Mapping the Distribution, Abundance and Risk Assessment of Marine Birds in the Northwest Atlantic Ocean: Phase I, Proof of Concept and Techniques Development

USFWS Northeast Region Science Seminar Series

Webinar – August 8, 2013

Presenters:

Beth Gardner

Andrew Gilbert

Brian Kinlan

Dick Veit

Outline

- **Intro and Overview** - Beth Gardner [NCSU]
- **Data Management** - Andrew Gilbert [BRI]
- **Data Review** - Dick Veit [CUNY]
- **Exposure Modeling** - Brian Kinlan [NOAA/NCCOS]
- **Risk Modeling** - Beth Gardner [NCSU]

Background

- Increased interest in sea bird distributions, habitat relationships, and carcass deposition rates
  - General ecology and management
  - Oil spills
  - Off shore wind farms

Chris Graythen

Background

- Off shore wind power garnering lots of interest
  - Many states have implemented a ~20% renewable energy by ~2020 mandate

- Increased risk due to collisions, anthropogenic activities, and habitat alteration

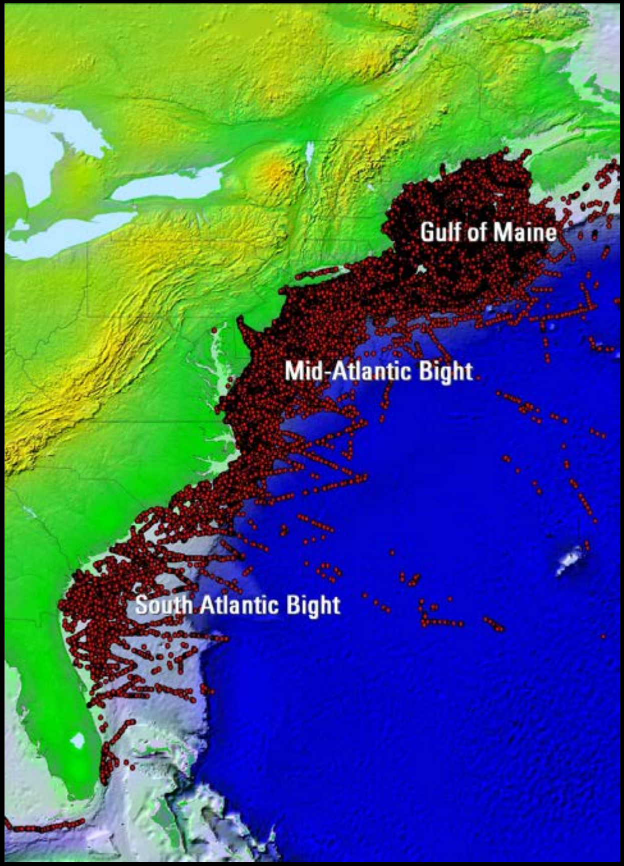
Spatial Risks?

How do we quantify the potential risks?

1. What species of birds are present in the vicinity of a wind farm and how many? (**Exposure**)

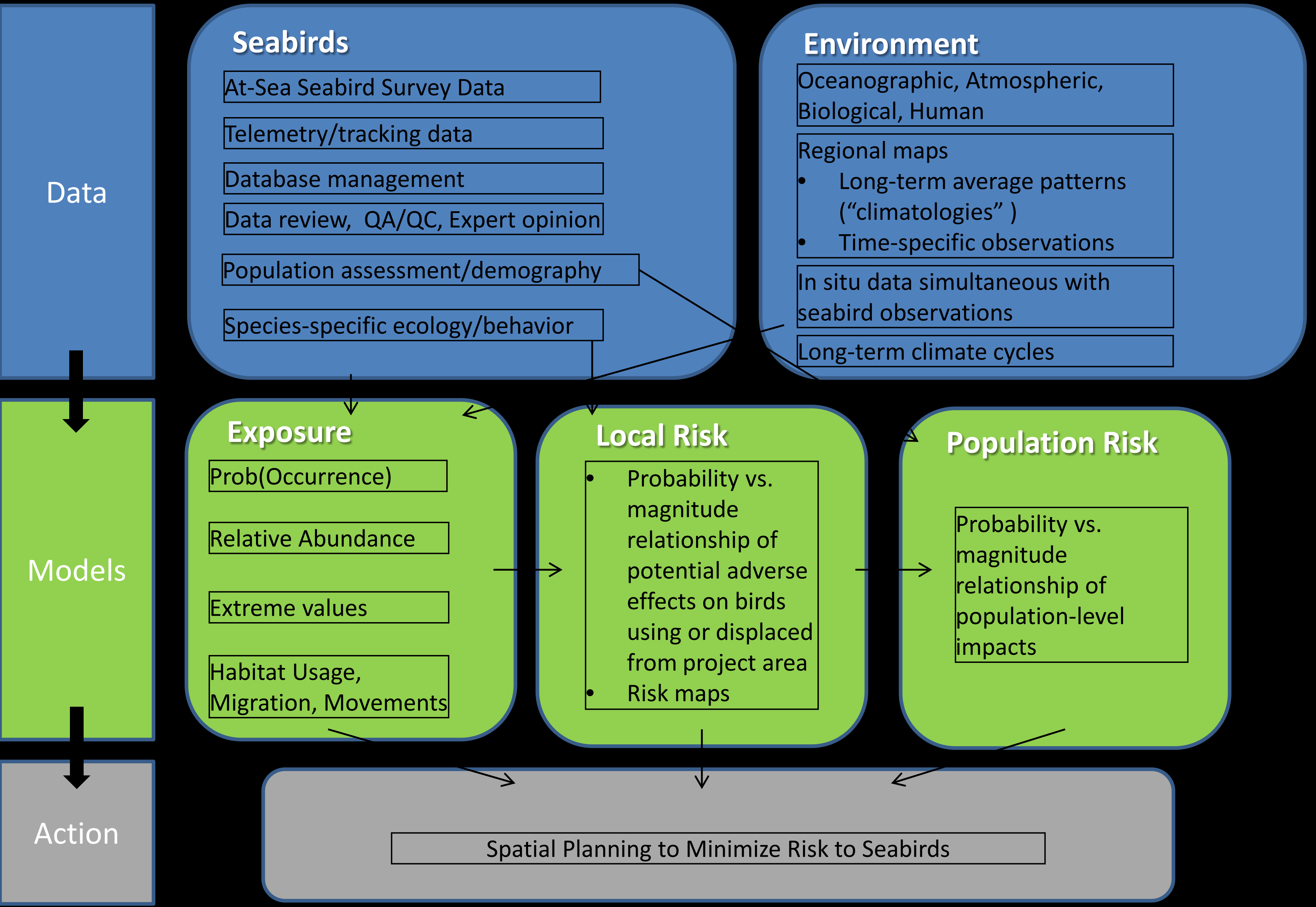
2. What is the per capita probability of an adverse effect of wind farms on birds of a given species, given that they are present in the area? (**Species-Specific Hazard** or **Sensitivity**)

3. How much are the potential adverse impacts from combining (1) and (2) likely to impact the population of each species, given its current status, trends, and ecological traits? (**Population Vulnerability**)



Modified from Crichton (1999) definition

Concept Map – Spatial Planning for Marine Bird Risk Assessment

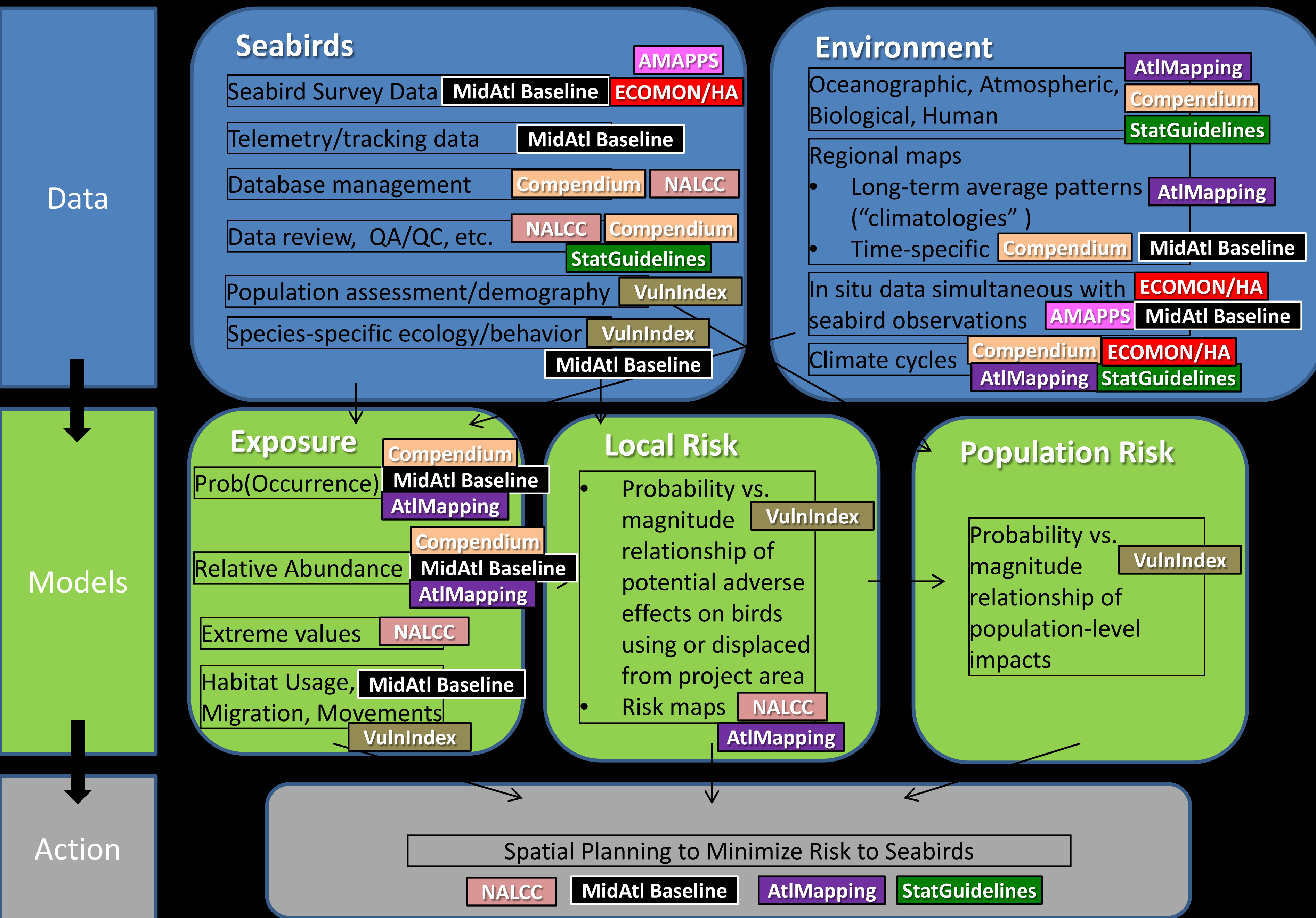


Marine Spatial Planning & Seabirds – Summary of US Atlantic Projects

Abbrev.	Project	Institutions	PI's	Primary Funding	Approx Timeframe
NALCC	Mapping the Distribution, Abundance and Risk Assessment of Marine Birds in the Northwest Atlantic Ocean: Phase I, Proof of Concept and Techniques Development	NCSU, NOAA/NCCOS, BRI, CUNY	Gardner, Kinlan, Gilbert, Veit	NALCC	2012-2013
MidAtl Baseline	Mid-Atlantic Baseline Studies Project	BRI, NCSU, CUNY, Duke, et al.	Williams, Gilbert, Gardner, Veit, et al.	DOE, Maryland DNR	2012-2015
Compendium	Atlantic Seabird Compendium Phases I & II	USGS PWRC, USFWS, NCSU, NOAA/NCCOS, Tufts, BRI, URI	O'Connell, Gilbert, Gardner, Kinlan, Wimer, Ellis, et al.	BOEM	2008-2013
AtlMapping	Atlantic Seabird Mapping and Modeling	NOAA/NCCOS	Kinlan	BOEM	2013-2016
AMAPPS	AMAPPS	NOAA/NMFS, USFWS, US Navy	Garrison, Palka, et al.	BOEM	2010-2013
ECOMON/HA	Ships of Opportunity Seabird Surveys	CUNY, NOAA/NMFS	Veit	NOAA	2006-2014
VulnIndex	Vulnerability Index	Normandeau	Willmot, Forcey, Kent	BOEM	2012-2013
StatGuidelines	Statistical Guidelines for Marine Bird Survey Effort for Hotspot and Coldspot Detection	NOAA/NCCOS	Kinlan, Zipkin, O'Connell	BOEM	2013-2014

NALCC MidAtl Baseline Compendium AMAPPS AtlMapping ECOMON/HA VulnIndex StatGuidelines

Concept Map – Spatial Planning for Marine Bird Risk Assessment



NALCC seabird modeling webinar

Andrew Gilbert - BRI

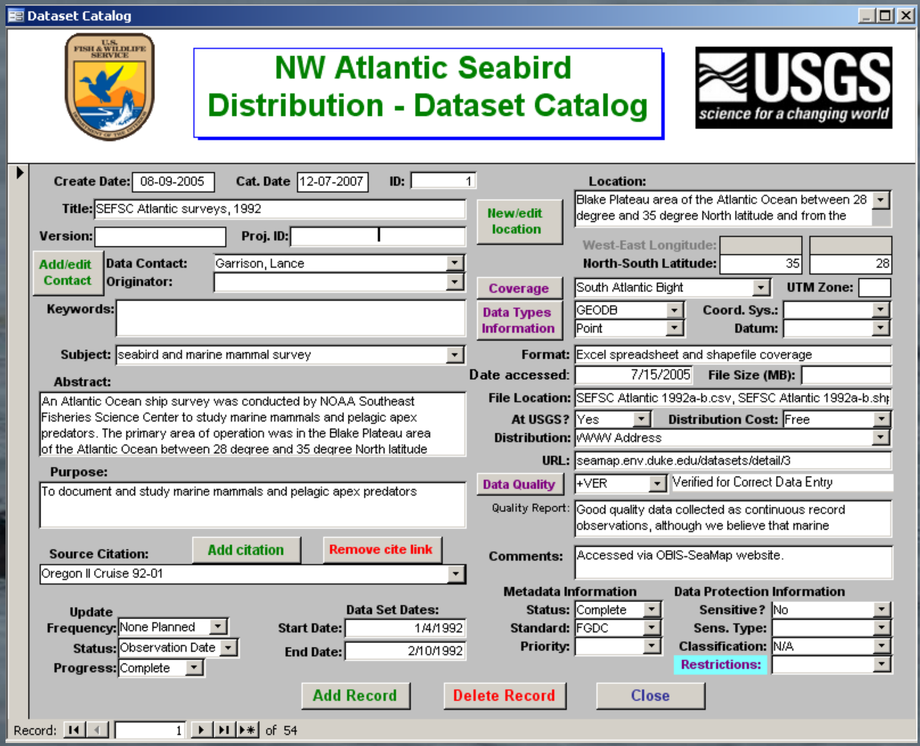
Data assistance for modelers and Atlantic Seabird Compendium

History of Atlantic Seabird Compendium

- Need to evaluate seabird distribution for offshore proposals
- No centralized repository of seabird data for the U.S. Atlantic.
- USFWS funded USGS to catalog seabird datasets in 2005
- USFWS further funded USGS to compile and standardize data into a single database in 2006
- BOEMRE added funds to continue work and add modeling component in 2008-to present
- BDBM and Atlantic Modeling projects use data from db
- USSG, PWRC maintains database

Seabird Dataset Catalog

- Created a catalog of seabird datasets
- Record information about datasets and information they contain (metadata catalog)
  - Coverage area
  - Abstract
  - Dates
  - Data type (digital, analog, text file, GIS)
- Locate data and archive where possible

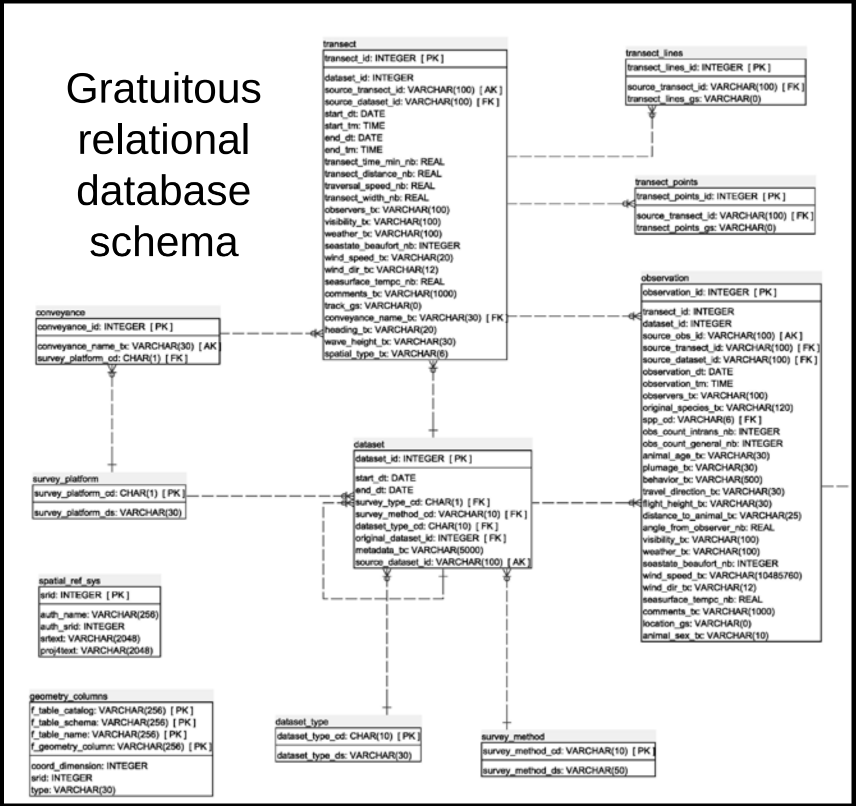


Example seabird surveys

Dataset	Years of surveys	Region of survey
Manomet Center for Conservation Sciences	1978-1980	Gulf of Maine, Mid-Atlantic Bight
Cetacean and Seabird Assessment Program	1980-1988	Gulf of Maine, Mid-Atlantic Bight
Georgia pelagic surveys	1982-1985	South Atlantic Bight
Southeast Fisheries Science Center surveys	1992,1998,1999	South Atlantic Bight
Winter Survey of the Mid-Atlantic	2001-2003	Mid-Atlantic Bight
Cape Wind, Mass Audubon	2002-2006	Nantucket Sound
North Carolina shelf—trophic predators	2004-2005	Offshore North Carolina
Bar Harbor whale watch	2005-2006	Offshore Mount Desert Island, ME
NOAA Ecosystem Monitoring Survey	2007-2012	Gulf of Maine, Mid-Atlantic Bight
NOAA Herring Acoustic Survey	2006-2012	Gulf of Maine, Mid-Atlantic Bight

