Mapping the distribution of marine birds using a space-time double-hurdle model

Earvin Balderama (LUC), Beth Gardner (NCSU), Brian Reich (NCSU)
Department of Mathematics & Statistics, Loyola University Chicago, Chicago, IL, USA

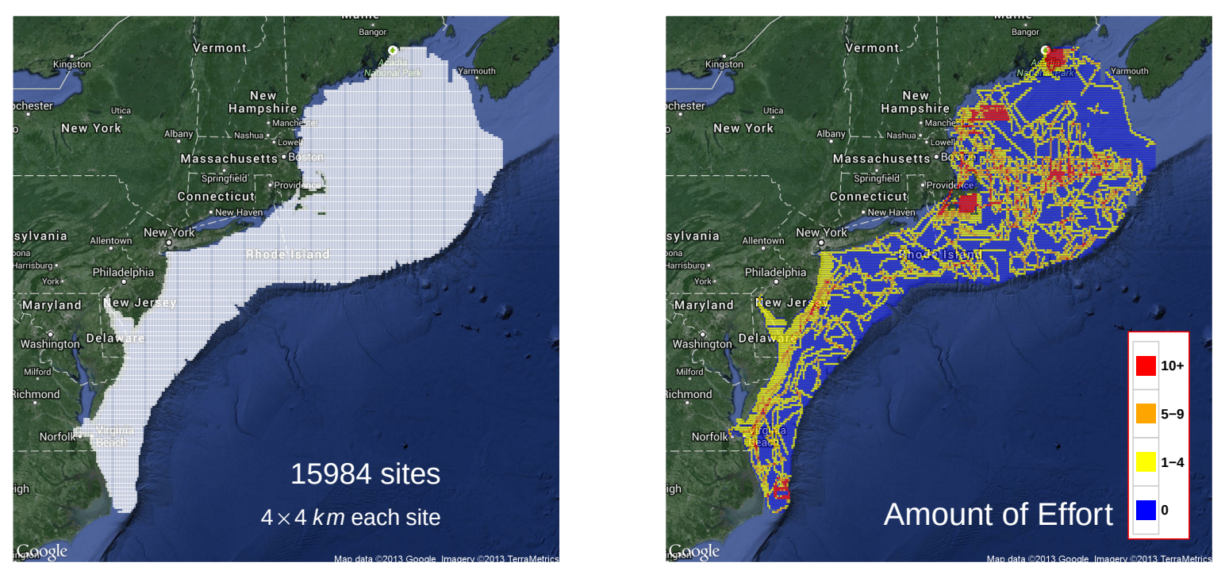
Motivation

- Construction of offshore wind energy facilities in U.S. Atlantic coast regions may impact marine life.

- **Goal:** Develop spatial-temporal models to assess avian distribution and abundance, and create maps to identify sensitive and high-use areas in need of protection.

Data

- Surveys: Boat/aerial continuous-time strip transects.
- Space: 15,984 4km×4km pixels.
- Time: July 2002—November 2010.
- E_{ij} = Effort (# surveys intersecting pixel i, month j)
- y_{ij} = Count (# individual birds in pixel i, month j)



Example:
Northern Gannet data

Count	Frequency
0	9553
1 – 10	1778
11 – 100	184
101 – 1000	11
1001+	1

Model

- Double-Hurdle model accounts for both excessive zero-inflation and extreme over-dispersion.

Likelihood of observing y_{ij} birds in pixel i during the j^{th} month of the year:

$$f(y_{ij}|\boldsymbol{\theta}) = \begin{cases} p_{ij}, & y_{ij} = 0,\\ [1-p_{ij}]\cdot[1-q_{ij}]\cdot \mathrm{NB}(\mu_{ij}, r), & 1 \le y_{ij} < \psi,\\ [1-p_{ij}]\cdot q_{ij}\cdot \mathrm{GPD}(\psi, \sigma, \xi), & y_{ij} \ge \psi. \end{cases}$$

- Negative binomial (NB) for small, "typical" counts.
  - left-truncated at 0 and right-truncated at a fixed ψ.
- Generalized Pareto (GPD) for large, right-tail counts.
  - GPD density is > 0 at threshold ψ or above.

