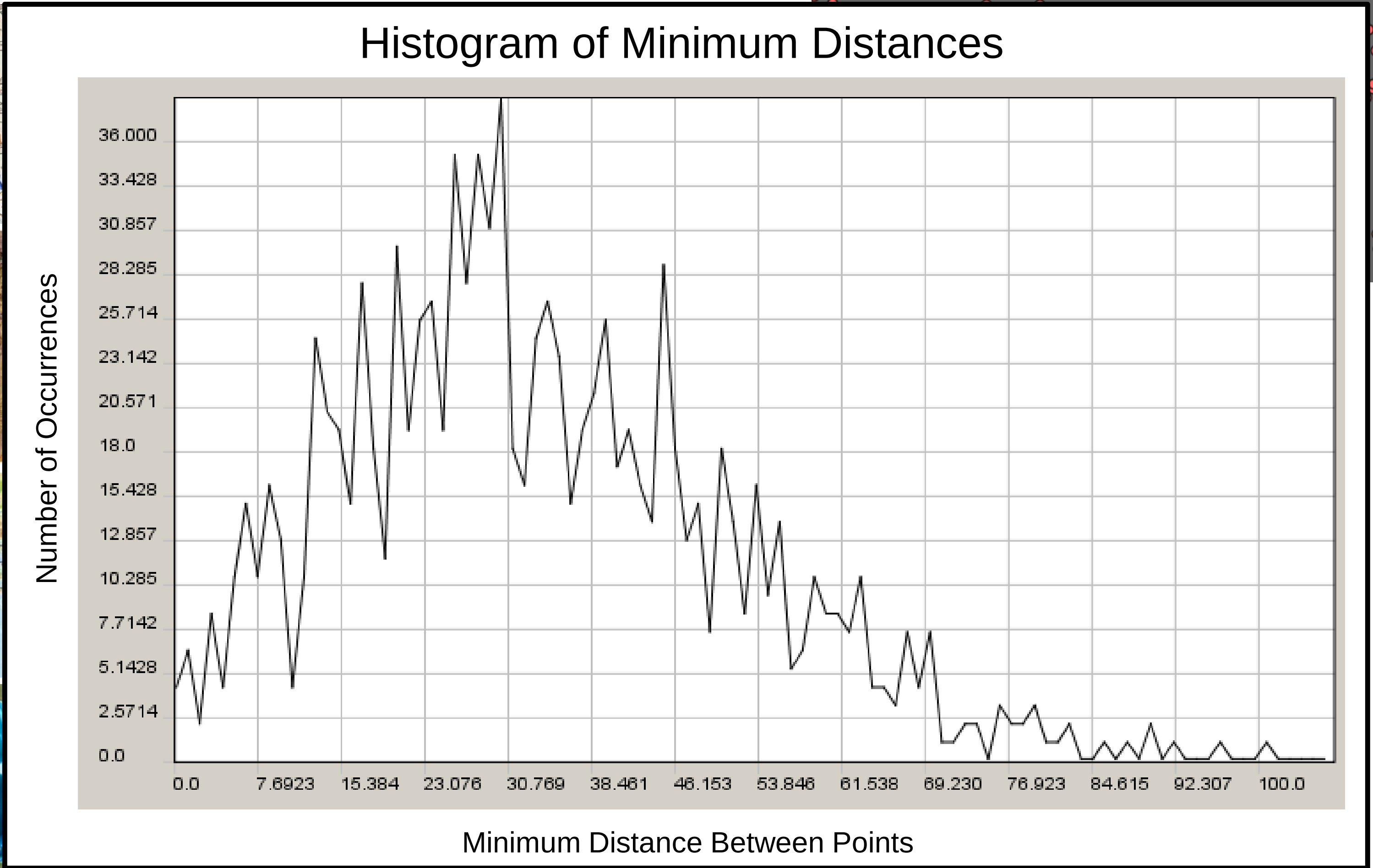




“Random” Data

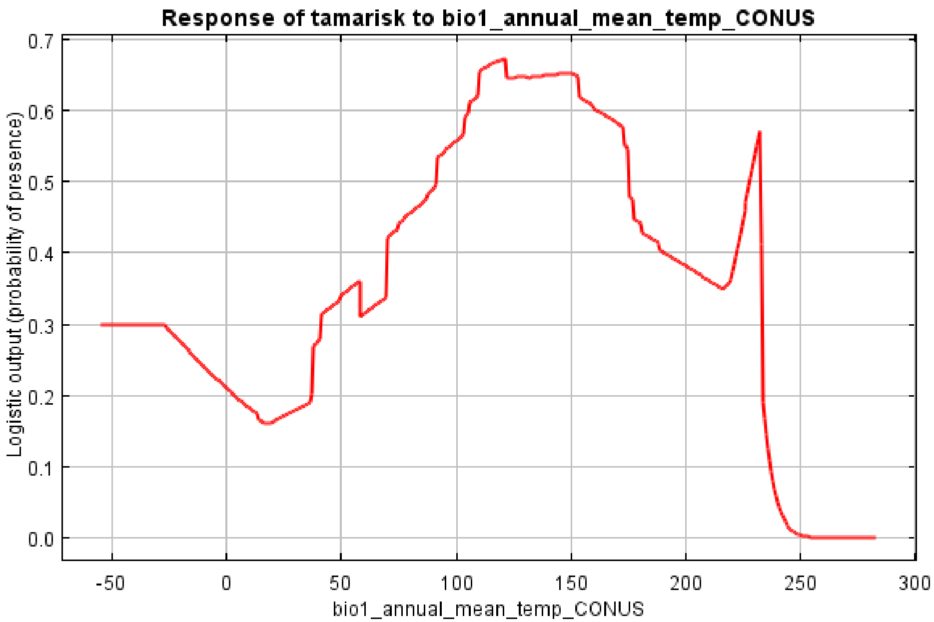






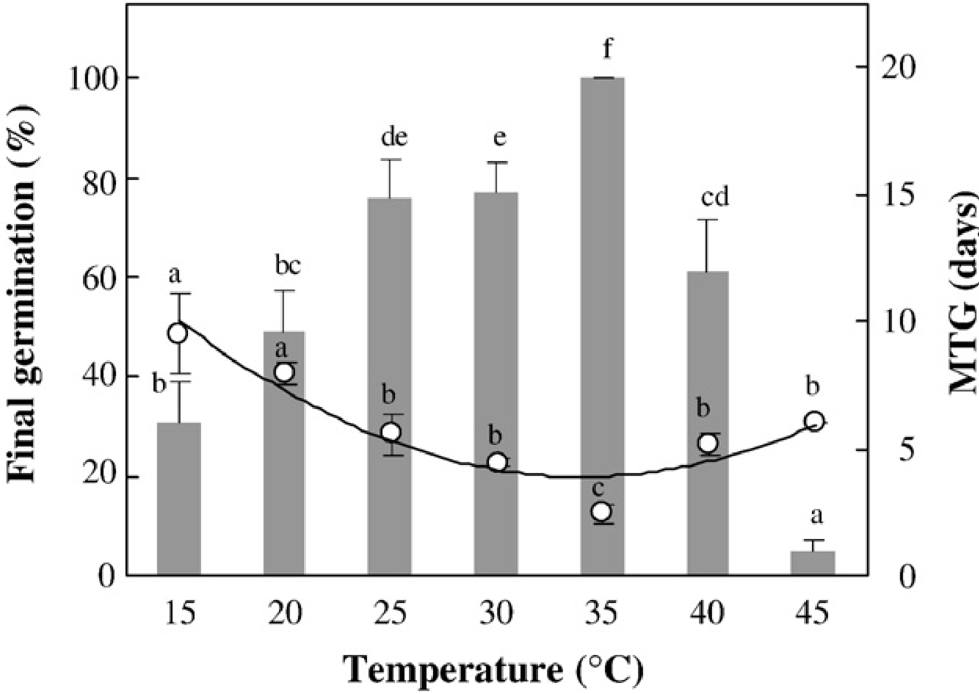
Histogram of Minimum Distances
Number of Occurrences
36.000
33.428
30.857
28.285
25.714
23.142
20.571
18.0
15.428
12.857
10.285
7.7142
5.1428
2.5714
0.0
0.0
7.6923
15.384
23.076
30.769
38.461
46.153
53.846
61.538
69.230
76.923
84.615
92.307
100.0
Minimum Distance Between Points

Over-fitting The Data?