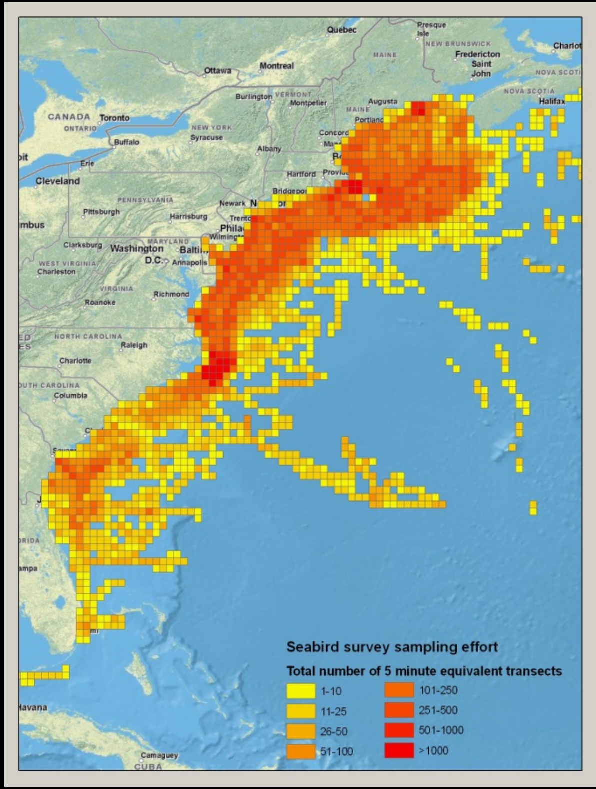
Database compilation

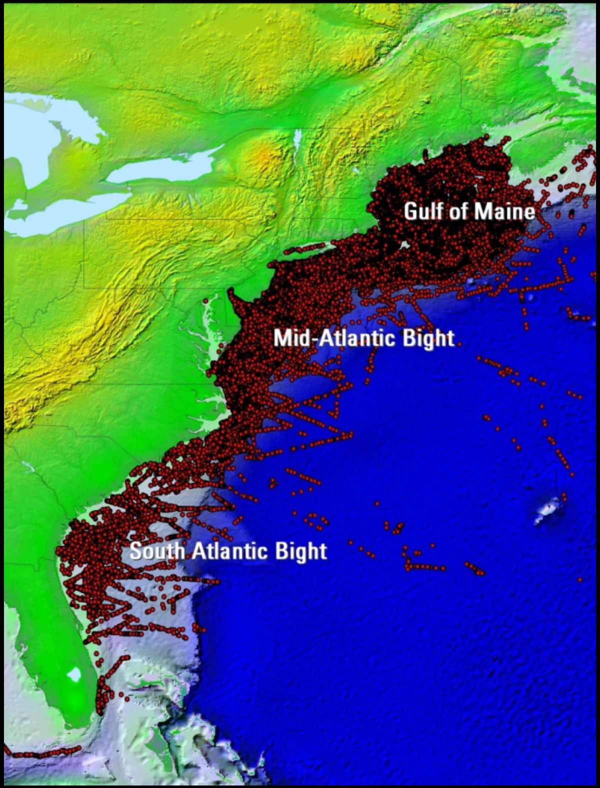
- Create master observation dataset
  - Create standard species lists
  - Create common data fields (date, time, observation ID, effort ID, etc.)
- Create effort dataset when possible and link to species observations
  - presence <u>AND</u> absence data
  - facilitates error detection
- Create Postgresql 8.4 relational database



Occurrence and effort data

- > >270,000 obs (~80) datasets
- >data spans the 1900’s, but most from the 1980s
- Data collected using a mix of scientific and non-scientific methods

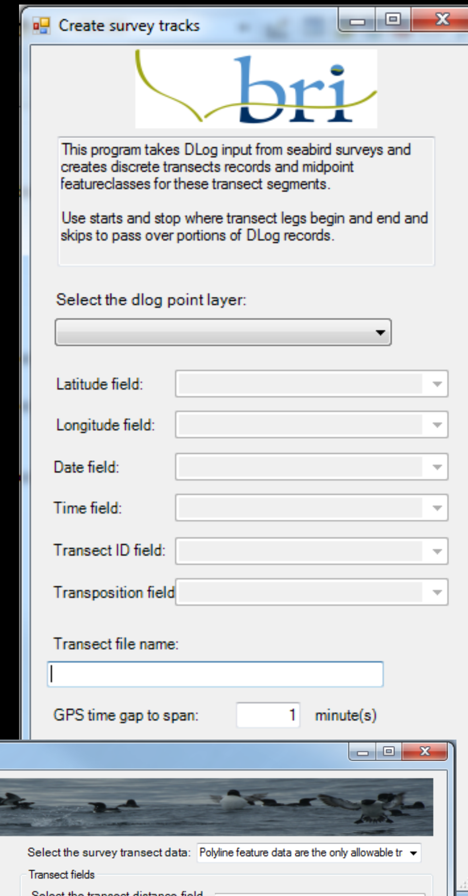


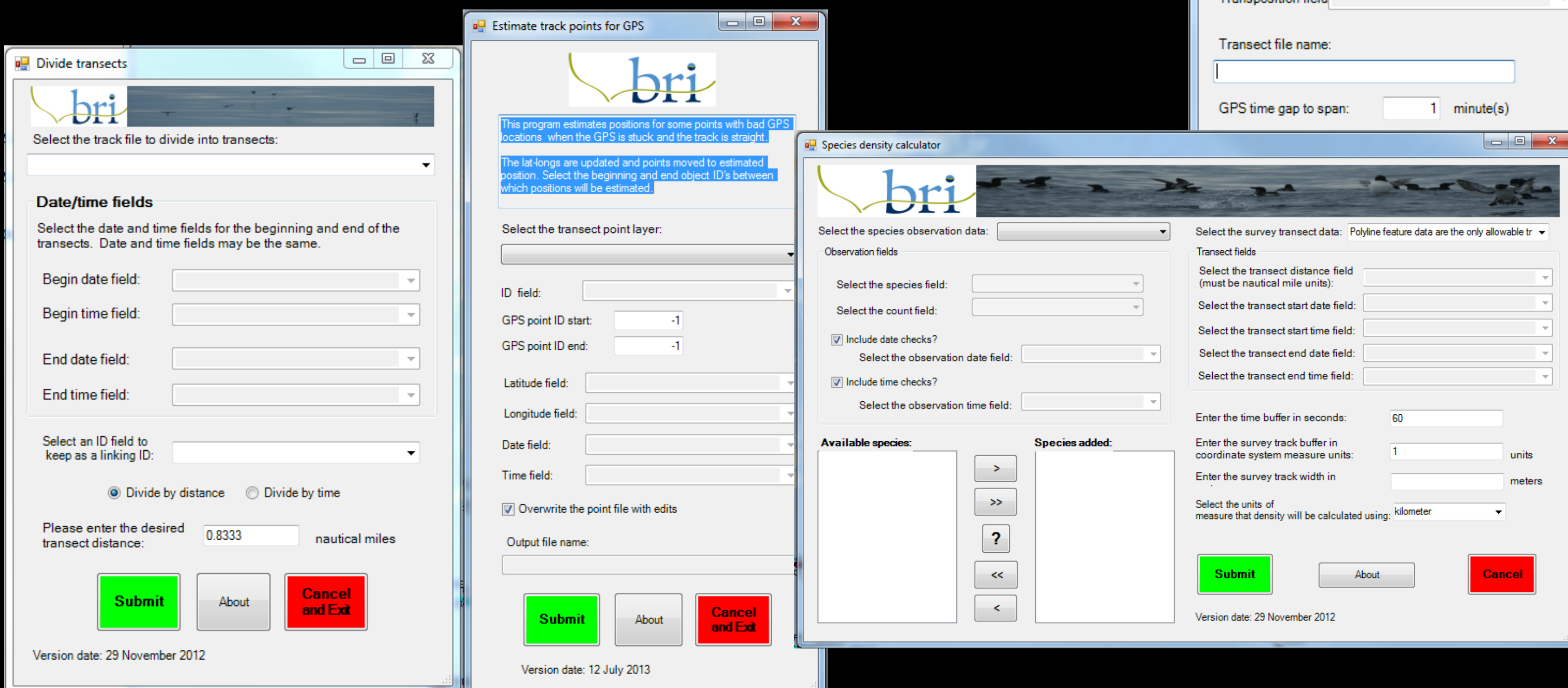


- Standardized survey effort in 5-min equivalents
  - Discrete time transects: 5 minute equivalents = # of 5 minute periods of survey
  - Continuous time transects: 5 minute equivalents = 0.8333 nautical mile survey segments (the distance traveled by a ship traveling 10 knots for 5 minutes)

Develop survey processing tools

- Tools developed to work with existing seabird datasets in ArcGIS 10.1
- Survey track divider, density estimator, Dlog track builder, GPS position estimation





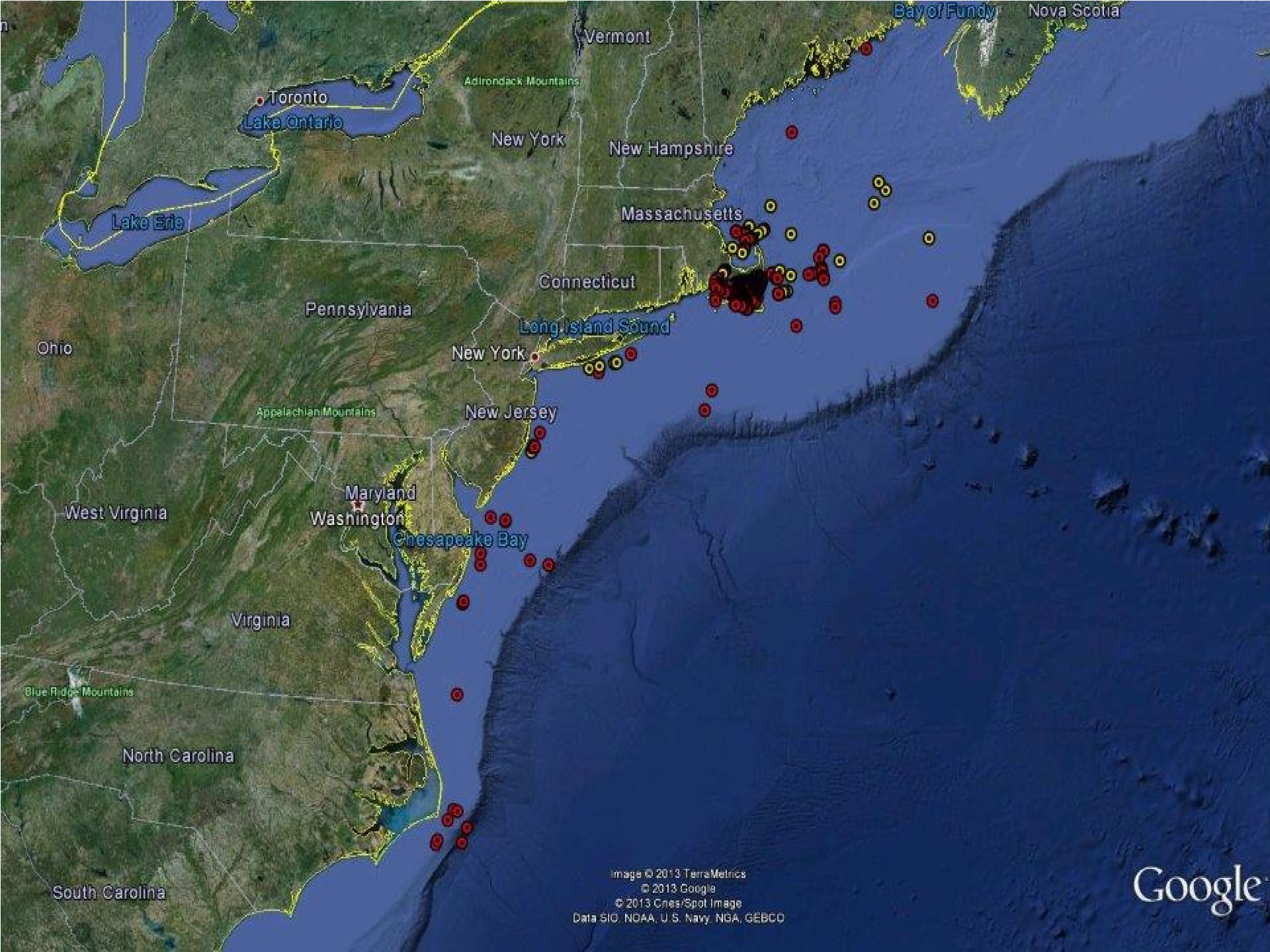
Current and future work

- Continue working with USGS database personnel to provide them with processed and QA/QC'd data from the most recent surveys available
  - DOE Mid-Atlantic baseline surveys – boat and hi-definition aerial
  - NOAA EcoMon and Herring Acoustic Surveys 2011/2012 boat surveys from Richard Veit
  - AMAPPS surveys (NOAA, BOEM, NAVY, USFWS)
- Work with modelers to update database and provide data support as needed
- Update tools as needed to make processing more efficient in current version of ArcGIS