Spatial Hierarchical Regression

- Can create (monthly/yearly) maps using estimates of:

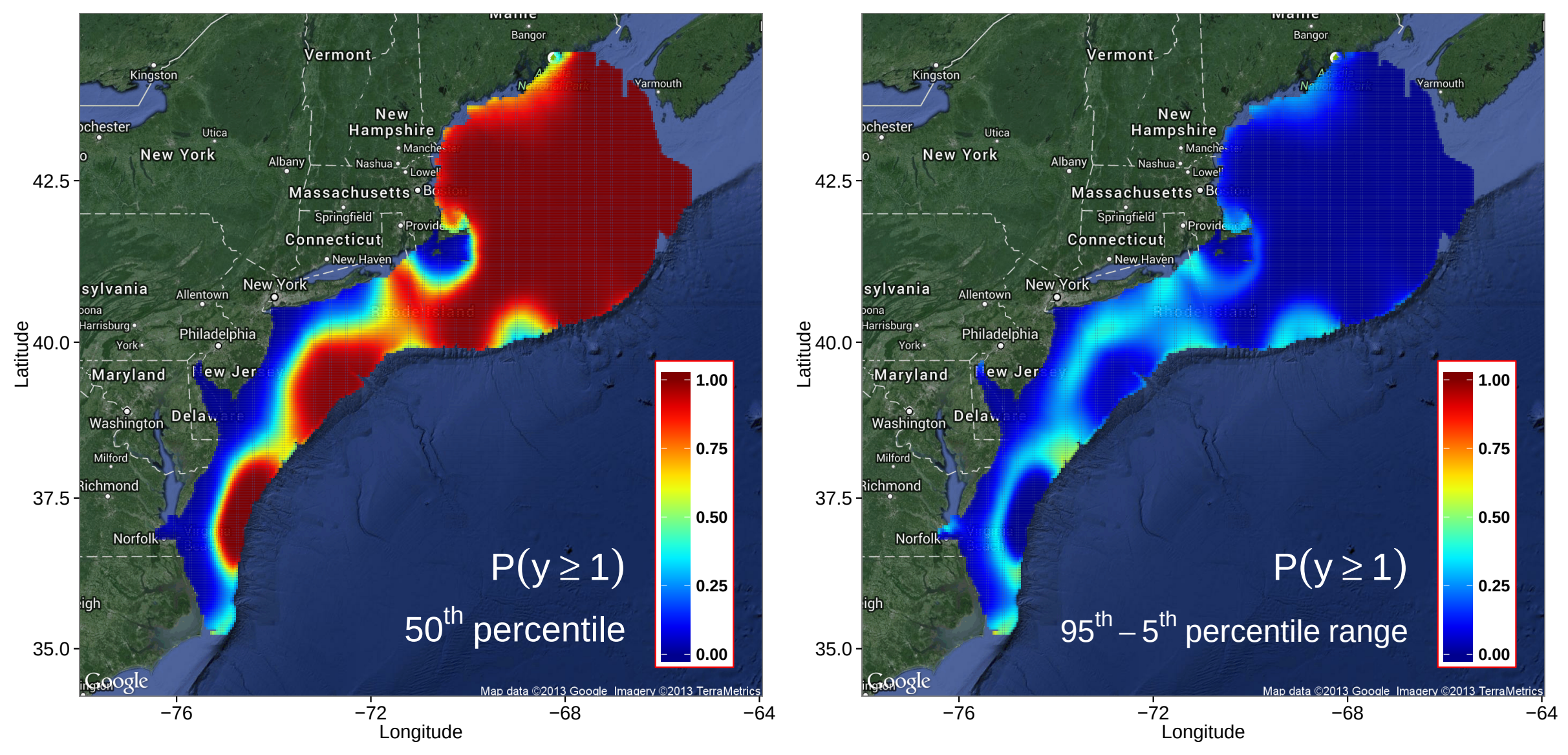
  $\mathbf{p}$ = *P(zero-count)*
  $$\mathrm{logit}(\mathbf{p}) = \mathbf{X}\boldsymbol{\gamma} + \mathbf{S}$$
  $\boldsymbol{\mu}$ = *mean of typical-count distribution.*
  $$\log(\boldsymbol{\mu}) = \log(E) + \mathbf{X}\boldsymbol{\beta} + \mathbf{S}$$
  $\mathbf{q}$ = *P(large-count | nonzero-count)*
  $$\mathrm{logit}(\mathbf{q}) = \mathbf{X}\boldsymbol{\delta}$$
- Environmental covariates
  - $\mathbf{x}_1$ = Sea surface temperature.
  - $\mathbf{x}_2$ = Ocean depth.
  - $\mathbf{x}_3$ = Chlorophyll-a level.
  - $\mathbf{x}_4$ = Distance-to-shore.
- Temporal effects (Fourier basis)
  - $\mathbf{x}_5 = sin(\frac{\pi}{6}\cdot \mathrm{Month})$.
  - $\mathbf{x}_6 = cos(\frac{\pi}{6}\cdot \mathrm{Month})$.
- Spatial effects (Guassian Markov random field)
  - Dimension reduction of 15984×15984 inverse covariance matrix $\mathbf{Q}$ (intrinsic CAR prior).
  - $\mathbf{S} = \mathbf{V}\boldsymbol{\alpha}$, where $\mathbf{V}$: eigenvectors from $\mathbf{Q} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$