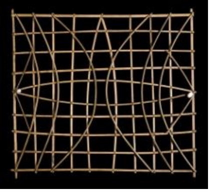
Maxent model for *Tamarix* in the US: response to temperature when modeled with temperature and precipitation

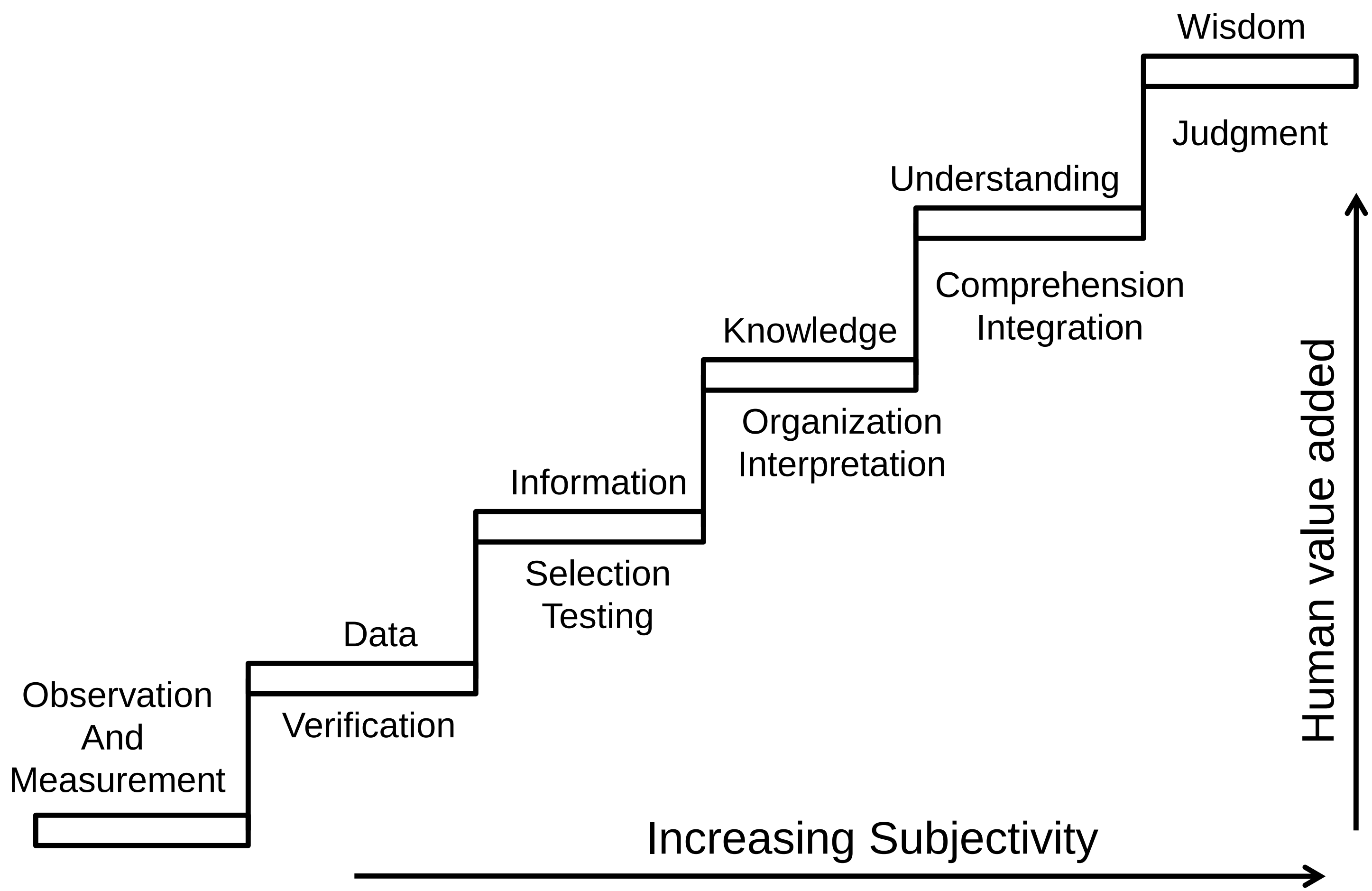
What should the model look like?



Maraghni, M., M. Gorai, and M. Neffati. 2010. Seed germination at different temperatures and water stress levels, and seedling emergence from different depths of Ziziphus lotus. South African Journal of Botany 76:453-459.

Staircase of Knowledge





Environmental Monitoring and Characterization, Aritola, Pepper, and Brusseau