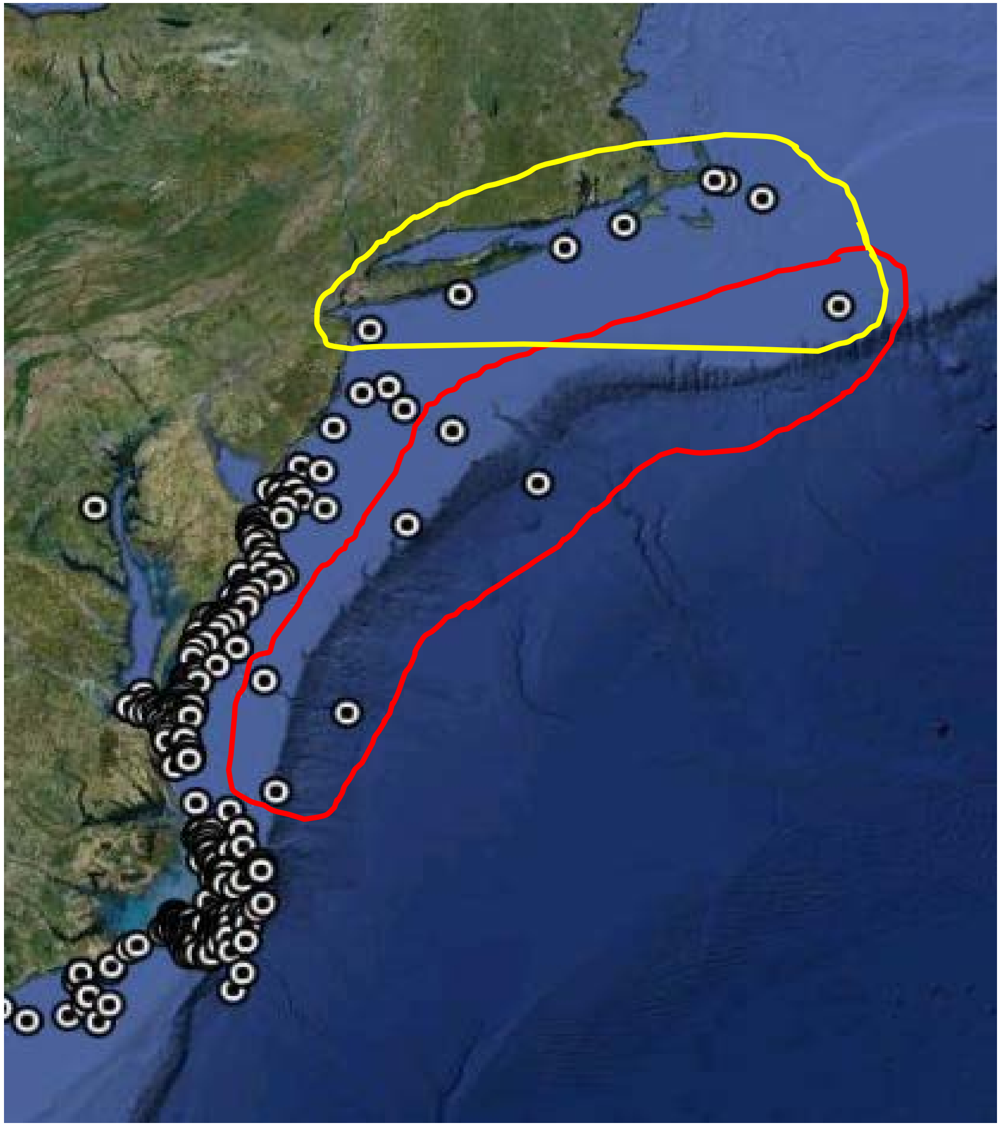
Error Checking of USGS Database

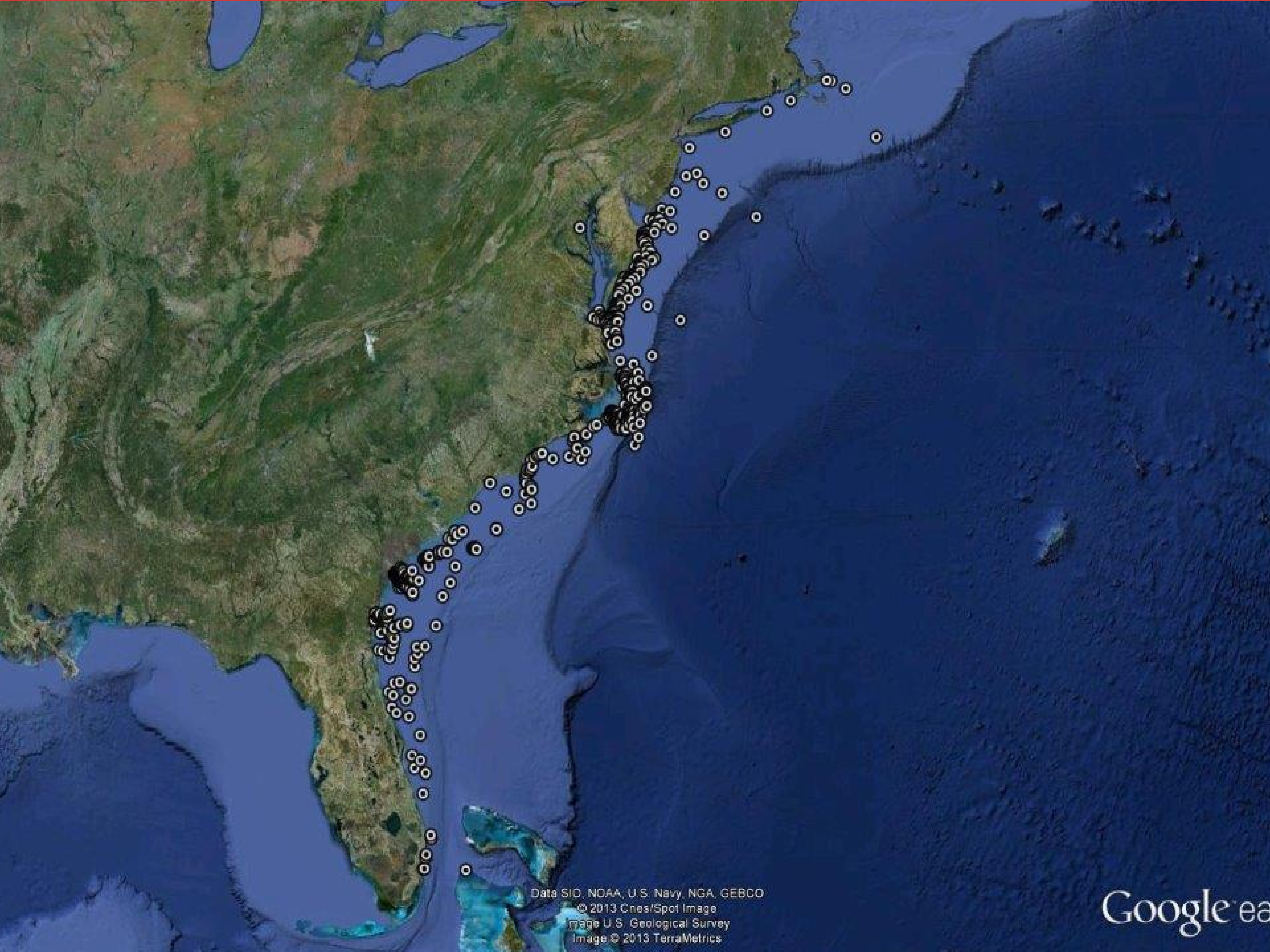
Richard Veit and Deborah Jaques

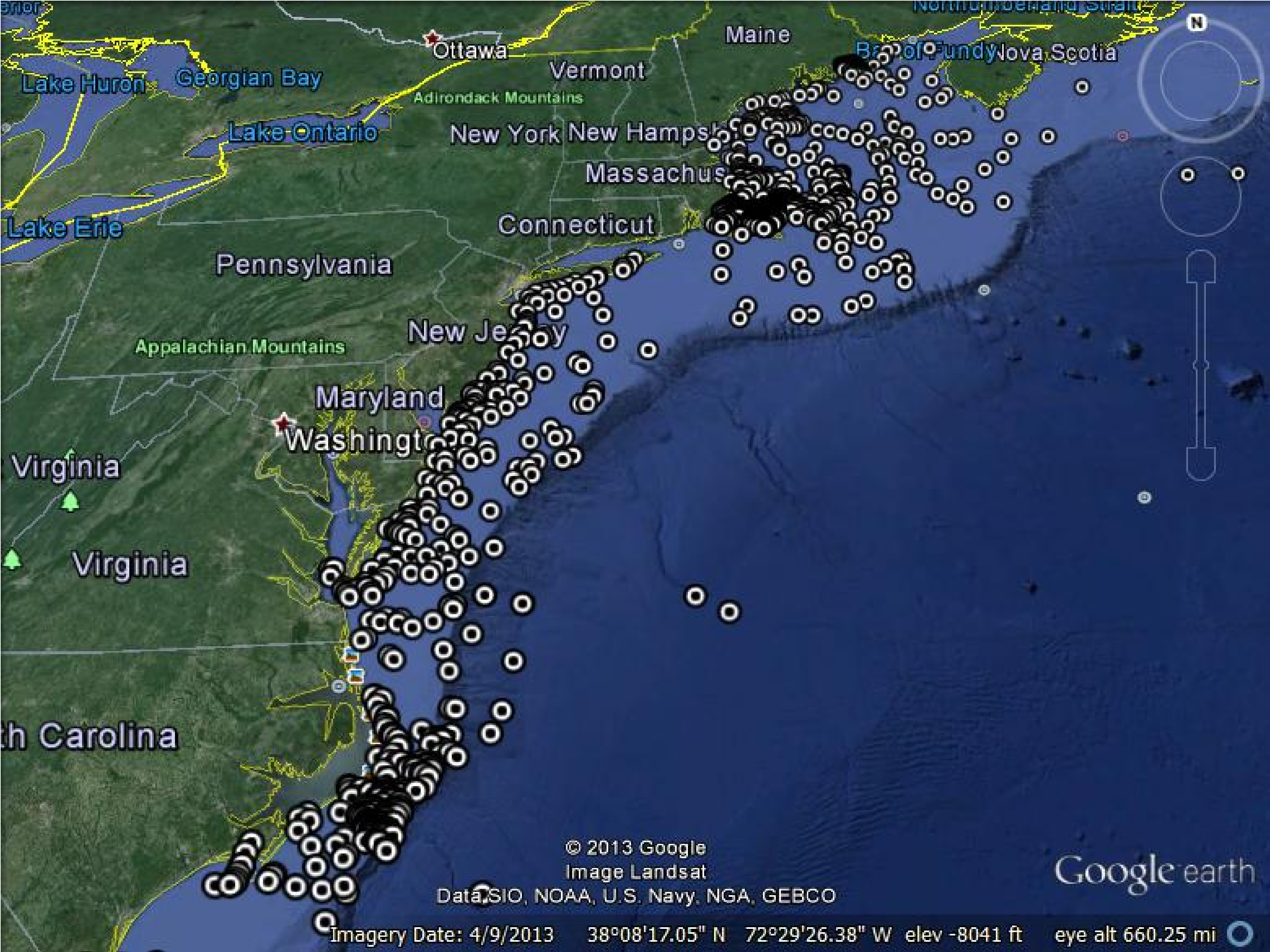
- 270,000 record database
- Checked for errors in “4 letter” codes
- Particular problem with “ROTE” meaning either Roseate Tern or Royal Tern
- We have made suggested changes
- Unidentified birds (e.g. UNTE, UNAL, UNSH) represent a large fraction of birds in database



Bay of Fundy
Nova Scotia
Vermont
Adirondack Mountains
Toronto
Lake Ontario
New York
New Hampshire
Massachusetts
Lake Erie
Connecticut
Pennsylvania
Long Island Sound
Ohio
New York
Appalachian Mountains
New Jersey
Maryland
West Virginia
Washington
Chesapeake Bay
Virginia
Blue Ridge Mountains
North Carolina
South Carolina
Image © 2013 TerraMetrics
© 2013 Google
© 2013 Cnes/Spot Image
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Google




Data SIO, NOAA, U.S. Navy, NGA, GEBCO
© 2013 Cnes/Spot Image
Image U.S. Geological Survey
Image © 2013 TerraMetrics



Maine
Ottawa
Vermont
Bay of Fundy
Nova Scotia
Lake Huron
Georgian Bay
Adirondack Mountains
Lake Ontario
New York
New Hampsh
Massachus
Connecticut
Lake Erie
Pennsylvania
New Jersey
Appalachian Mountains
Maryland
Washingt
Virginia
Virginia
th Carolina
© 2013 Google
Image Landsat
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Google earth
Imagery Date: 4/9/2013
38°08'17.05" N 72°29'26.38" W elev -8041 ft
eye alt 660.25 mi

Environmental and bird data collected separately

Bird data collected from ships, most modeled data long term averages from satellite

Predictive models of at-sea marine bird occurrence and abundance in the U.S. Mid-Atlantic

USFWS Northeast Region Science Seminar Series
August 8, 2013

Funding/collaborators:

North Atlantic Landscape Conservation Cooperative

USGS science for a changing world

BOEM Bureau of Ocean Energy Management

Brian Kinlan

brian.kinlan@noaa.gov

Robert Rankin

Chris Caldow

NCCOS

NOAA / NOS / National Centers for Coastal Ocean Science (NCCOS)
Silver Spring, MD

Study Areas

Phased approach

Phase I (2010-2012): New York Bight
NOAA, NY Dept. of State

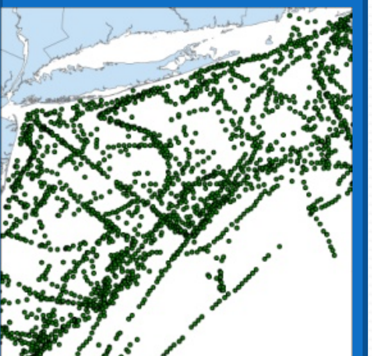
Phase II (2011-2013): U.S. Mid-Atlantic
NOAA, BOEM, USGS

Phase III (2013-2015): U.S. Atlantic
NOAA, BOEM, USGS

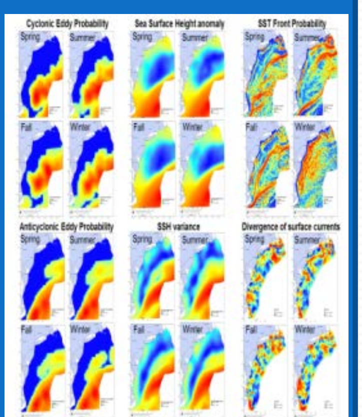
Statistical modeling approach

Zero-inflated Negative Binomial GAMLSS model (Generalized Additive Modeling of location, scale, and shape) fit using Component-wise Ensemble Gradient Boosting (*mboost, gamboostLSS*)

Data



Predictors



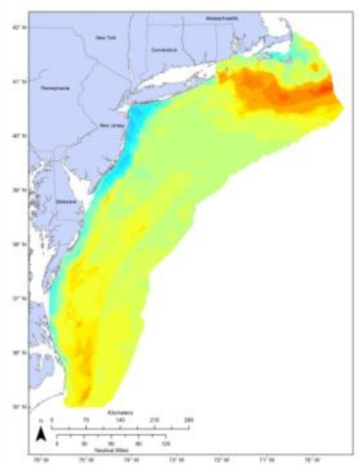
Occurrence Component

Boosted GAMLSS Ensemble

Bernoulli component

Tune with

Cross-validation Brier Score



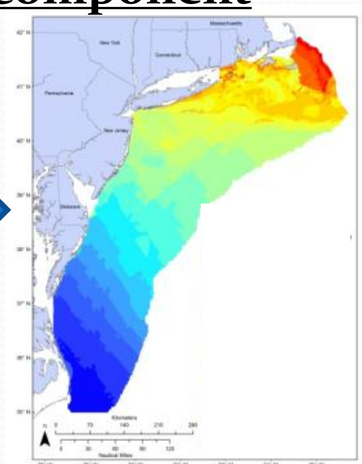
Abundance Component

Boosted GAMLSS Ensemble

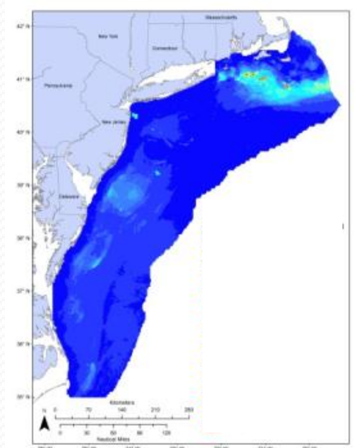
Negative binomial count component

Tune with

Multi-threshold Cross-validation Brier Score



Prediction: Expected Number of Birds Sighted in standardized survey transect



- Assess uncertainty via bootstrap

Details: Hierarchical ZIP/NB GAMLSS Model, Component-wise Ensemble Boosting

- Technique based on Schimd et al. (2010); Component-wise modification of ‘mboost’ R package (Bühlmann et al 2007)
- Loss function: negative log likelihood of the ZIP/NB (l)
- Iteratively fit base-learners to the partial derivative of the negative log likelihood w.r.t each parameter’s fitted function, $f_k(x)$.
- Cross-validation (Brier Score) to tune size of ensemble and shrinkage rate

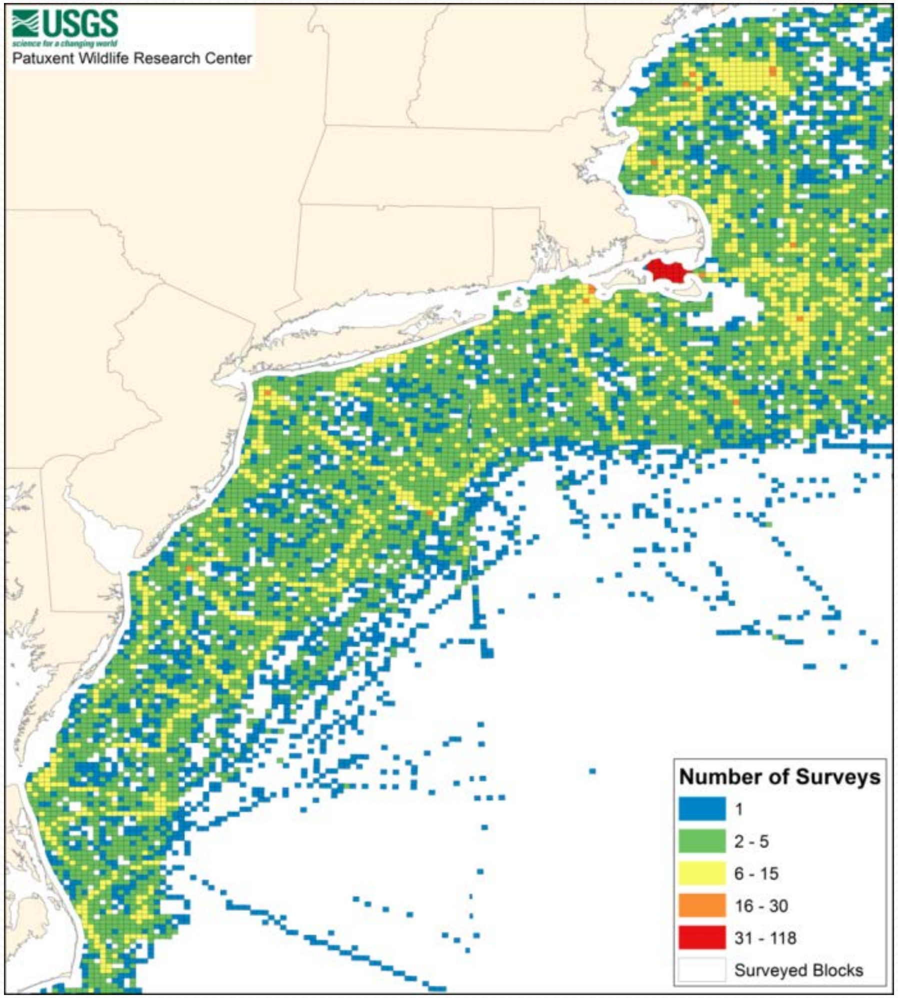
---

- *Schapire, Robert E. “The Strength of Weak Learnability.” Machine Learning 5, no. 2 (1990): 197–227. doi:10.1023/A:1022648800760.*
- *Friedman, Jerome H. “Greedy Function Approximation: a Gradient Boosting Machine.” Annals of Statistics 29, no. 5 (2001): 1189–1232.*
- *Elith, J, J R Leathwick, and T Hastie. “A Working Guide to Boosted Regression Trees.” The Journal of Animal Ecology 77, no. 4 (July 2008): 802–813. doi:10.1111/j.1365-2656.2008.01390.x.*
- *Schmid, M., S. Potapov, A. Pfahlberg, and T. Hothorn. “Estimation and Regularization Techniques for Regression Models with Multidimensional Prediction Functions.” Statistics and Computing 20, no. 2 (2010): 139–150.*
- *Borisov, Alexander, G. Runger, E. Tuv, and Nutta Lurponglukana-Strand. “Zero-inflated Boosted Ensembles for Rare Event Counts.” Advances in Intelligent Data Analysis VIII 5572 (2009): 225–236. doi:10.1007/978-3-642-03915-7_20.*
- *Hutchinson, R.A., L.P. Liu, and T.G. Dietterich. “Incorporating Boosted Regression Trees into Ecological Latent Variable Models.” In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, edited by W Burgard and D. Roth, 1343–1348. Association for the Advancement of Artificial Intelligence, 2011. http://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/viewFile/3711/4086.*
- Mayr, Andreas, Nora Fenske, Benjamin Hofner, Thomas Kneib, and Matthias Schmid. “Generalized Additive Models for Location, Scale and Shape for High Dimensional Data—a Flexible Approach Based on Boosting.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61, no. 3 (2012): 403–427. doi:10.1111/j.1467-9876.2011.01033.x.
- *Bühlmann, Peter, and Torsten Hothorn. “Boosting Algorithms: Regularization, Prediction and Model Fitting.” Statistical Science 22, no. 4 (November 2007): 477–505. doi:10.1214/07-STS242.*

Compendium of Avian Information in the U.S. Atlantic Outer Continental Shelf (USGS, BOEM)

For modeling purposes:

- 32 scientific data sets – 28 ship-based, 4 aerial
- Transects were standardized to 4.63km
- 44,176 survey transects representing 463 species



Summary of modeled species

- **Modeled 26 species individually**
- **Otherwise grouped taxonomically and ecologically similar species**
- **Up to four seasonal models for each species/group: Spring, Summer, Fall, Winter**

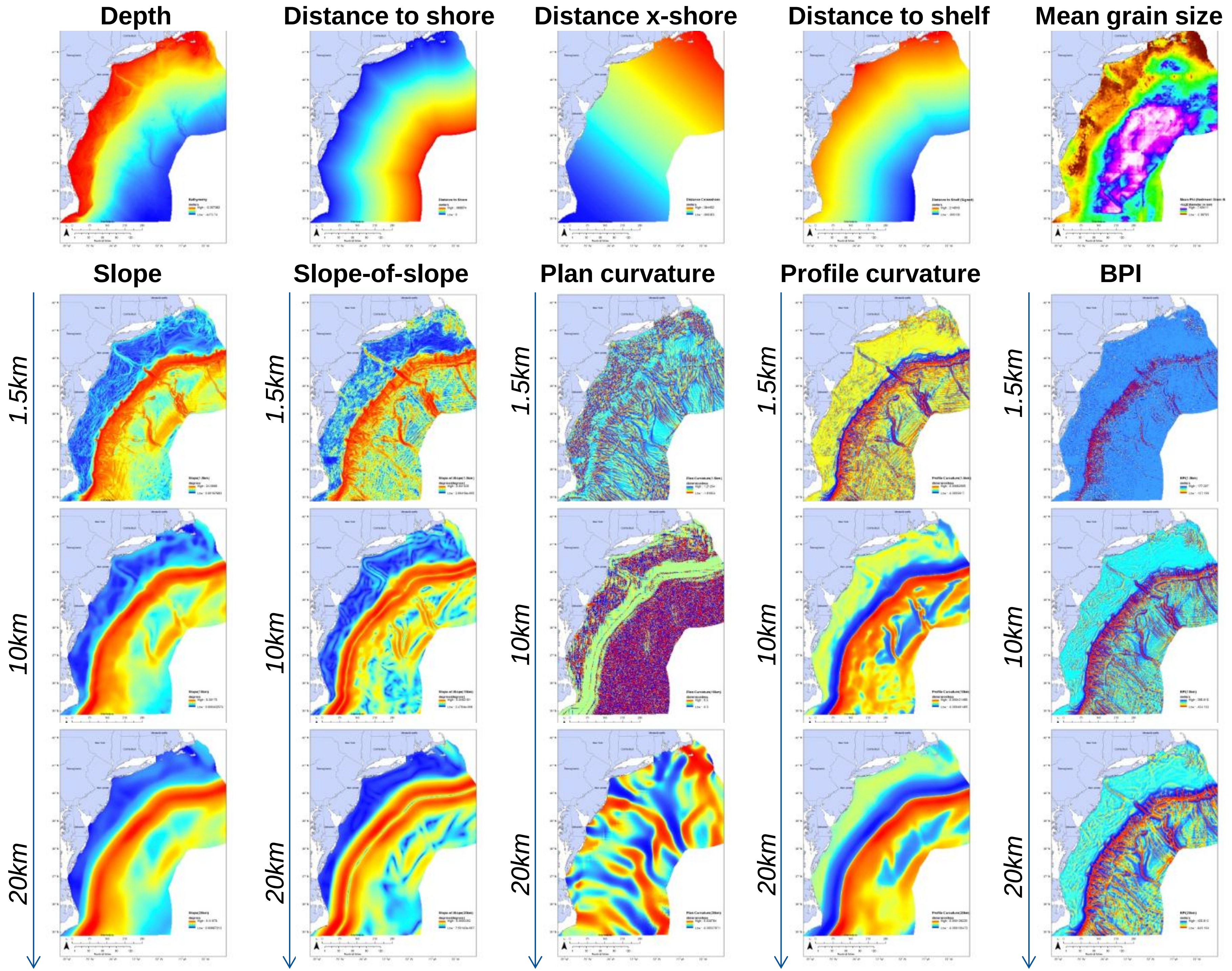
				Modeled?			
Species code	Common name	Scientific name	Family	Spring	Summer	Fall	Winter
aush	Audubon's Shearwater	*Puffinus lherminieri*	Procellariidae	No	Yes	Yes	No
blki	Black-Legged Kittiwake	*Rissa tridactyla*	Laridae	Yes	No	Yes	Yes
blsc	Black Scoter	*Melanitta americana*	Anatidae	Yes	No	Yes	Yes
bogu	Bonaparte's Gull	*Chroicocephalus philadelphia*	Laridae	No	No	No	Yes
coei	Common Eider	*Somateria mollissima*	Anatidae	Yes	Yes	Yes	Yes
colo	Common Loon	*Gavia immer*	Gaviidae	Yes	Yes	Yes	Yes
cosh	Cory's Shearwater	*Calonectris diomedea*	Procellariidae	No	Yes	Yes	No
cote	Common Tern	*Sterna hirundo*	Sternidae	Yes	Yes	Yes	No
dove	Dovekie	*Alle alle*	Alcidae	No	No	Yes	Yes
gbbg	Great Black-Backed Gull	*Larus marinus*	Laridae	Yes	Yes	Yes	Yes
grsh	Great Shearwater	*Puffinus gravis*	Procellariidae	Yes	Yes	Yes	No
herg	Herring Gull	*Larus argentatus*	Laridae	Yes	Yes	Yes	Yes
lagu	Laughing Gull	*Leucophaeus atricilla*	Laridae	Yes	Yes	Yes	No
lesp	Leach's Storm-Petrel	*Oceanodroma leucorhoa*	Hydrobatidae	No	Yes	Yes	No
ltdu	Long-tailed Duck	*Clangula hyemalis*	Anatidae	Yes	Yes	Yes	Yes
nofu	Northern Fulmar	*Fulmarus glacialis*	Procellariidae	Yes	Yes	Yes	Yes
noga	Northern Gannet	*Morus bassanus*	Sulidae	Yes	Yes	Yes	Yes
poja	Pomarine Jaeger	*Stercorarius pomarinus*	Stercorariidae	No	No	Yes	No
razo	Razorbill	*Alca torda*	Alcidae	Yes	Yes	Yes	Yes
reph	Red Phalarope	*Phalaropus fulicarius*	Scolopacidae	Yes	No	No	No
rtlo	Red-Throated Loon	*Gavia stellata*	Gaviidae	Yes	No	Yes	Yes
sosh	Sooty Shearwater	*Puffinus griseus*	Procellariidae	Yes	Yes	No	No
susc	Surf Scoter	*Melanitta perspicillata*	Anatidae	Yes	Yes	Yes	Yes
wisp	Wilson's Storm-Petrel	*Oceanites oceanicus*	Hydrobatidae	Yes	Yes	Yes	No
wwsc	White-Winged Scoter	*Melanitta fusca*	Anatidae	Yes	Yes	Yes	Yes