Predicted Annual Exposure Maps

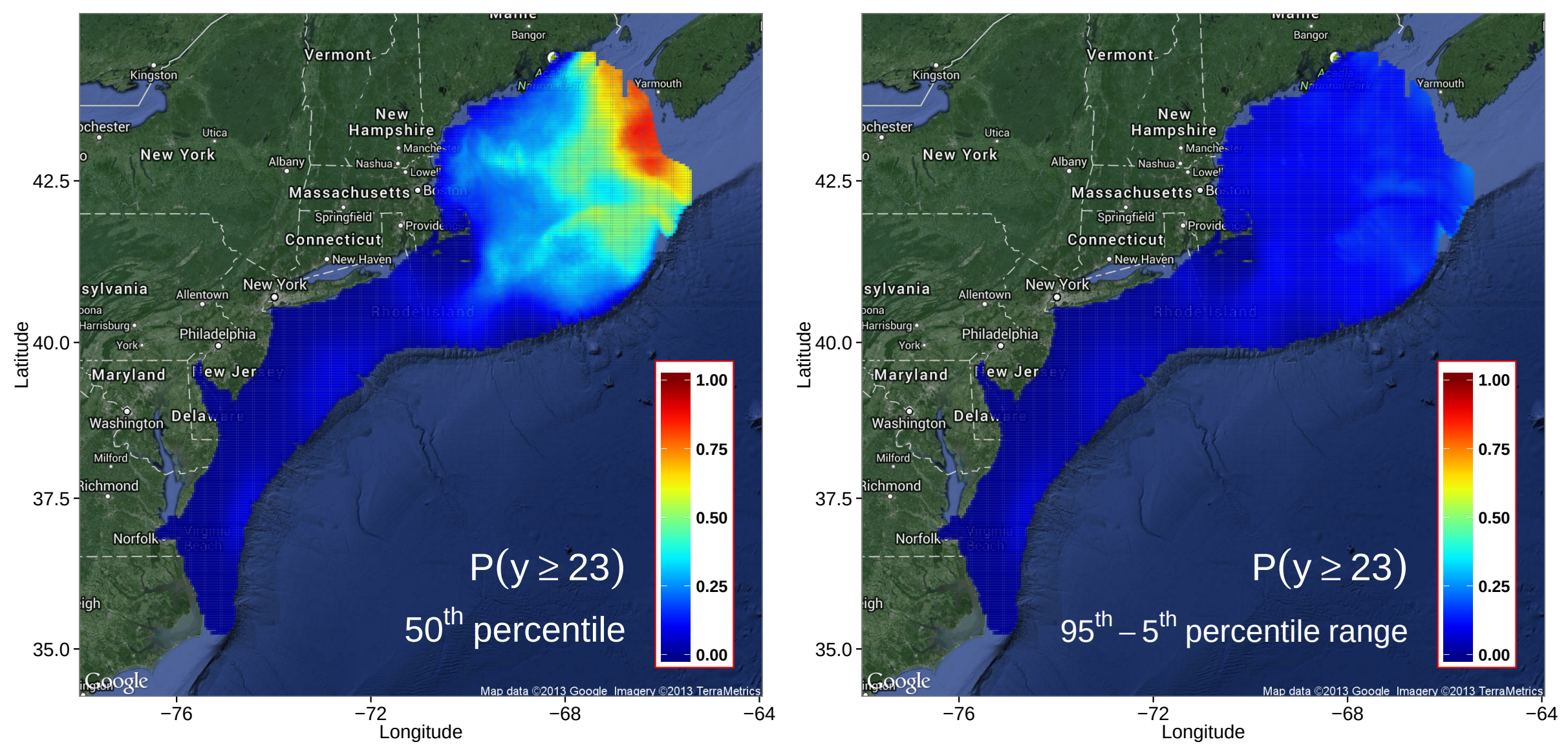
- Maps of exposure probability during a calendar year.
- Maps of uncertainty based on a 90% credible interval.

Great Shearwater

Occupancy probability ⟷ *Map of uncertainty*

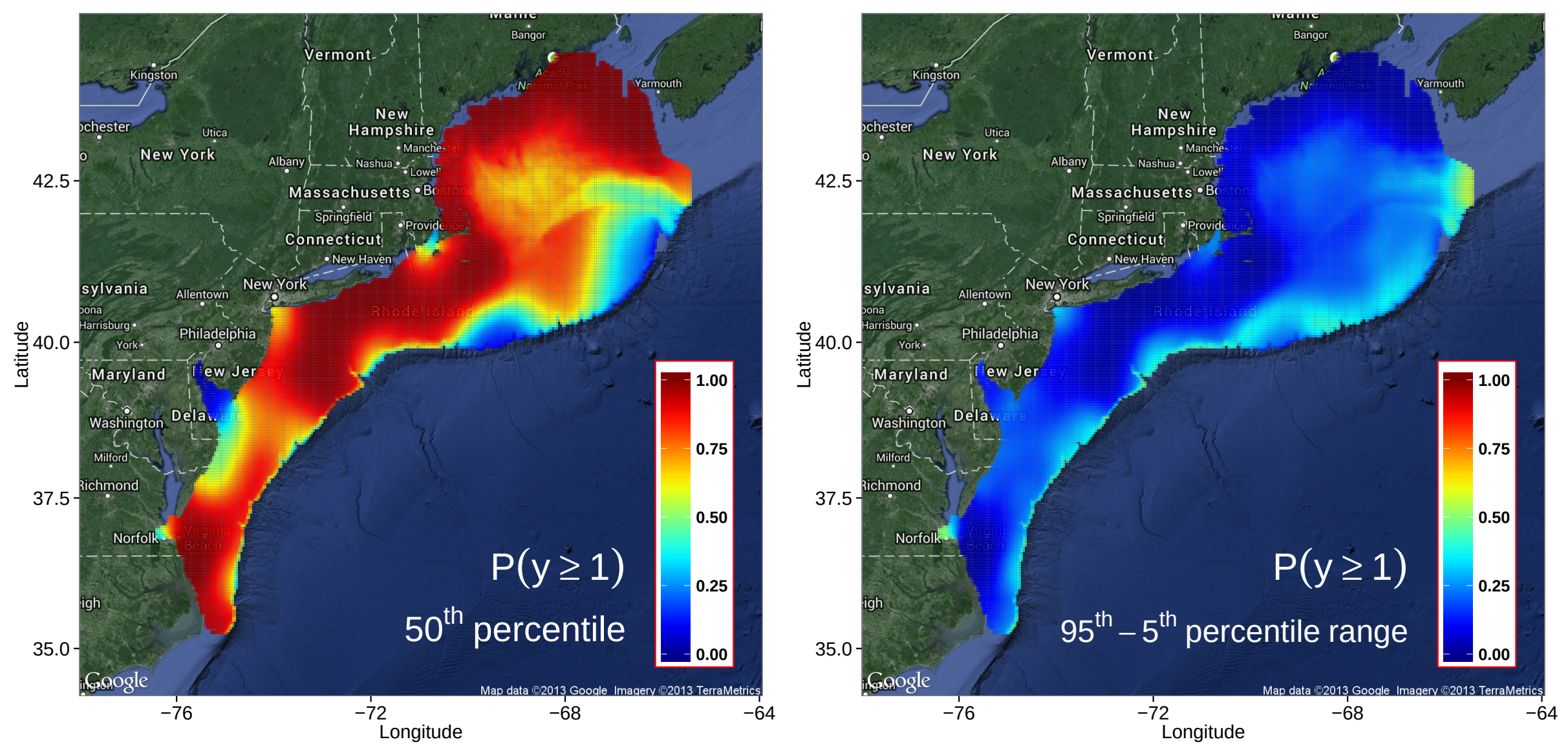


Large-count probability ⟷ *Map of uncertainty*

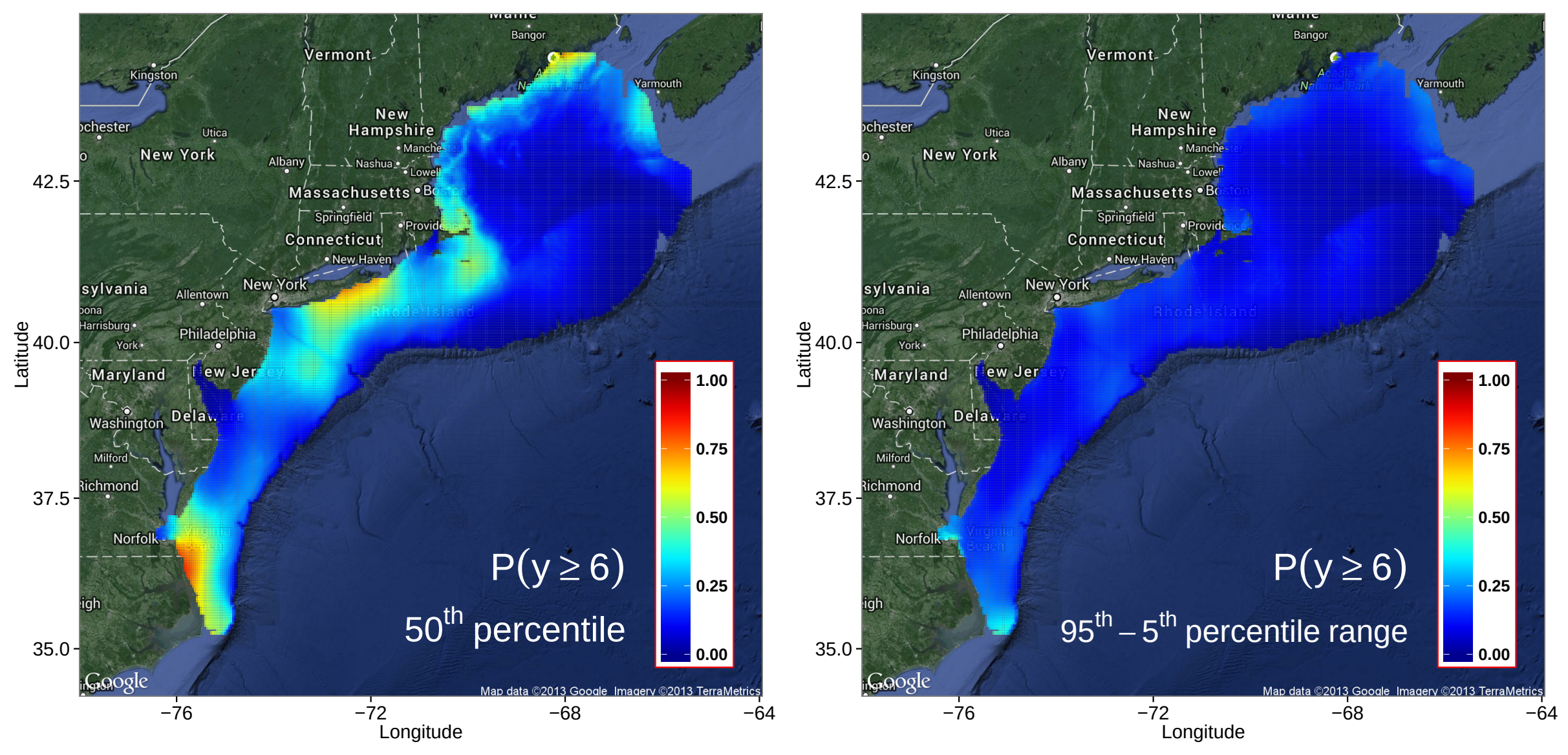


Northern Gannet

Occupancy probability ⟷ *Map of uncertainty*

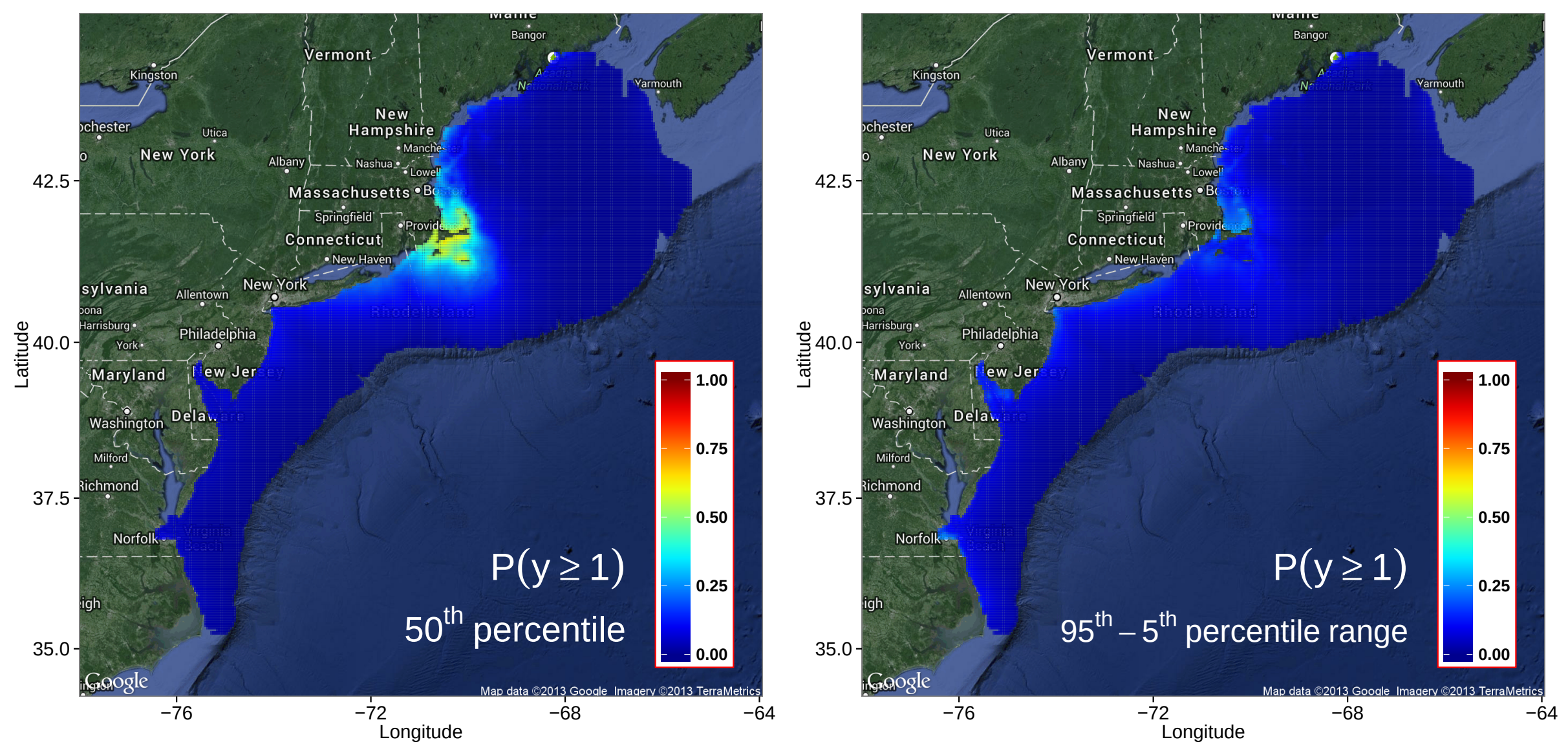


Large-count probability ⟷ *Map of uncertainty*



Roseate Tern

Occupancy probability ⟷ *Map of uncertainty*



Analyses

- Parameter estimation in a Bayesian MCMC framework.
- Threshold values considered:
  - $\psi = 1$ ⟸ single GPD-hurdle
  - $\psi = \{97.5th \text{ percentile}\}$
  - $\psi = \{99th \text{ percentile}\}$
  - $\psi = \infty$ ⟸ single NB-hurdle
- Considered models with & w/out spatial effect $\mathbf{S}$.
- Model comparison: DIC and LPML for goodness-of-fit.

Data Summary and Resulting Best Model