Potential Predictor Variables

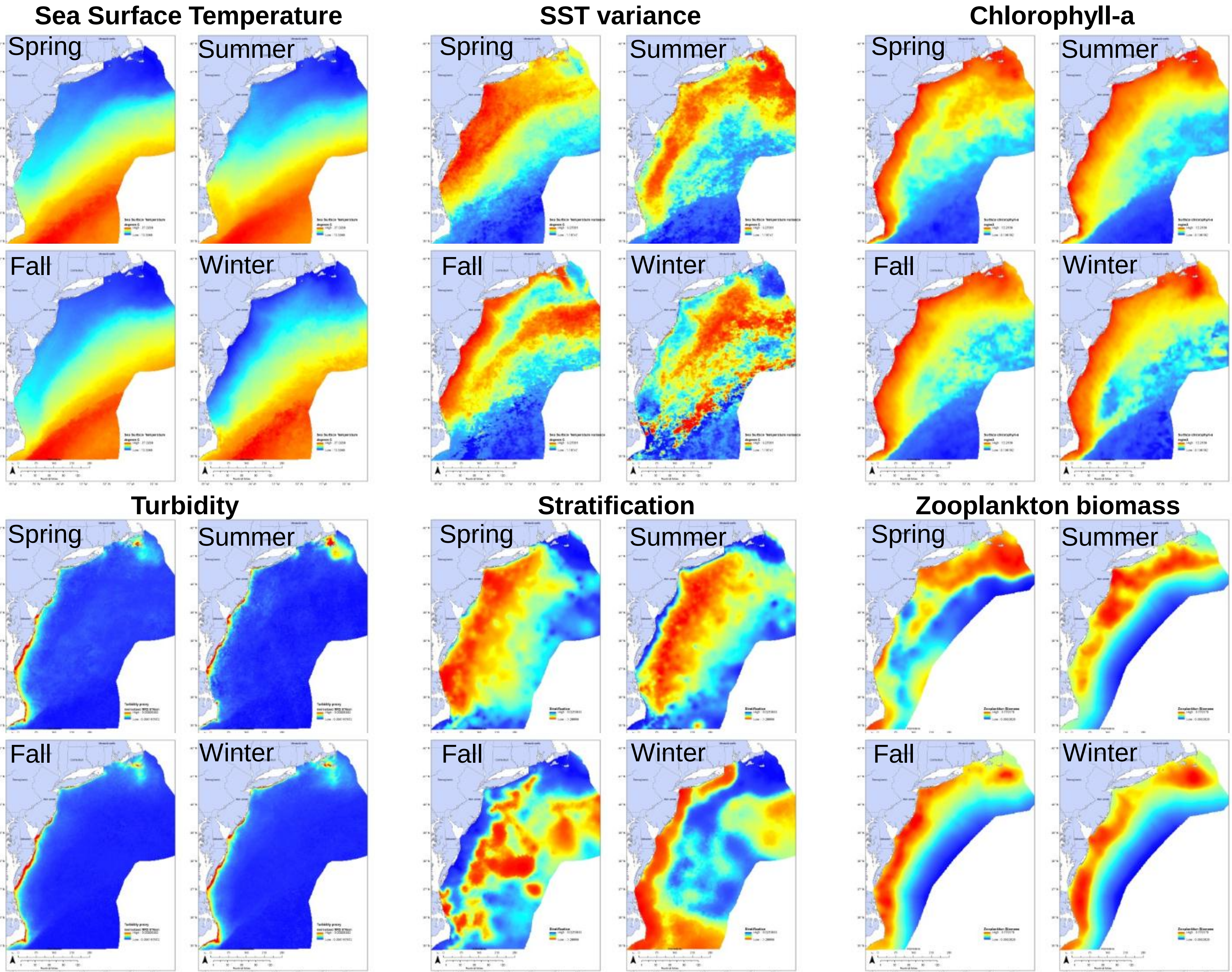
- **Spatial variables**
  - Static (e.g., Bathymetry)
  - Dynamic (e.g., SST) – Seasonal Climatologies

- **Temporal variables**
  - Climate Indices (with and without lags)
  - Year
  - Julian Day
  - Models fit separately by season (Spring, Summer, Fall, Winter)

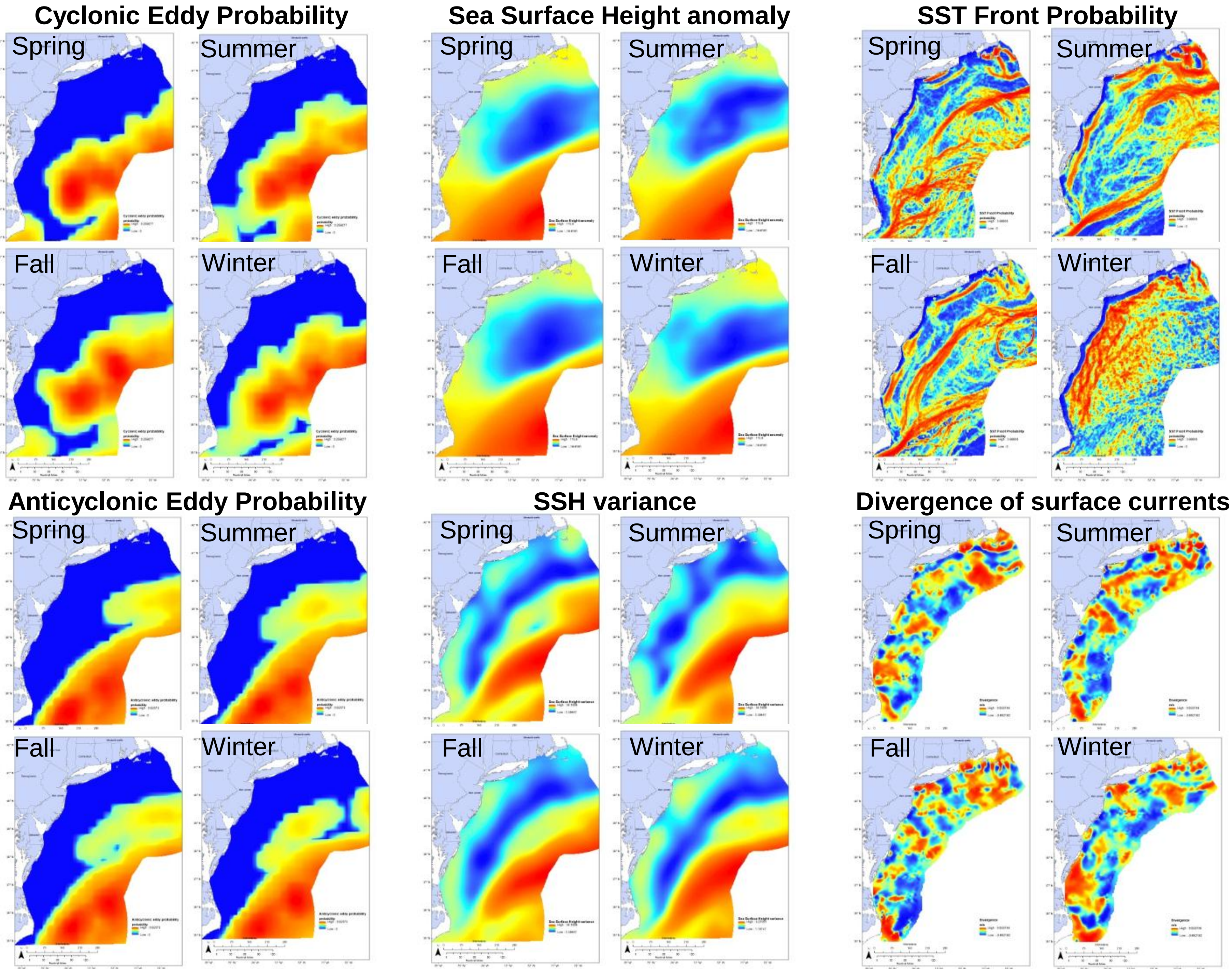
- **Survey variables**
  - Transect
  - Dataset
  - Platform



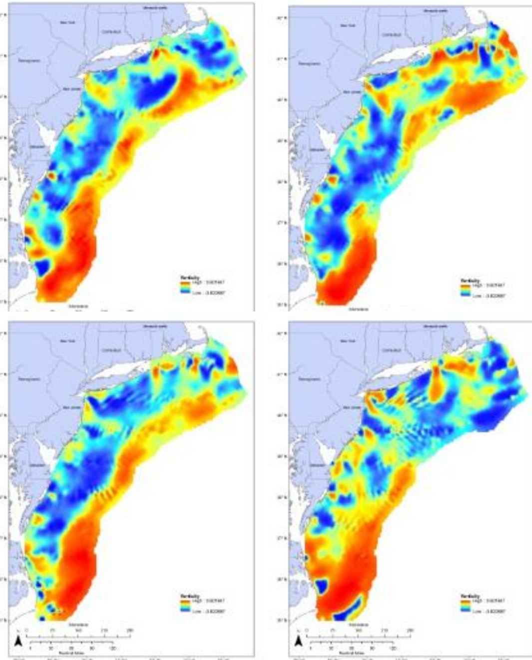
Depth
Distance to shore
Distance x-shore
Distance to shelf
Mean grain size
Slope
Slope-of-slope
Plan curvature
Profile curvature
BPI
1.5km
10km
20km



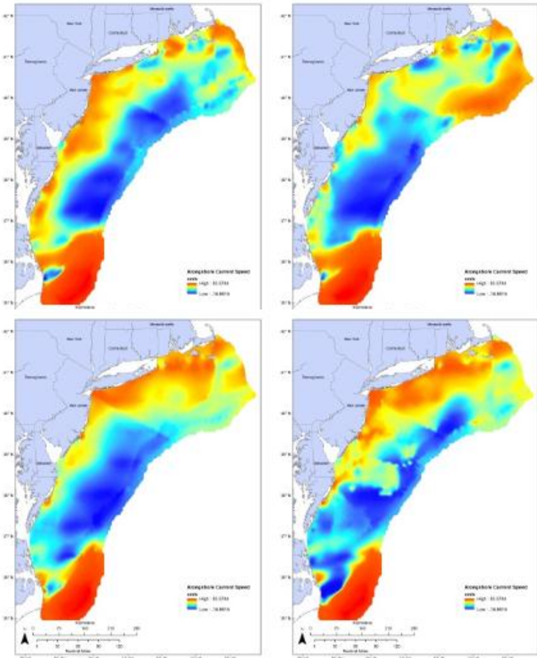
Sea Surface Temperature
Spring
Summer
Fall
Winter
SST variance
Spring
Summer
Fall
Winter
Chlorophyll-a
Spring
Summer
Fall
Winter
Turbidity
Spring
Summer
Fall
Winter
Stratification
Spring
Summer
Fall
Winter
Zooplankton biomass
Spring
Summer
Fall
Winter


Cyclonic Eddy Probability
Spring
Summer
Fall
Winter
Sea Surface Height anomaly
Spring
Summer
Fall
Winter
SST Front Probability
Spring
Summer
Fall
Winter
Anticyclonic Eddy Probability
Spring
Summer
Fall
Winter
SSH variance
Spring
Summer
Fall
Winter
Divergence of surface currents
Spring
Summer
Fall
Winter

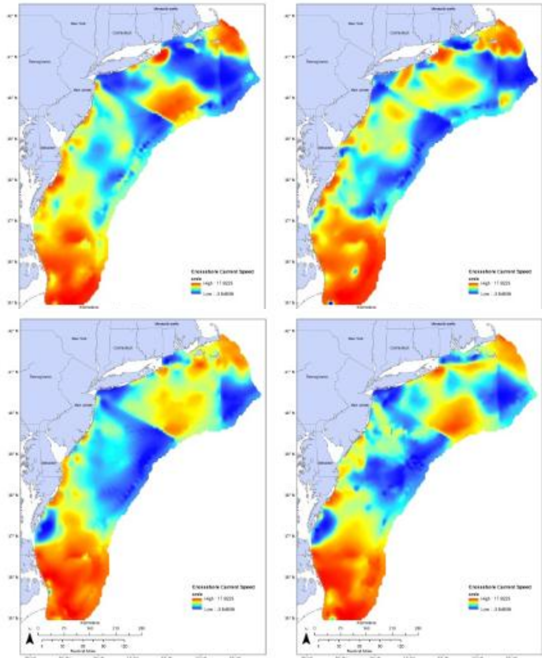
Vorticity



Alongshore Current Speed

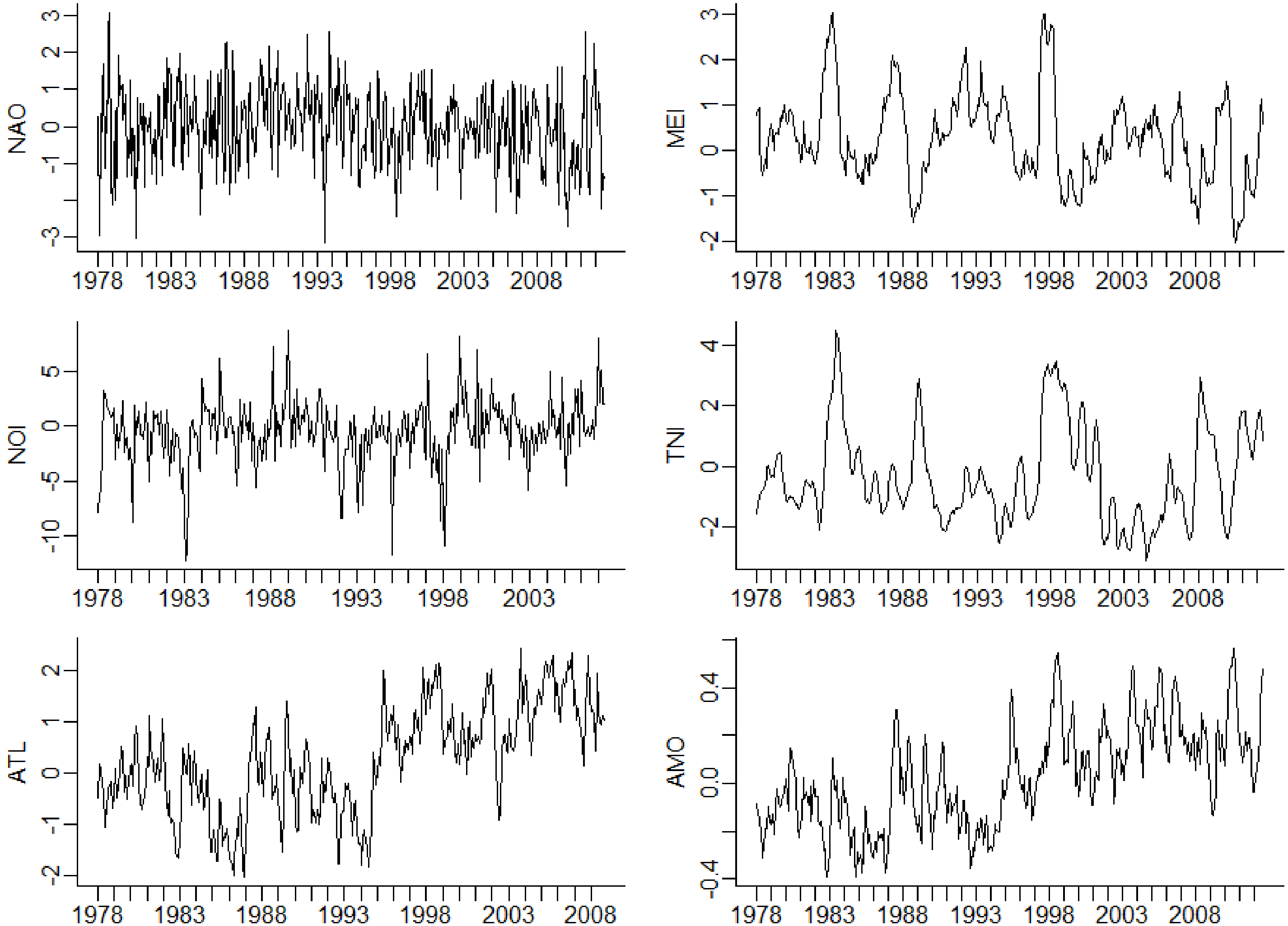


Crossshore Current Speed



Temporal variables

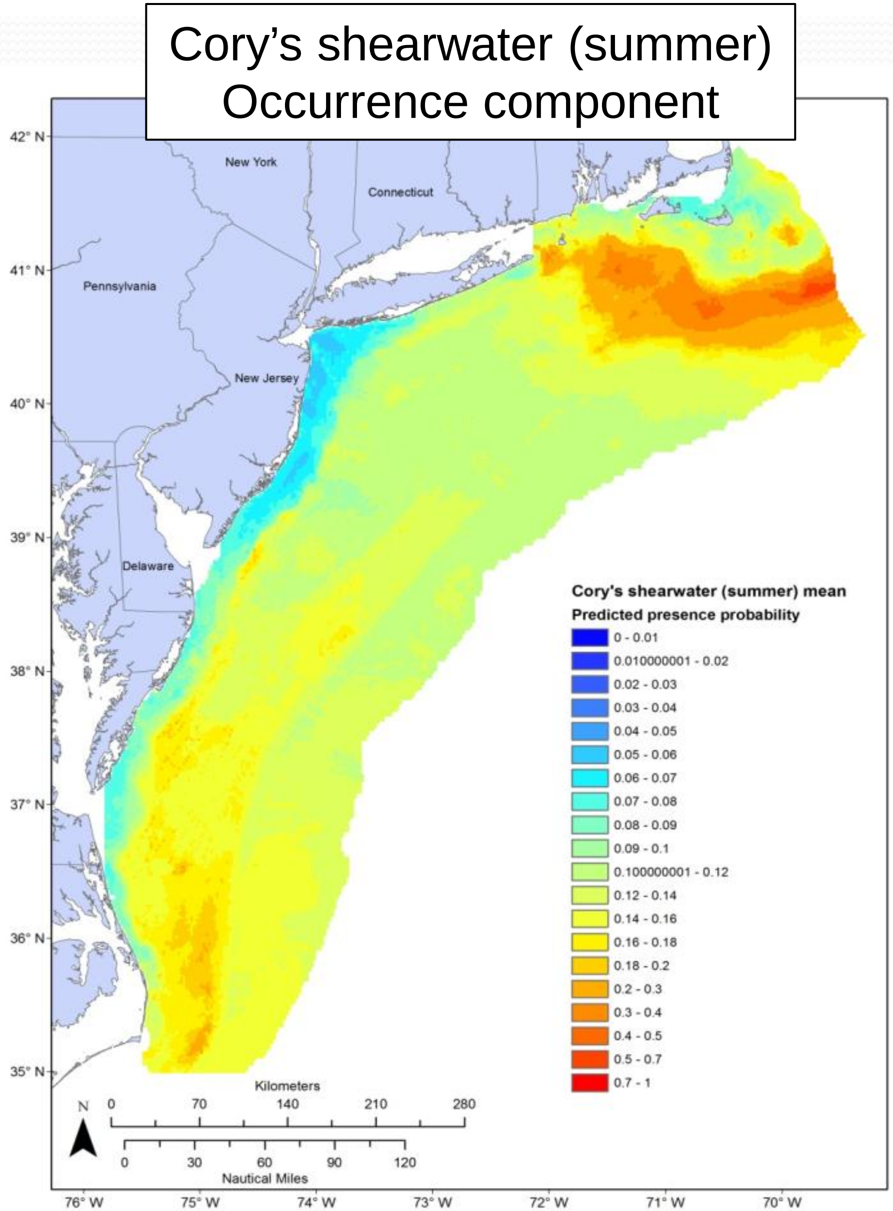
- Year
- Julian day
- Climate index time series (at lag 0 and lag 1 year)

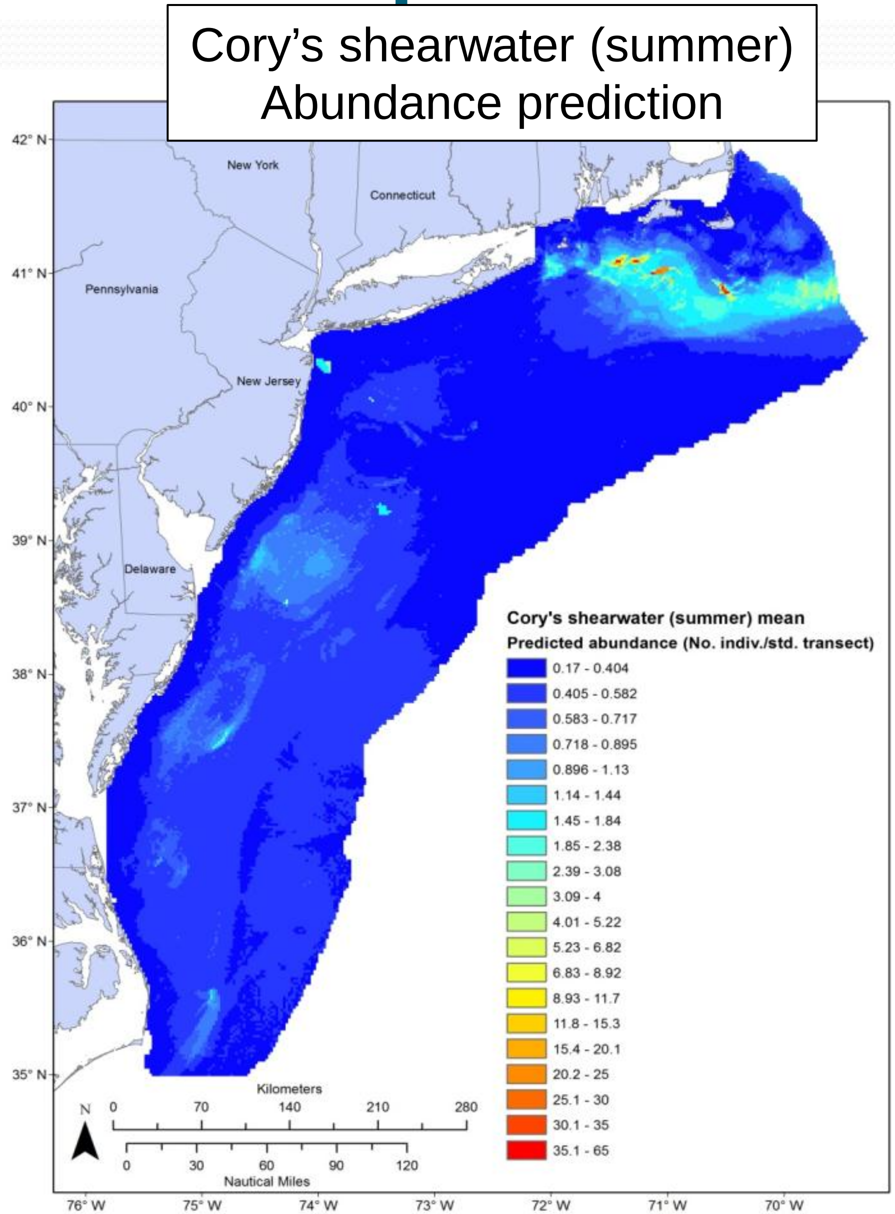


Survey variables

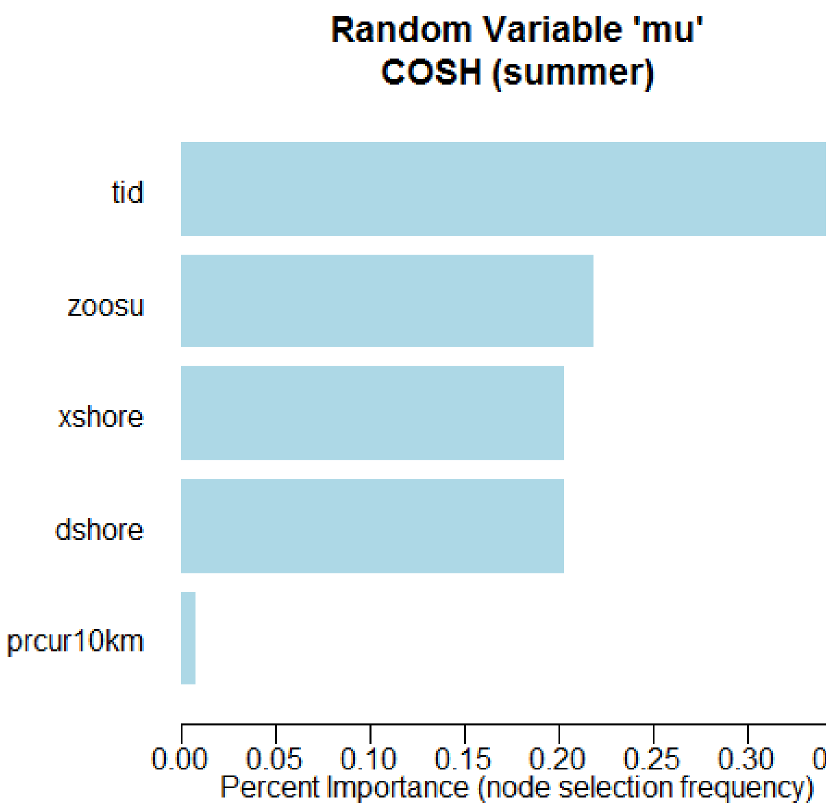
- Transect ID
- Dataset ID
- Platform (Boat/Plane)

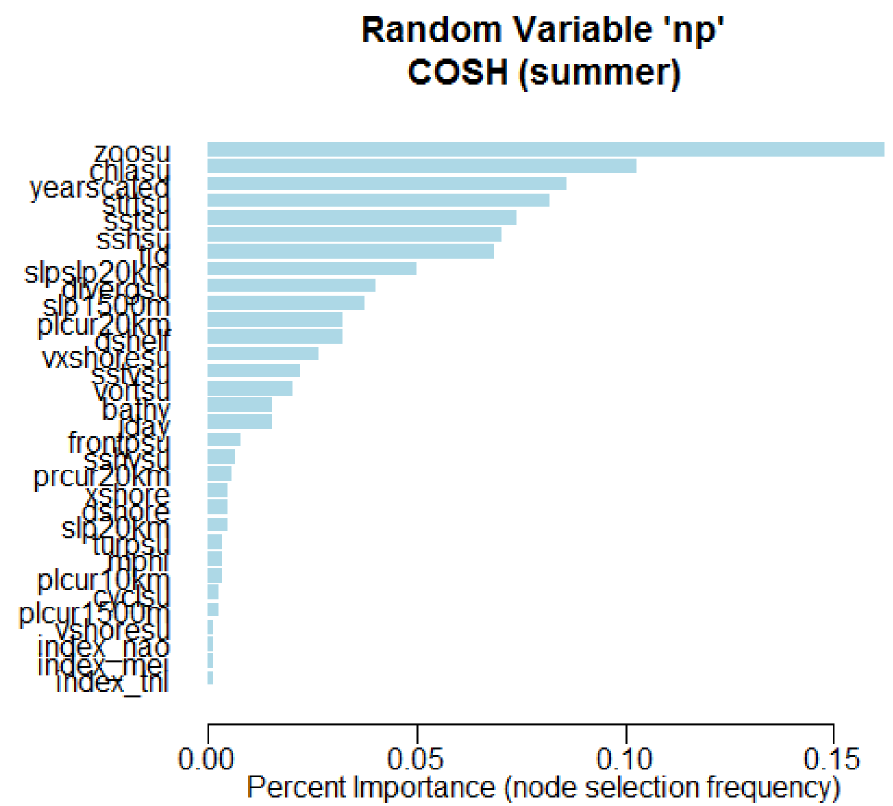
Predictive maps of long-term occurrence and abundance patterns



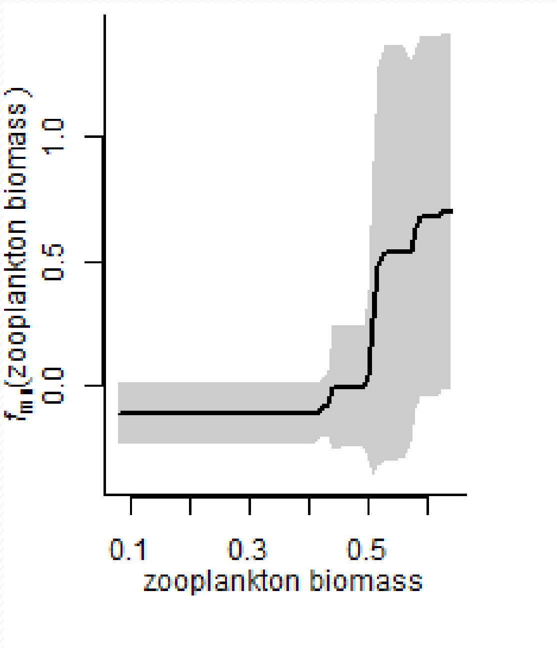


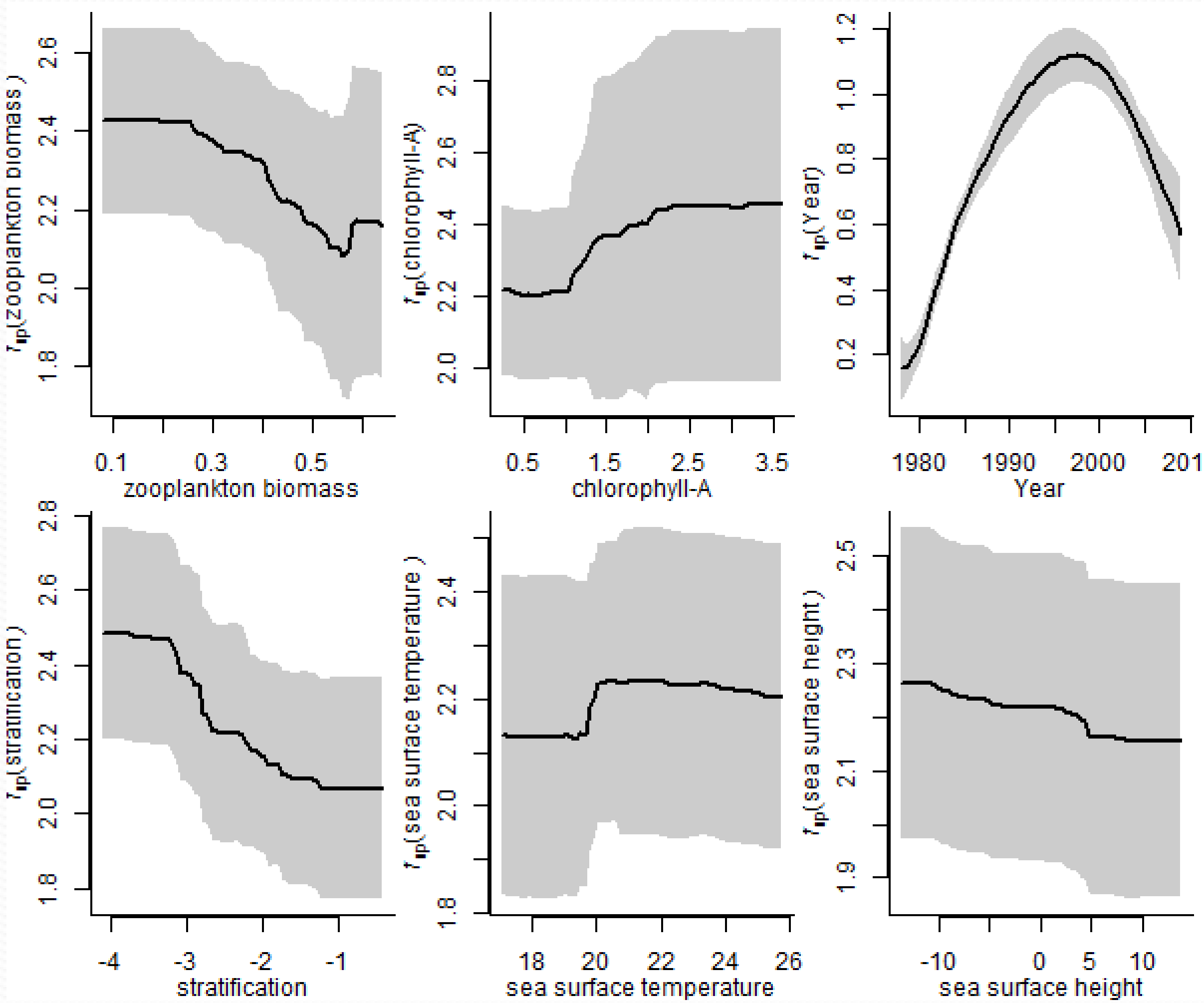
Insights into important predictors





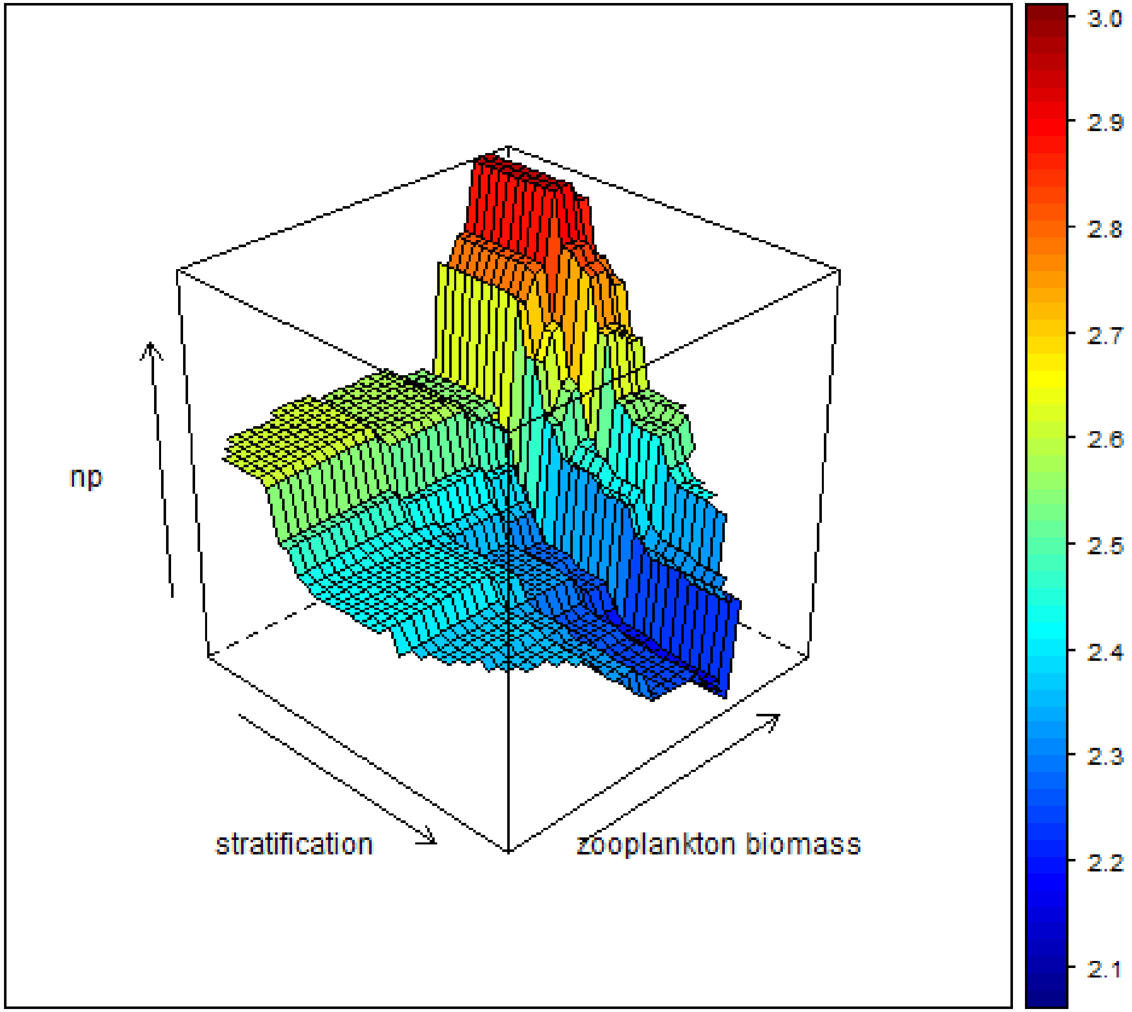
Flexible estimation of non-linear effects





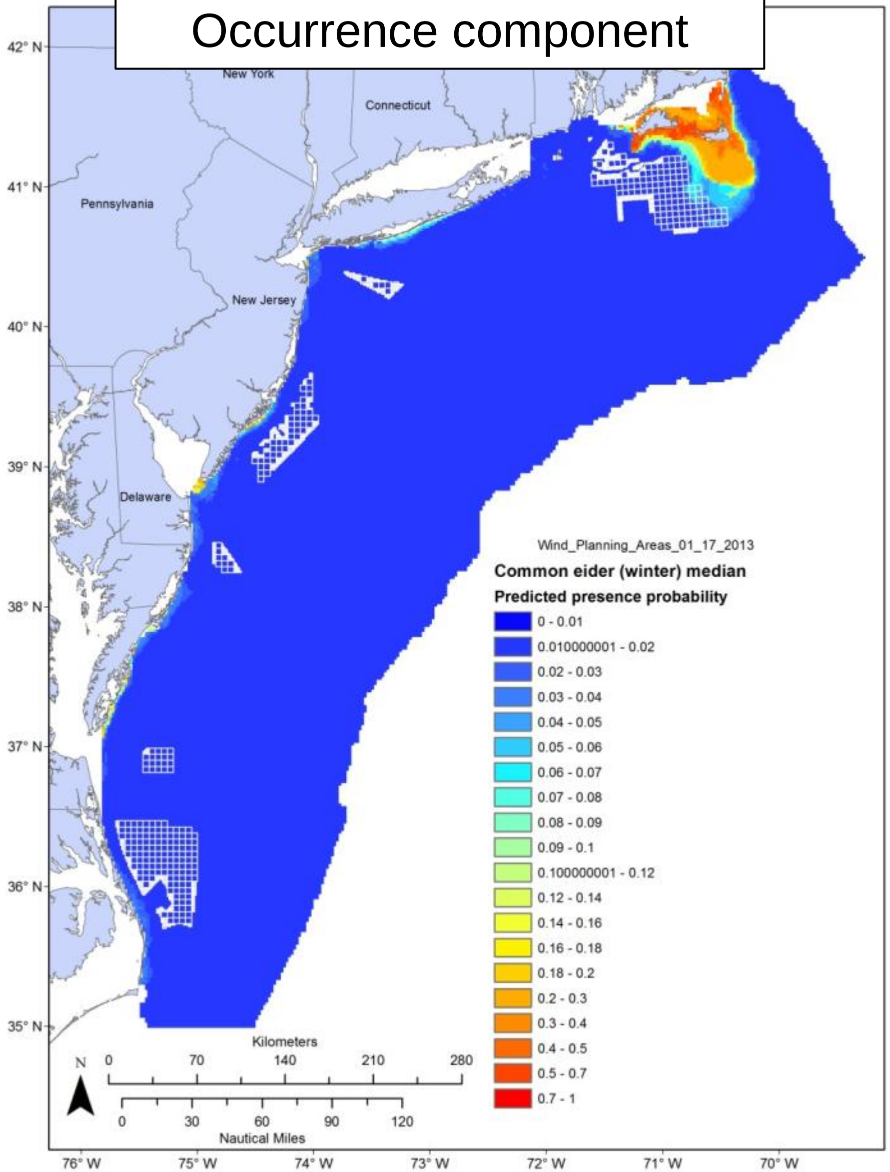
Characterization of complex interactions