Species	#Obs.	Mean (SD)	Med	Max	Best Model (ψ)
Atlantic Puffin	248	2.0 (1.5)	1	10	Double-hurdle (5)
Black-capped Petrel	540	23.1 (56.0)	6	605	Double-hurdle (5)
Black Tern	738	4.9 (9.3)	2	105	Double-hurdle (5)
Bonaparte's gull	376	5.8 (18.3)	2	262	Double-hurdle (5)
Common Eider	1432	572.1 (3019.8)	15	50025	**GPD-hurdle (1)**
Common Loon	1319	3.3 (3.9)	2	40	Double-hurdle (7)
Common Tern	809	11.8 (51.7)	3	1094	Double-hurdle (5)
Cory's Shearwater	634	4.9 (16.9)	2	266	Double-hurdle (5)
Double-crested Cormorant	232	13.9 (43.9)	2	501	Double-hurdle (5)
Dovekie	550	7.6 (17.6)	3	299	Double-hurdle (7)
Great Black-backed Gull	3188	4.8 (25.8)	2	1300	Double-hurdle (7)
Great Shearwater	3195	12.2 (35.8)	4	950	Double-hurdle (23)
Herring Gull	3249	5.7 (31.0)	2	1300	Double-hurdle (8)
Laughing Gull	464	3.2 (6.3)	2	88	Double-hurdle (5)
Leach's Storm-petrel	840	6.2 (21.3)	2	345	Double-hurdle (5)
Long-tailed Duck	1443	94.0 (432.6)	17	11000	**GPD-hurdle (1)**
Northern Fulmar	1330	7.8 (43.2)	2	1352	Double-hurdle (5)
Northern Gannet	2248	6.1 (41.7)	2	1775	Double-hurdle (6)
Razorbill	1002	10.7 (19.7)	4	293	Double-hurdle (7)
Roseate Tern	196	7.1 (16.8)	2	137	**GPD-hurdle (1)**
Sooty Shearwater	729	9.4 (38.8)	2	700	Double-hurdle (6)
Surf Scoter	1135	60.2 (146.3)	15	1400	Double-hurdle (30)
White-winged Scoter	885	24.5 (73.1)	4	1027	Double-hurdle (7)
Wilson's Storm-petrel	1790	13.4 (92.2)	2	3061	Double-hurdle (6)

Conclusions

- Spatial models (with $\mathbf{S}$) fits better than non-spatial.
- ψ at 97.5*th* percentile fits better than 99*th* percentile.
- For most species, the double-hurdle model fits better than any single-hurdle model.
  - If the double-hurdle model is not the best-fit, then the GPD-hurdle is the best-fit.
- If considering only single-hurdle models, then GPD-hurdle fits better than NB-hurdle for most species.

Current Work & Future Considerations

- Investigate other distributions, i.e., log-normal models.
- Expand study area and incorporate new data.
- Treat threshold parameter ψ as unknown.
- Maps of "persistence", "vulnerability", ...
- Incorporate climate models.
- Point-process models.
- R Shiny app.

Acknowledgements

- Software used: R (`www.r-project.org`)
- Maps made with: ggmap package in R.
- Data acquired from: Avian Compendium (NOAA)
- Thanks: Allison Sussman, Mark Wimer, Brian Kinlan, and the following organizations:

`http://math.luc.edu/~ebalderama/`
`ebalderama@luc.edu`