Cory's shearwater (summer) – Occurrence component

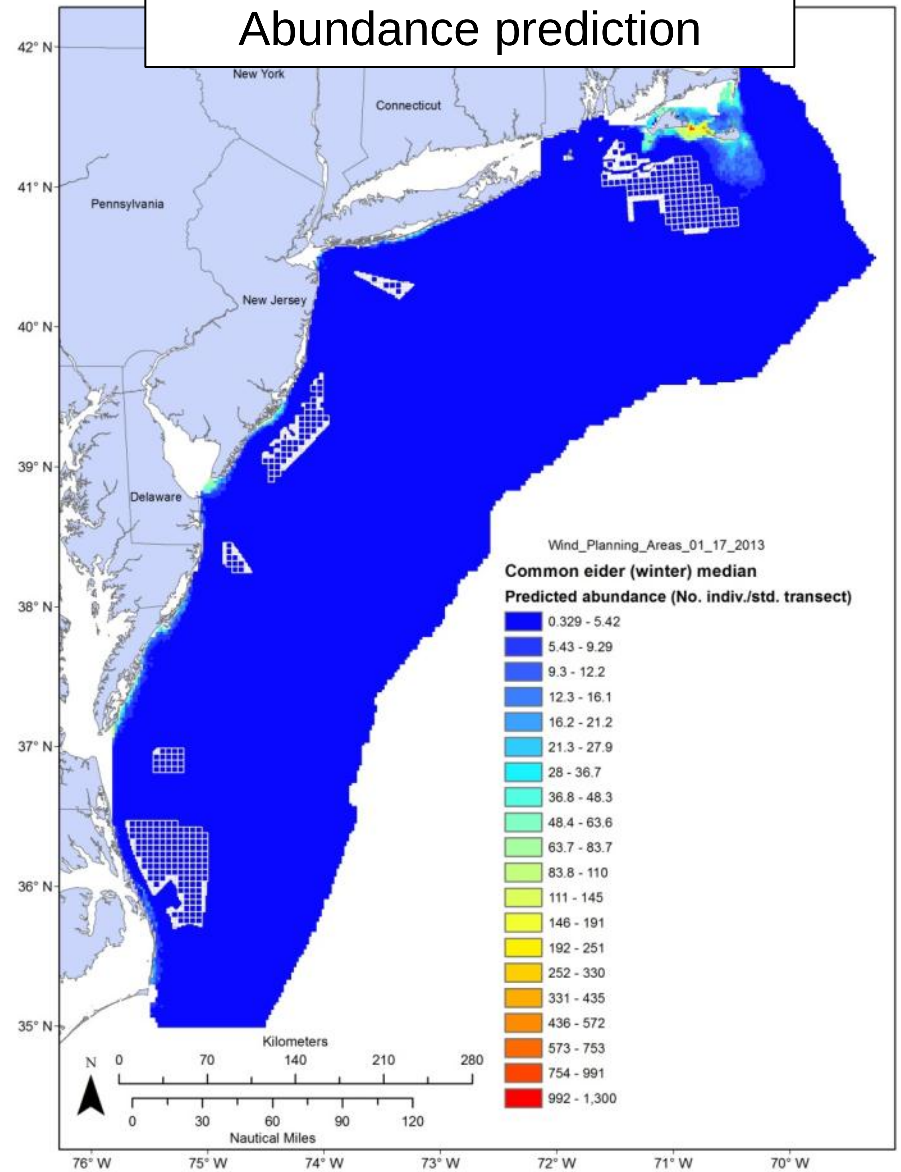


Input into wind energy planning

Common eider (winter)
Occurrence component

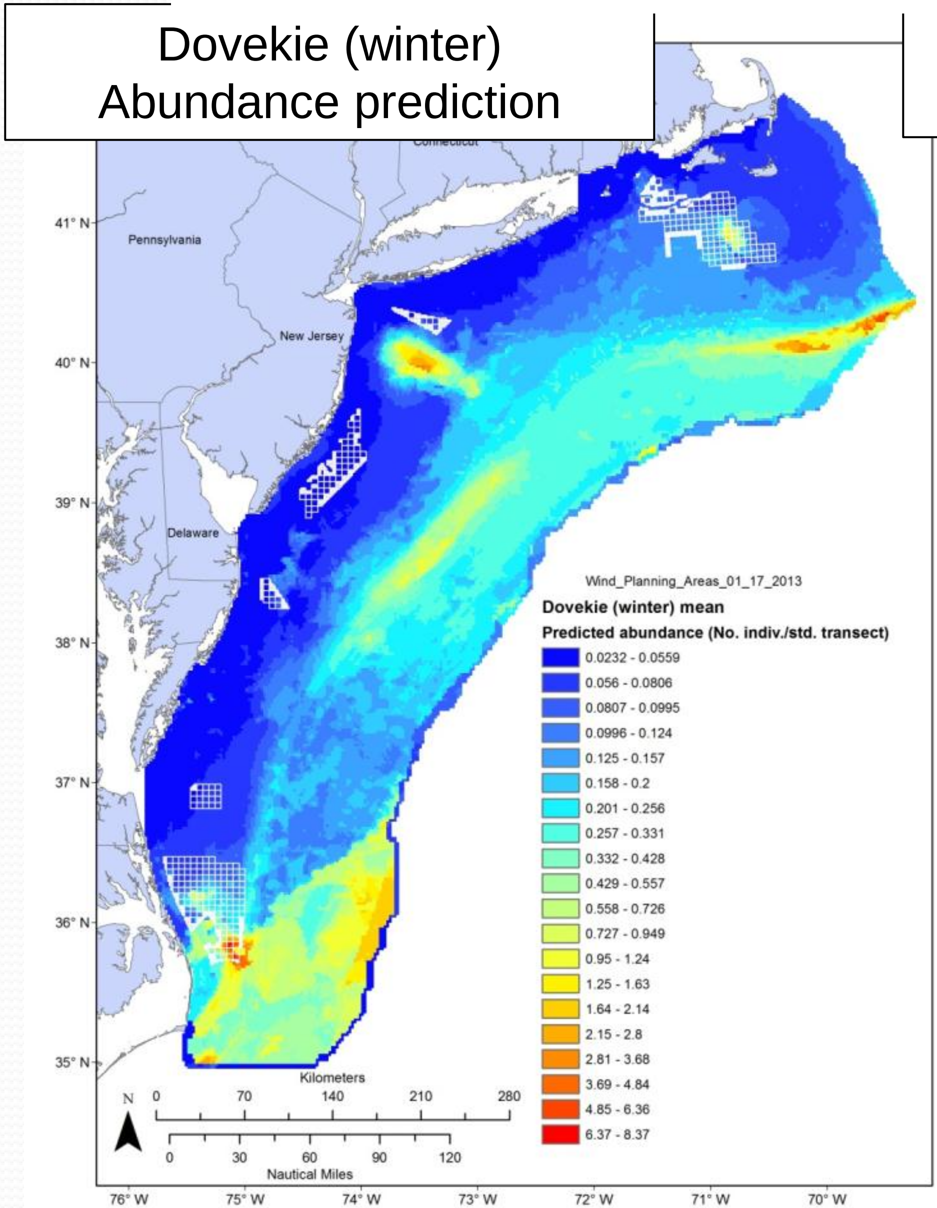


Common eider (winter)
Abundance prediction

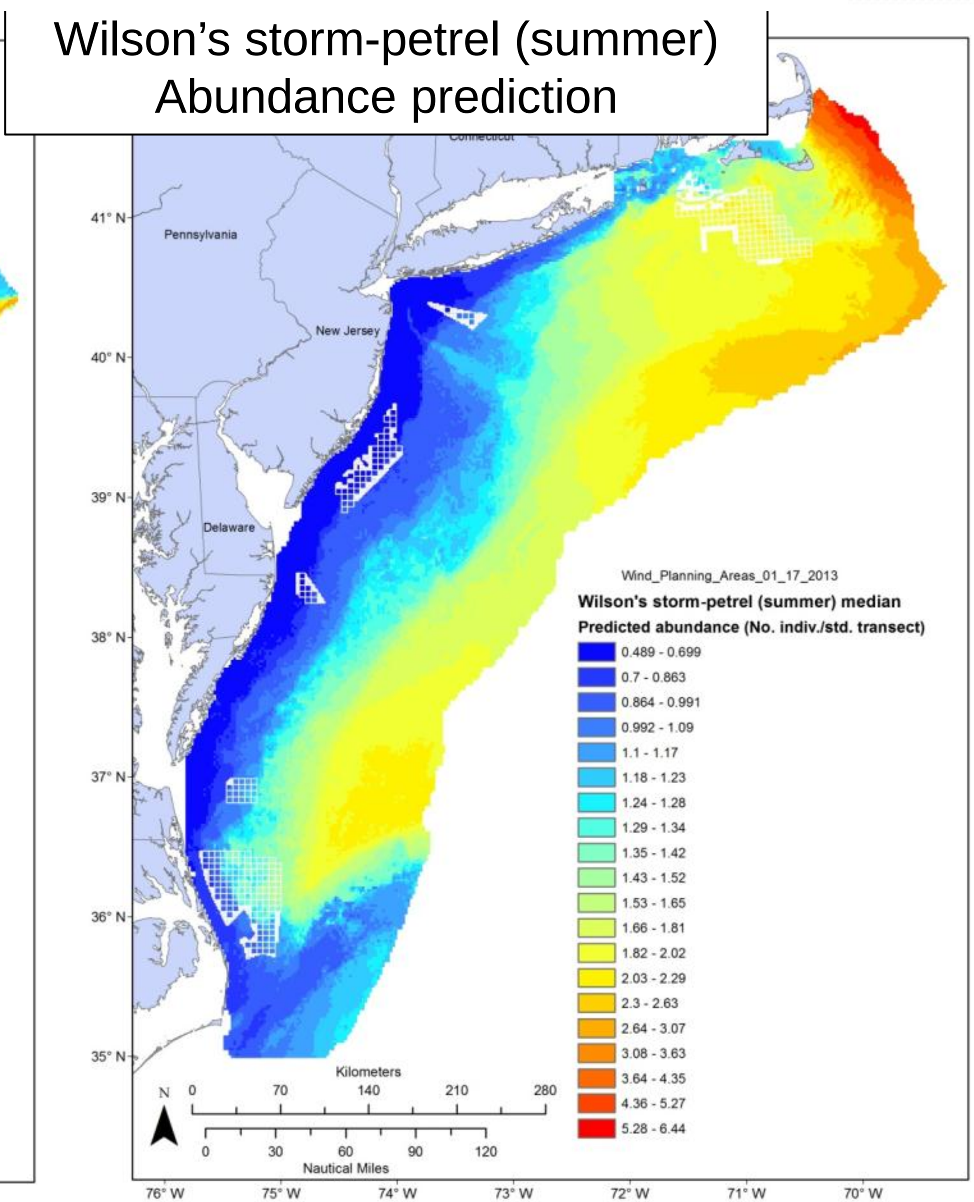


Can't avoid all species; need for weighting by sensitivity/vulnerability

Dovekie (winter)
Abundance prediction

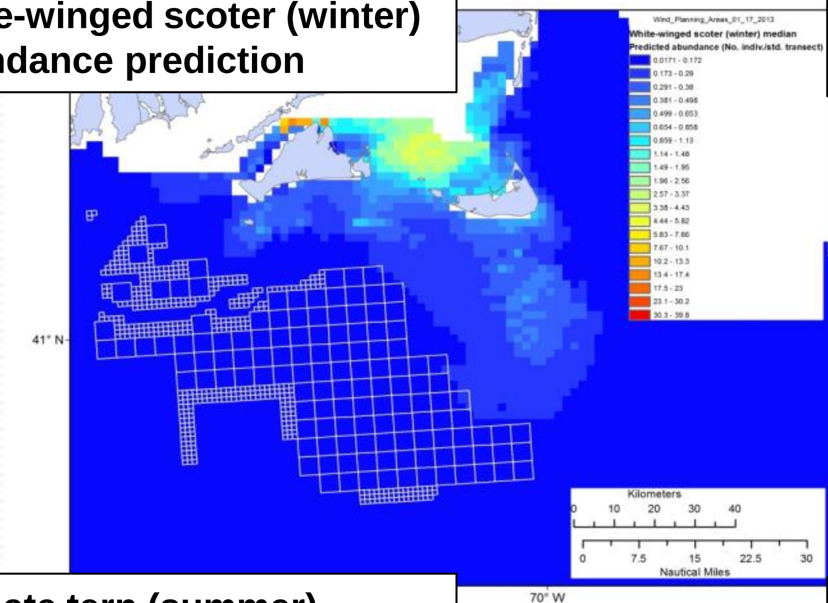


Wilson's storm-petrel (summer)
Abundance prediction

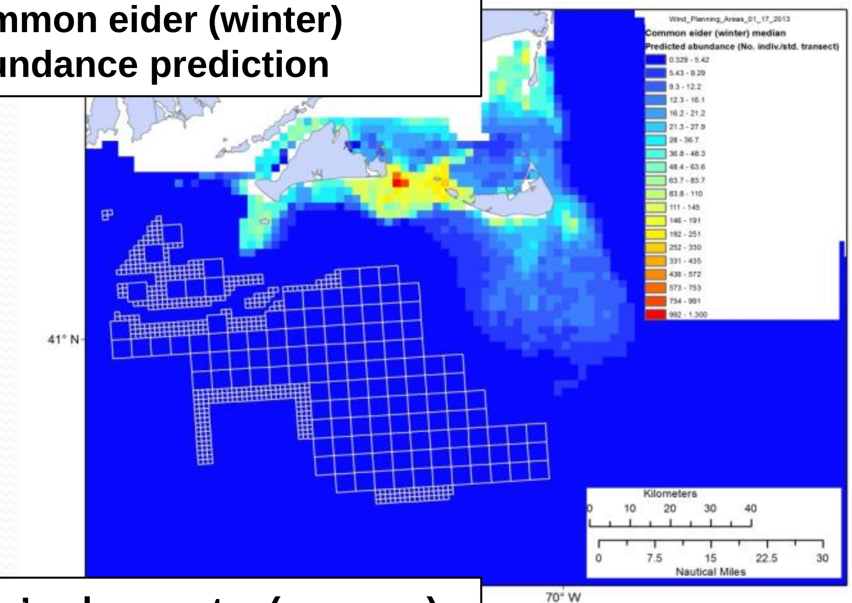


Value of fine-scale information

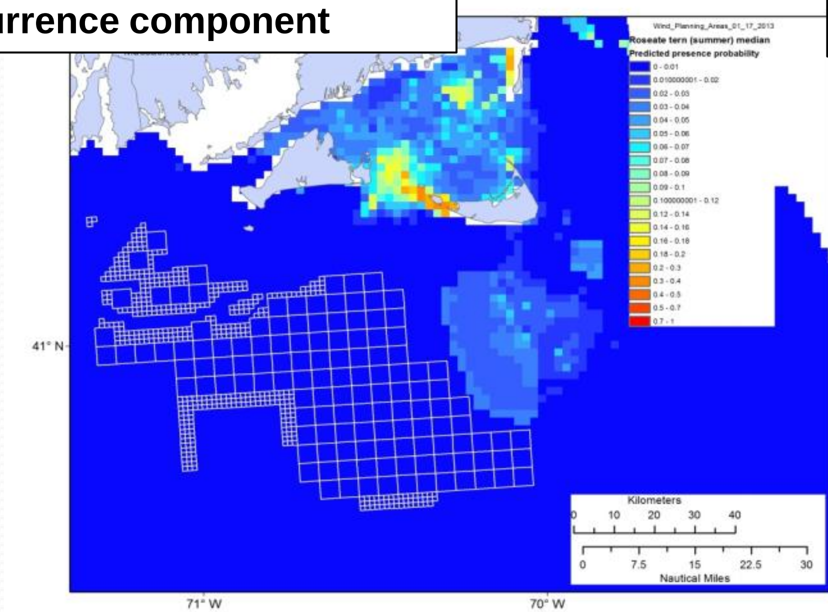
White-winged scoter (winter)
Abundance prediction



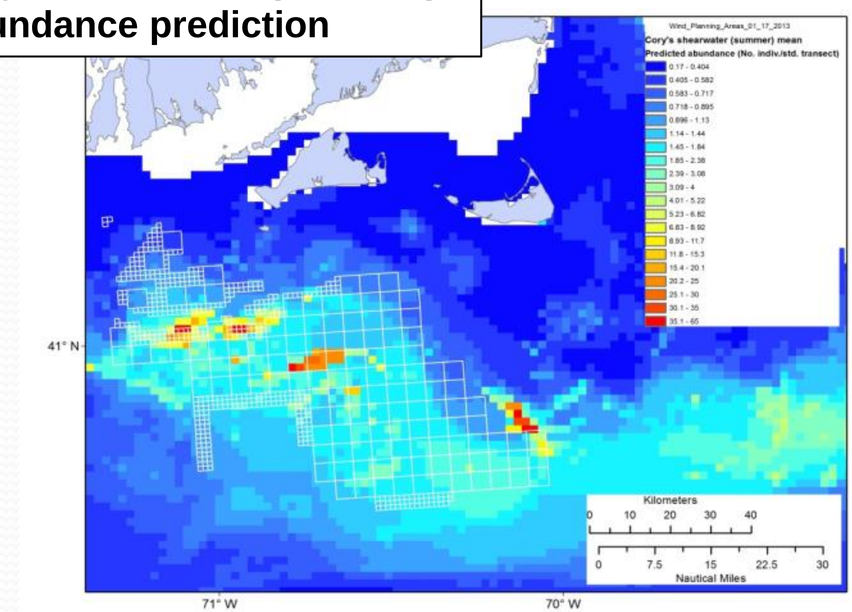
Common eider (winter)
Abundance prediction



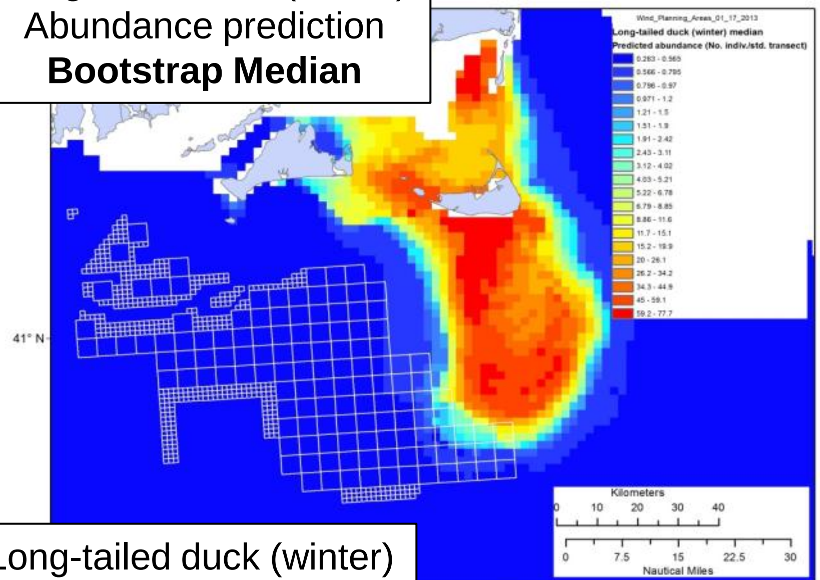
Roseate tern (summer)
Occurrence component



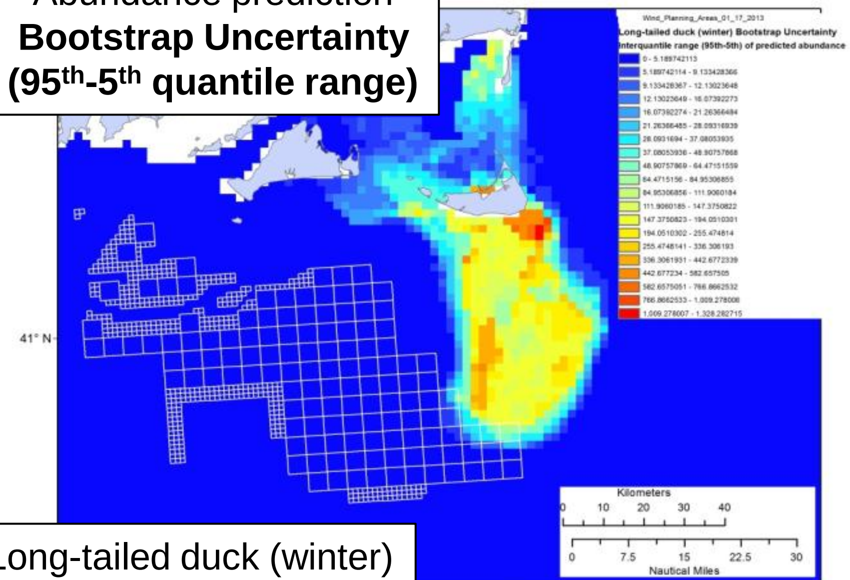
Cory's shearwater (summer)
Abundance prediction



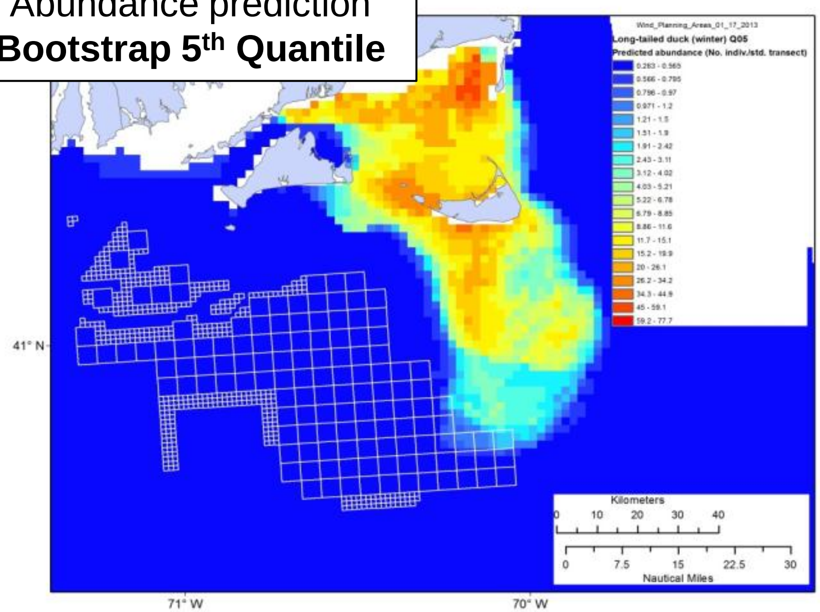
Value of model uncertainty assessment



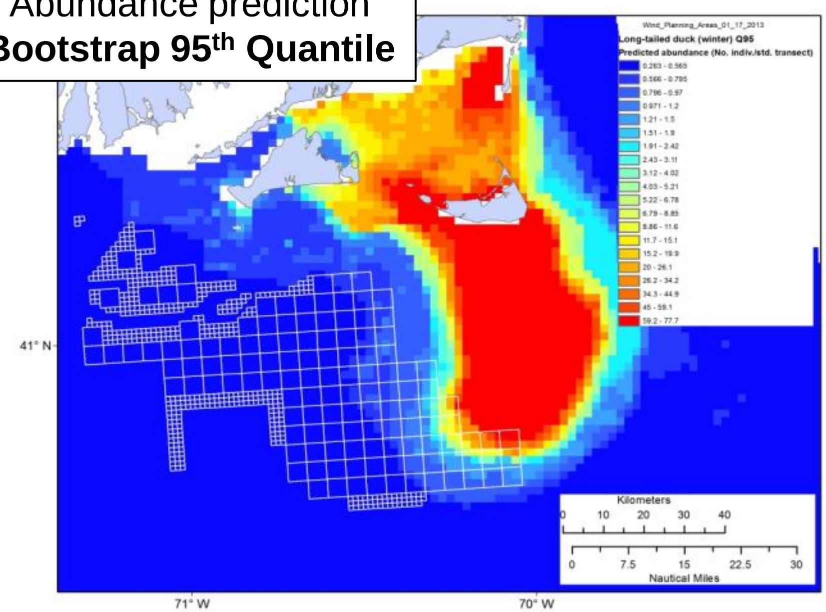
Long-tailed duck (winter)
Abundance prediction
Bootstrap Median



Long-tailed duck (winter)
Abundance prediction
Bootstrap Uncertainty
(95th-5th quantile range)

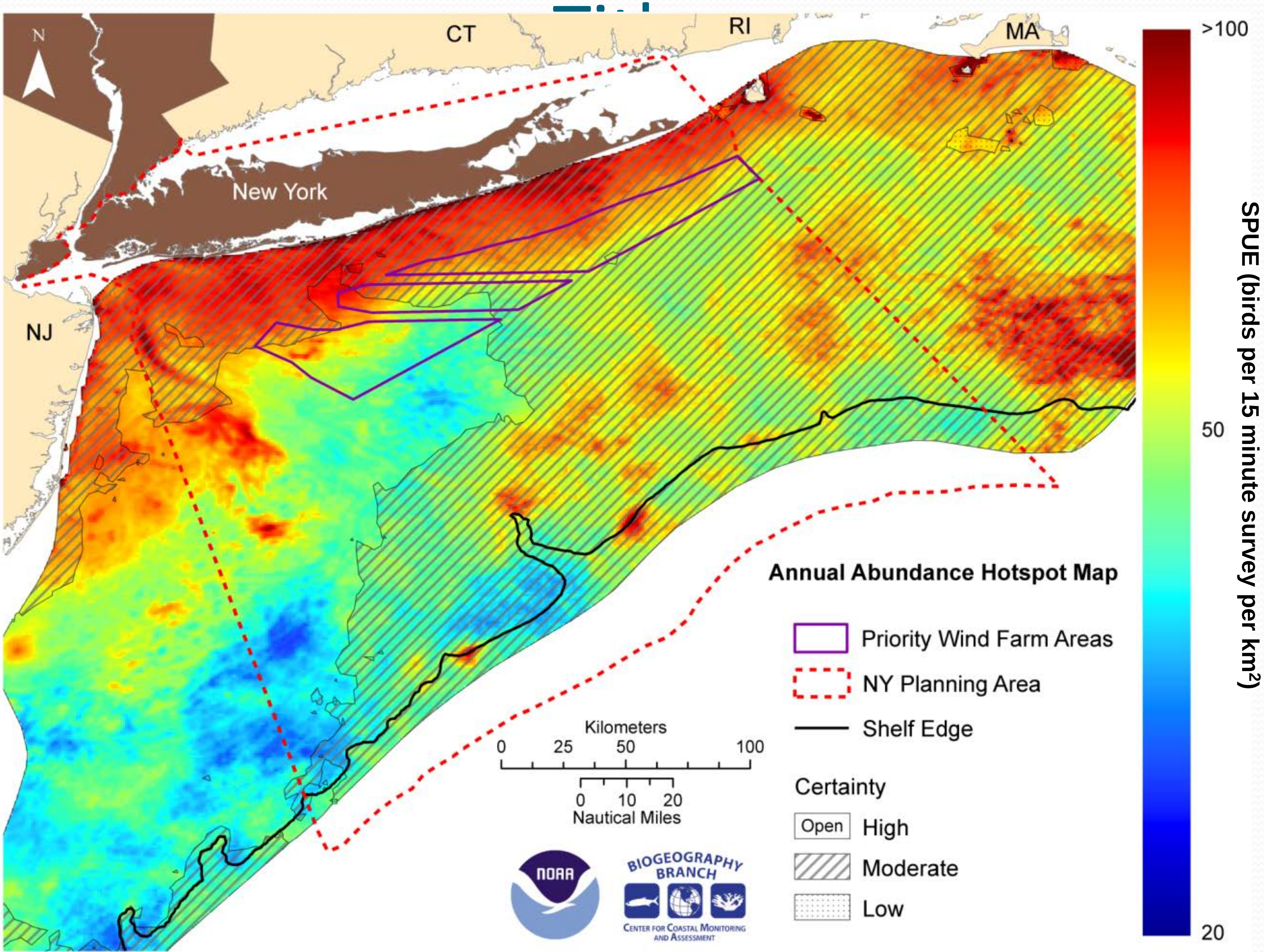


Long-tailed duck (winter)
Abundance prediction
Bootstrap 5th Quantile

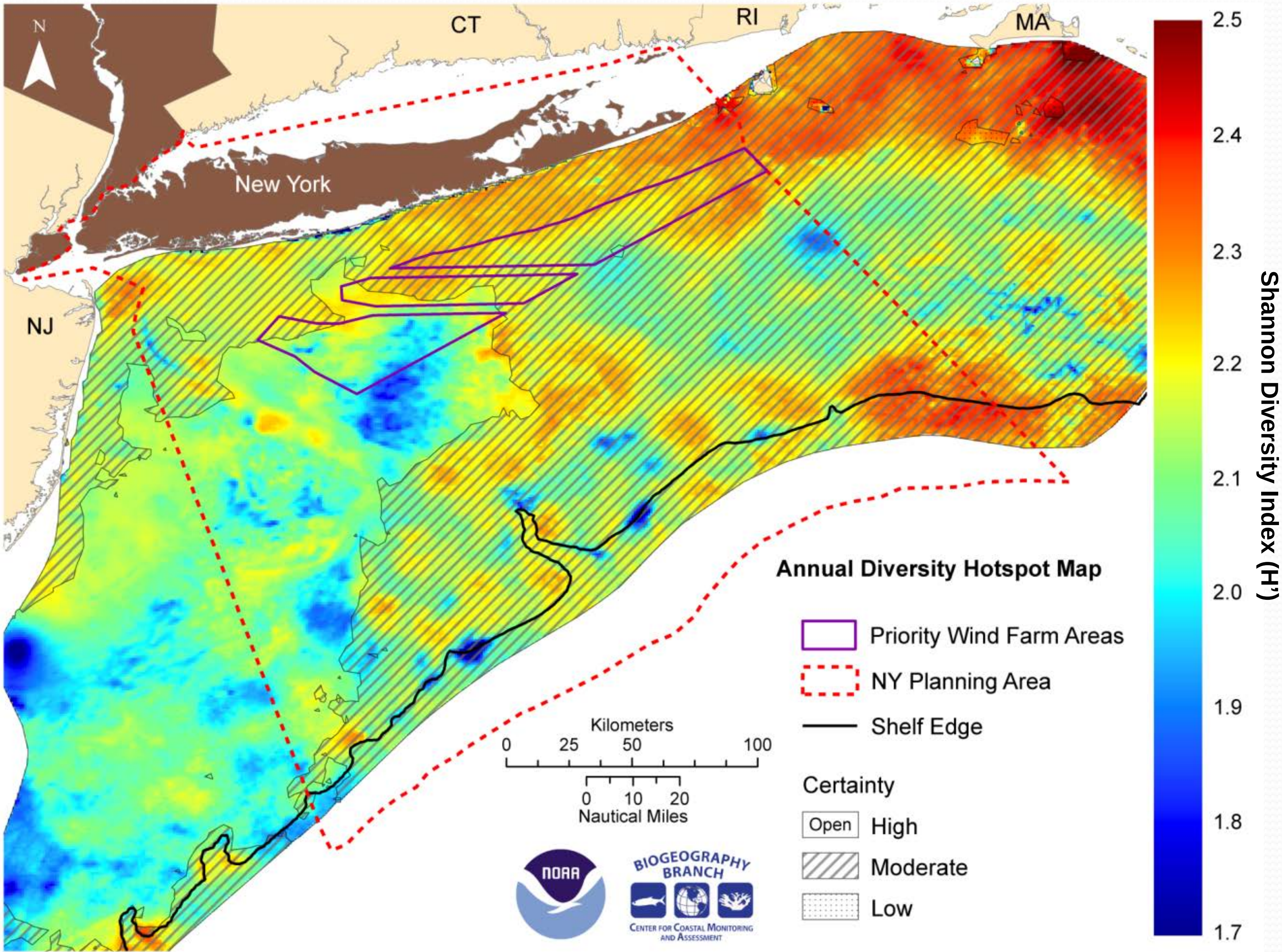


Long-tailed duck (winter)
Abundance prediction
Bootstrap 95th Quantile

Synthetic map products: abundance hotspots



Synthetic map products: diversity hotspots



Next steps for risk assessment

1. What species of birds are present in the vicinity of a wind farm and how many? (**Exposure**)

2. What is the per capita probability of an adverse effect of wind farms on birds of a given species, given that they are present in the area? (**Species-Specific Hazard** or **Sensitivity**)

3. How much are the potential adverse impacts from combining (1) and (2) likely to impact the population of each species, given its current status, trends, and ecological traits? (**Population Vulnerability**)

Mapping the distribution, abundance and risk assessment of marine birds in the Northwest Atlantic

Earvin Balderama[1,2], Beth Gardner[1], Brian Reich[2]

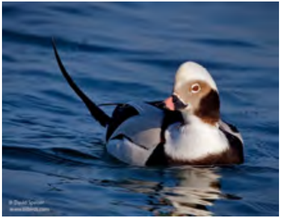
[1]Dept. of Forestry and Environmental Resources
North Carolina State University

[2]Dept. of Statistics
North Carolina State University

August 8, 2013

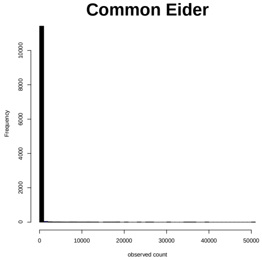
Outline

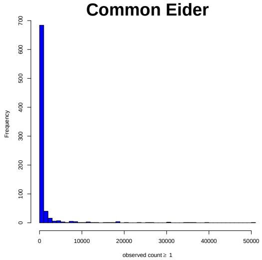
- Motivation
- Seabird data
    - **Common Eider**
    - **Greater Shearwater**
    - Herring Gull
    - Long-tailed Duck
    - Northern Gannett
    - Wilson's Storm Petrel
- Models and Methods
- Results
- Conclusion



Motivation

- Development of offshore wind energy facilities
  - Increased mortality from wind turbine collisions
  - Effects of anthropogenic activities
  - Altered habitat, source of food
- Statistical motivation
  - Distribution & relative abundance
  - Spatial heterogeneity
  - High-risk areas

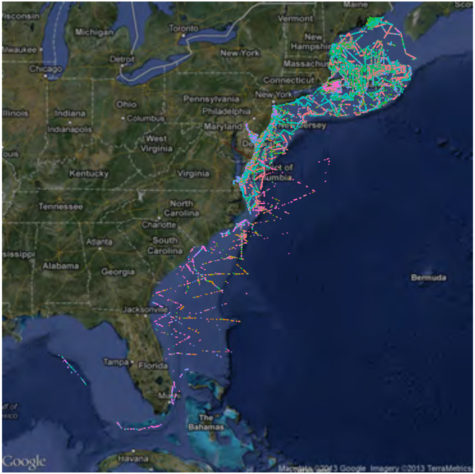




Data Collection

Boat and aerial transects from 1992 – 2010

- 43,701 transects
- 133,890 separate sightings
- > 2 million total birds
- $\sim$ 150 unique species



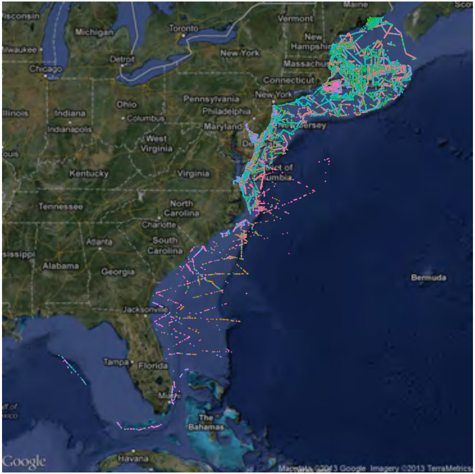
Bird Sightings

Boat and aerial transects from 1992 – 2010

- 43,701 transects
- 133,890 separate sightings
- $>$ 2 million total birds
- $\sim$ 150 unique species

Observer information

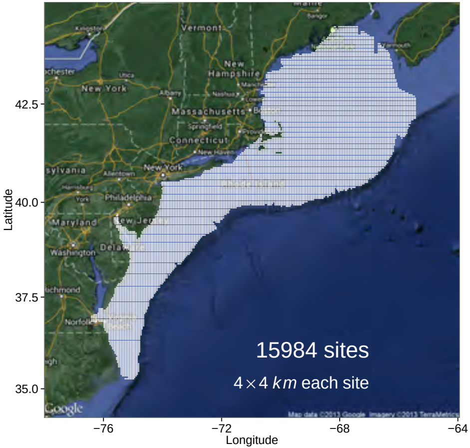
- Species
- Count
- Longitude, Latitude
- Date, Time



Spatial domain

Space-time window

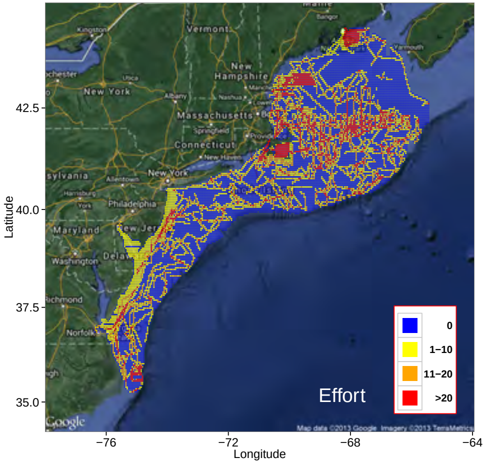
- 15,984 sites
- $4km \times 4km$ each site
- July 2002 - November 2010
- Group data by calendar month ***(MON)***



Effort

Space-time window

- 15,984 sites
- $4km \times 4km$ each site
- July 2002 - November 2010
- Group data by calendar month ***(MON)***



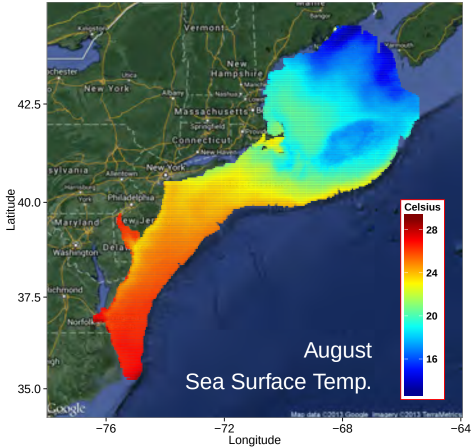
Covariates

Space-time window

- 15,984 sites
- $4km \times 4km$ each site
- July 2002 - November 2010
- Group data by calendar month ***(MON)***

Site-level variables

- Sea surface temp. ***(SST)***



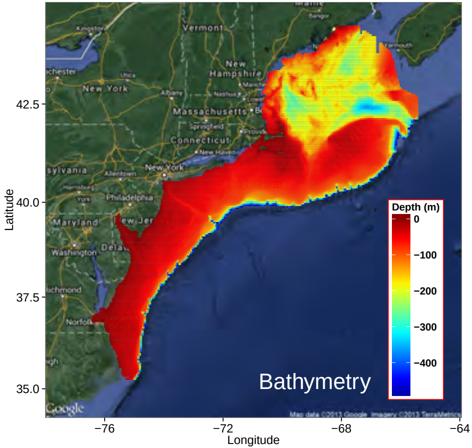
Covariates

Space-time window

- 15,984 sites
- $4km \times 4km$ each site
- July 2002 - November 2010
- Group data by calendar month ***(MON)***

Site-level variables

- Sea surface temp. ***(SST)***
- Bathymetry ***(BATH)***



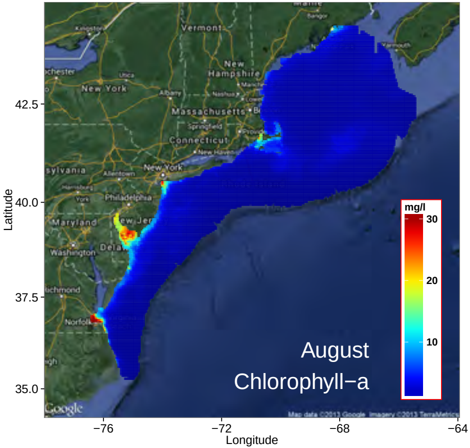
Covariates

Space-time window

- 15,984 sites
- $4km \times 4km$ each site
- July 2002 - November 2010
- Group data by calendar month ***(MON)***

Site-level variables

- Sea surface temp. ***(SST)***
- Bathymetry ***(BATH)***
- Chlorophyll-a ***(CHL)***



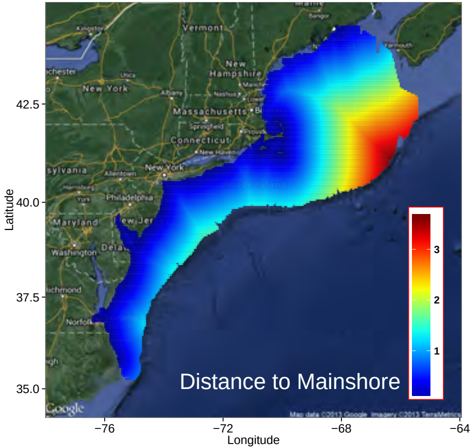
Covariates

Space-time window

- 15,984 sites
- $4km \times 4km$ each site
- July 2002 - November 2010
- Group data by calendar month ***(MON)***

Site-level variables

- Sea surface temp. ***(SST)***
- Bathymetry ***(BATH)***
- Chlorophyll-a ***(CHL)***
- Distance to shoreline ***(DTS)***



Common Eider data

Observed count	**Frequency**
0	10,736
$1-100$	537
$101-1000$	147
$1,001-10,000$	83
$10,001-50,000$	23
$50,001+$	1

Need a model that accounts for:

Zero-inflation

Extreme counts

Everything in between

Mixture model

Zero-Inflated Negative Binomial (ZINB)

$$
f(y|m,r) = \begin{cases} \pi + (1-\pi)\cdot\left(\dfrac{r}{r+m}\right)^r & \text{if } y = 0, \\[2ex] (1-\pi)\cdot\dfrac{\Gamma(y+r)}{\Gamma(r)y!}\left(\dfrac{r}{r+m}\right)^r\left(\dfrac{m}{r+m}\right)^y & \text{if } y \geq 1 \end{cases}
$$

Mixture model

Zero-Inflated Negative Binomial (ZINB)

$$f(y|m,r) = \begin{cases} \pi + (1-\pi)\cdot\left(\frac{r}{r+m}\right)^r & \text{if } y = 0, \\ (1-\pi)\cdot\frac{\Gamma(y+r)}{\Gamma(r)y!}\left(\frac{r}{r+m}\right)^r\left(\frac{m}{r+m}\right)^y & \text{if } y \geq 1 \end{cases}$$

Generalized Pareto Distribution (GPD)

$$g(y|\mu,\sigma,\xi) = \frac{\sigma^{1/\xi}}{\left(\sigma+\xi\times(y-\mu)\right)^{\frac{1}{\xi}+1}}$$

Models

Pure Mixture Models

- ZINB, for example,
- Models *extra* zero-counts in addition to NB zeros.

Models

Pure Mixture Models

- ZINB, for example,
- Models *extra* zero-counts in addition to NB zeros.

Hurdle Model

- Works as a zero-inflated model.
- f is truncated at 0, and normalized to be a density.
- Zeros generated by a separate mechanism.

Models

Pure Mixture Models

- ZINB, for example,
- Models *extra* zero-counts in addition to NB zeros.

Hurdle Model

- Works as a zero-inflated model.
- f is truncated at 0, and normalized to be a density.
- Zeros generated by a separate mechanism.

Double-Hurdle Model

- f is truncated at 0 and censored at $\boldsymbol{\mu}$
- Each component generated by a separate mechanism (density function).

Double-Hurdle model

$$P(\mathbf{y}|\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } \mathbf{y} = 0, \\ f(\mathbf{y}|\boldsymbol{\theta}) \sim \mathsf{NB}(\mathbf{m}, \mathbf{r}) & \text{if } 1 \leq \mathbf{y} < \boldsymbol{\mu}, \\ g(\mathbf{y}|\boldsymbol{\theta}) \sim \mathsf{GPD}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) & \text{if } \mathbf{y} \geq \boldsymbol{\mu}, \end{cases}$$

$\boldsymbol{\theta}$ represents the set of model parameters,

f is truncated at 0 and censored at $\boldsymbol{\mu}$, and normalized,

Double-Hurdle model

$$P(\mathbf{y}|\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } \mathbf{y} = 0, \\ f(\mathbf{y}|\boldsymbol{\theta}) \sim \mathsf{NB}(\mathbf{m}, \mathbf{r}) & \text{if } 1 \leq \mathbf{y} < \boldsymbol{\mu}, \\ g(\mathbf{y}|\boldsymbol{\theta}) \sim \mathsf{GPD}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) & \text{if } \mathbf{y} \geq \boldsymbol{\mu}, \end{cases}$$

$\boldsymbol{\theta}$ represents the set of model parameters,

f is truncated at 0 and censored at $\boldsymbol{\mu}$, and normalized,

Each component can be modeled separately.

Double-Hurdle model

Full data log-likelihood

$$\ell = \sum \log \Big[\mathbf{p}_Z \cdot \mathbb{1}_{[\mathbf{y}=0]} + (1-\mathbf{p}_E)\cdot(1-\mathbf{p}_Z)\cdot f(\mathbf{y}|\mathbf{m},\mathbf{r})\cdot \mathbb{1}_{[1\le \mathbf{y}<\boldsymbol{\mu}]} + \mathbf{p}_E \cdot (1-\mathbf{p}_Z)\cdot g(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi})\cdot \mathbb{1}_{[\mathbf{y}\ge\boldsymbol{\mu}]} \Big]$$

where $\mathbb{1}$ is the indicator function,

and the parameters that depend on site-level and spatial covariates are

- $\mathbf{p}_Z$ = Pr(site contains a Zero-count)
- $\mathbf{p}_E$ = Pr(site contains an Extreme-count)
- $\mathbf{m}$ = mean of Typical-count distribution.

Spatial random effects

Represent a model parameter θ through some link function Φ in a general spatial regression model

$$\Phi(\theta) = \mathbf{X}\boldsymbol{\beta} + \mathbf{S} + \mathbf{e}$$

- $\mathbf{S}$ is a Guassian process that accounts for spatial autocorrelation.

Spatial random effects

Represent a model parameter θ through some link function Φ in a general spatial regression model

$$\Phi(\theta) = \mathbf{X}\boldsymbol{\beta} + \mathbf{S} + \mathbf{e}$$

- $\mathbf{S}$ is a Guassian process that accounts for spatial autocorrelation.

Use the Guassian Markov Random Field (GMRF) prior distribution

$$\pi(\mathbf{S}|\tau) \propto \tau^{\text{rank}(\mathbf{Q})/2} \exp\left(-\frac{\tau}{2}\mathbf{S}'\mathbf{Q}\mathbf{S}\right)$$

- $\mathbf{Q} = \mathbf{D} - \rho\mathbf{A}$ is the CAR precision and τ is a smoothing parameter.
- $\mathbf{D}$ is a diagonal matrix with entries the number of each site's neighbors,
- $\mathbf{A}$ is the adjacency matrix.
- Let $\rho = 1$ to specify the intrinsic CAR prior.

Dimension reduction

$$\mathrm{logit}(\mathbf{p}_Z) \approxeq \mathbf{X}\boldsymbol{\beta}_Z + \mathbf{S}$$
$$\log(\mathbf{m}) \approxeq \mathbf{X}\boldsymbol{\beta}_T + \mathbf{S}$$
$$\mathrm{logit}(\mathbf{p}_E) \approxeq \mathbf{X}\boldsymbol{\beta}_E + \mathbf{S}$$

Dimension reduction

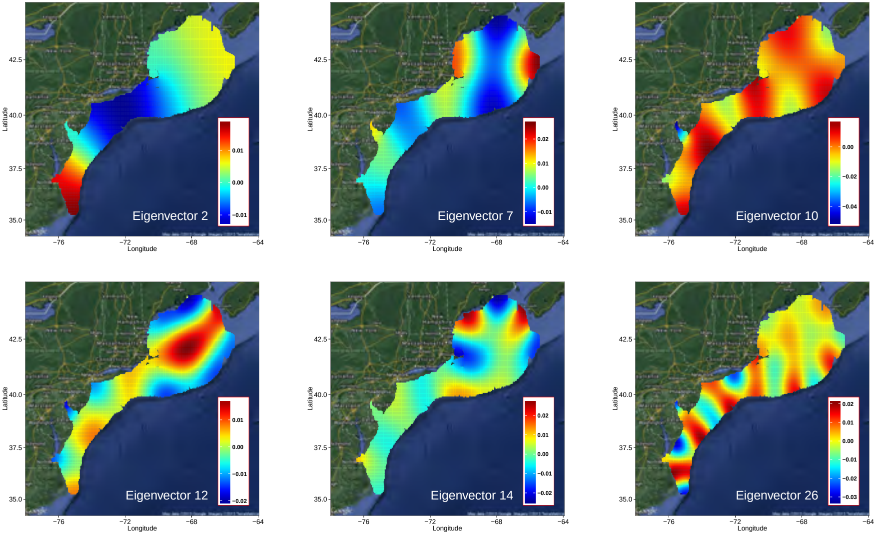
$$\text{logit}(\mathbf{p}_Z) \approxeq \mathbf{X}\boldsymbol{\beta}_Z + \mathbf{S}$$
$$\log(\mathbf{m}) \approxeq \mathbf{X}\boldsymbol{\beta}_T + \mathbf{S}$$
$$\text{logit}(\mathbf{p}_E) \approxeq \mathbf{X}\boldsymbol{\beta}_E + \mathbf{S}$$

With 15984 sites, want to reduce dimensionality of $\mathbf{S}$.

$$\mathbf{S} = \underset{n\times n}{\mathbf{V}} \cdot \underset{n\times 1}{\boldsymbol{\alpha}} \approx \underset{n\times q}{\mathbf{V}} \cdot \underset{q\times 1}{\boldsymbol{\alpha}}$$

$\mathbf{V}$ is the set of (orthonormal) eigenvectors from the decomposition of $\mathbf{Q}$,
$\boldsymbol{\alpha}$ is the vector of coefficients, and $q \ll n$
$q = 50$ in the following analyses.

Greater Shearwater significant spatial effects for **m**



Parameter Estimation

Uninformative Priors

$\boldsymbol{\beta}, \alpha \overset{iid}{\sim} \mathsf{Normal}(0, 1000000)$

$\boldsymbol{\tau} \sim \mathsf{Gamma}(0.5, 0.005)$

$r, \sigma \sim \text{log-}\mathsf{Normal}(5, 100)$

$\xi \sim \mathsf{Normal}(0, 1)$

μ set to a fixed constant.

Updated one-at-a-time

Gibbs updates for $\boldsymbol{\tau}$

Metropolis-Hastings updates for all others.

Significant predictors

Common Eider

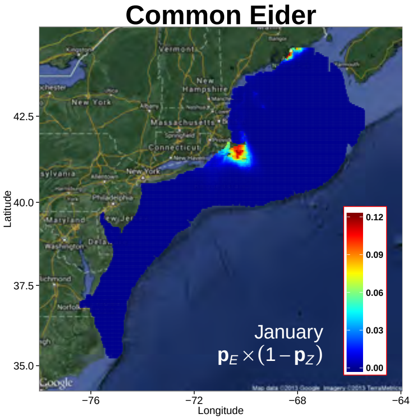
Covariate	**Parameter** $\mathbf{p}_Z$	$\mathbf{m}_T$	$\mathbf{p}_E$
Sea surface temp	+		–
Bathymetry		+	
Chlorophyll-a	–	+	
Distance to shore	+	+	
Month			
Month^2	–		

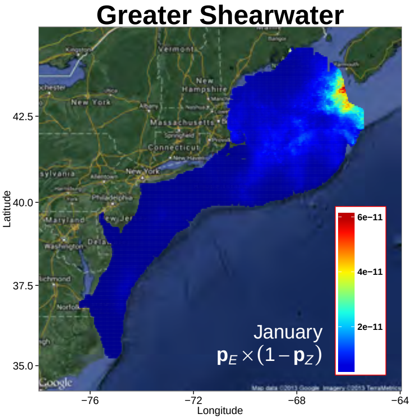
where +/– indicates positive/negative relationship.

Greater Shearwater

Covariate	**Parameter** $\mathbf{p}_Z$	$\mathbf{m}_T$	$\mathbf{p}_E$
Sea surface temp	+	–	–
Bathymetry	–		
Chlorophyll-a	+		
Distance to shore	–		
Month	–	+	+
Month^2	+	–	–

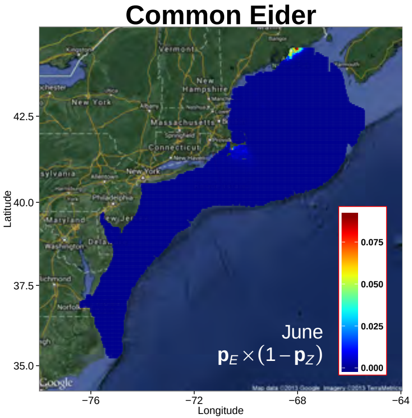
Extreme-Count Risk Map: January

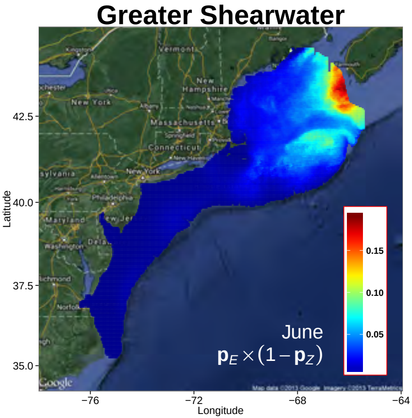




Red: higher probability of an extreme count.
Blue: lower probability.

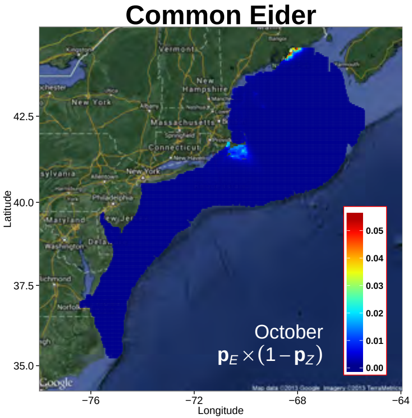
Extreme-Count Risk Map: June

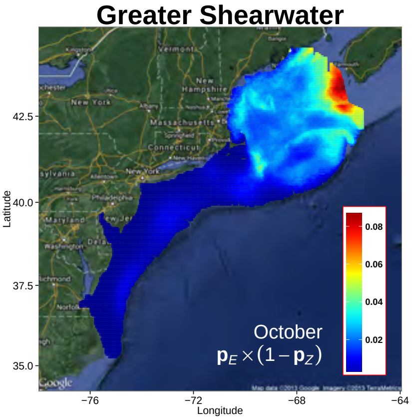




Red: higher probability of an extreme count.
Blue: lower probability.

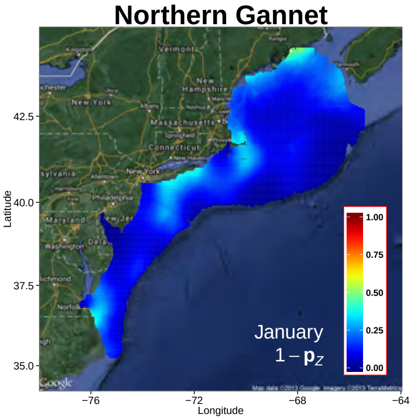
Extreme-Count Risk Map: October





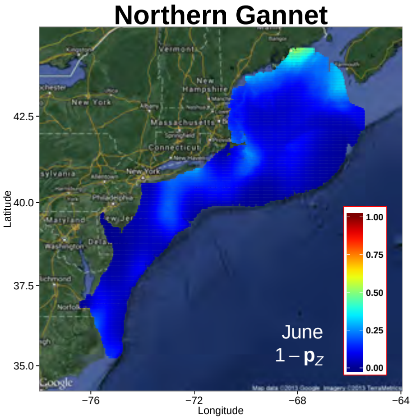
Red: higher probability of an extreme count.
Blue: lower probability.

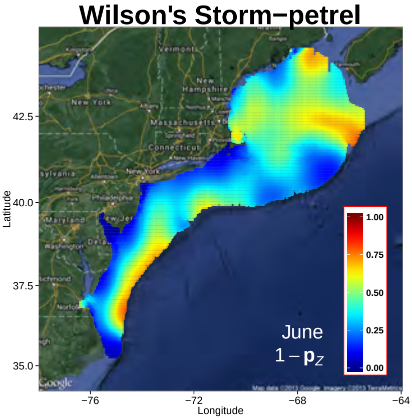
Non-Zeros Risk Map: January



Red: higher probability of a positive count.
Blue: lower probability.

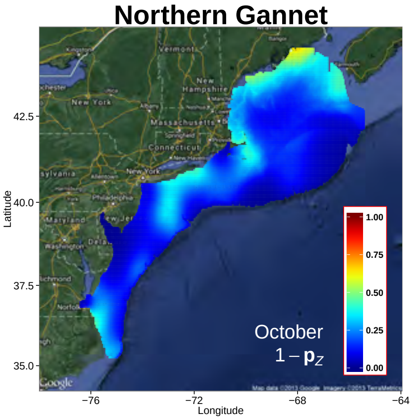
Non-Zeros Risk Map: June

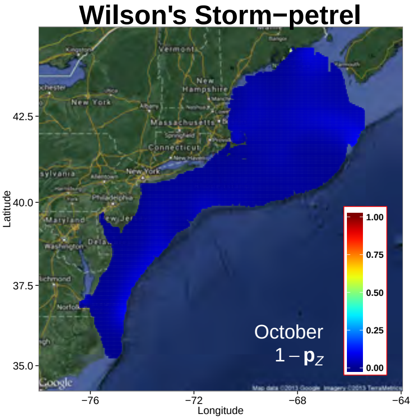




Red: higher probability of a positive count.
Blue: lower probability.

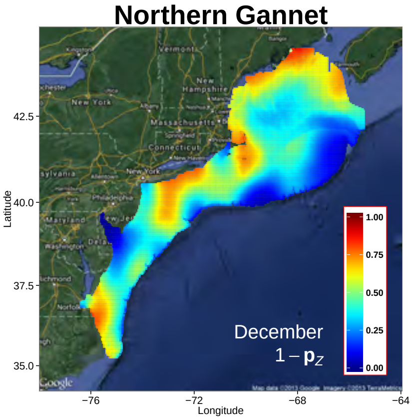
Non-Zeros Risk Map: October

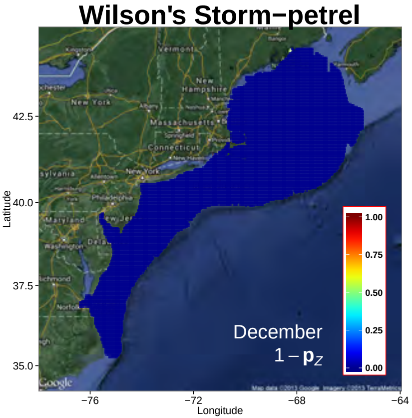




Red: higher probability of a positive count.
Blue: lower probability.

Non-Zeros Risk Map: December





Red: higher probability of a positive count.

Blue: lower probability.

Model Summary

Double-Hurdle model for zero-inflation and extreme counts.

Dimension reduction on spatial random effects.

Easy to interpret and map risk with $\mathbf{p}_Z, \mathbf{m}, \mathbf{p}_E$.

Future Work

- Other mixtures of distributions
- Data without "Extremes"
- Model comparison/diagnostics using...DIC?
- Automatic extremes threshold estimation
- Multiple hurdles

Thank you!

Acknowledgements

North Atlantic Landscape Conservation Cooperative (NALCC)

Wildlife Management Institute (WMI)

Allison Sussman (USGS Patuxent Wildlife Research Center)

Mark Wimer (USGS Patuxent Wildlife Research Center)

References

Menza, C., Kinlan, B. P., Dorfman, D. S., Poti, M., and Caldow, C. (2012), "A Biogeographic Assessment of Seabirds, Deep Sea Corals and Ocean Habitats of the New York Bight: Science to Support Offshore Spatial Planning," NOAA Technical Memorandum NOS NCCOS 141, NOAA National centers for coastal Ocean Science, Silver Spring, MD.

O'Connell, A., Gardner, B., Gilbert, A., and Ellis, J. (2011), "Compendium of Avian Occurrence Information for the Continental Shelf Waters along the Atlantic Coast of the United States," Tech. rep., U.S. Fish and Wildlife Service (FWS).

Zipkin, E. F., Gardner, B., Gilbert, A. T., O'Connell, A. F., Royle, J. A., and Silverman, E. D. (2010), "Distribution patterns of wintering sea ducks in relation to the North Atlantic Oscillation and local environmental characteristics," *Oecologia*, 163, 893–902.