Statistical analyses to support guidelines for marine avian sampling

Brian Kinlan (NOAA)
Elise F. Zipkin (USGS)
Allan F. O'Connell (USGS)
Chris Caldow (NOAA)
Allison Sussman (USGS)
Mark Wimer (USGS)

Special thanks to our NOAA Hollings Scholar, Diana Rypkema (Cornell University)

NOAA/NOS National Centers for Coastal Ocean Science (NCCOS)
USGS Patuxent Wildlife Research Center

Atlantic Marine Bird Conservation Cooperative
March 6, 2013

Objectives

Develop a framework for assessing:

1) which lease blocks are hotspots and coldspots
2) survey effort required to have sufficient statistical power to detect hotspots and coldspots

What is a hot/coldspot?

Hot spot = A lease block with an average species specific abundance that is **some multiple >1 (e.g., 3x)** the mean of the region

Cold spot = A lease block with an average species specific abundance that is **some multiple <1 (e.g., 1/3x)** the mean of the region

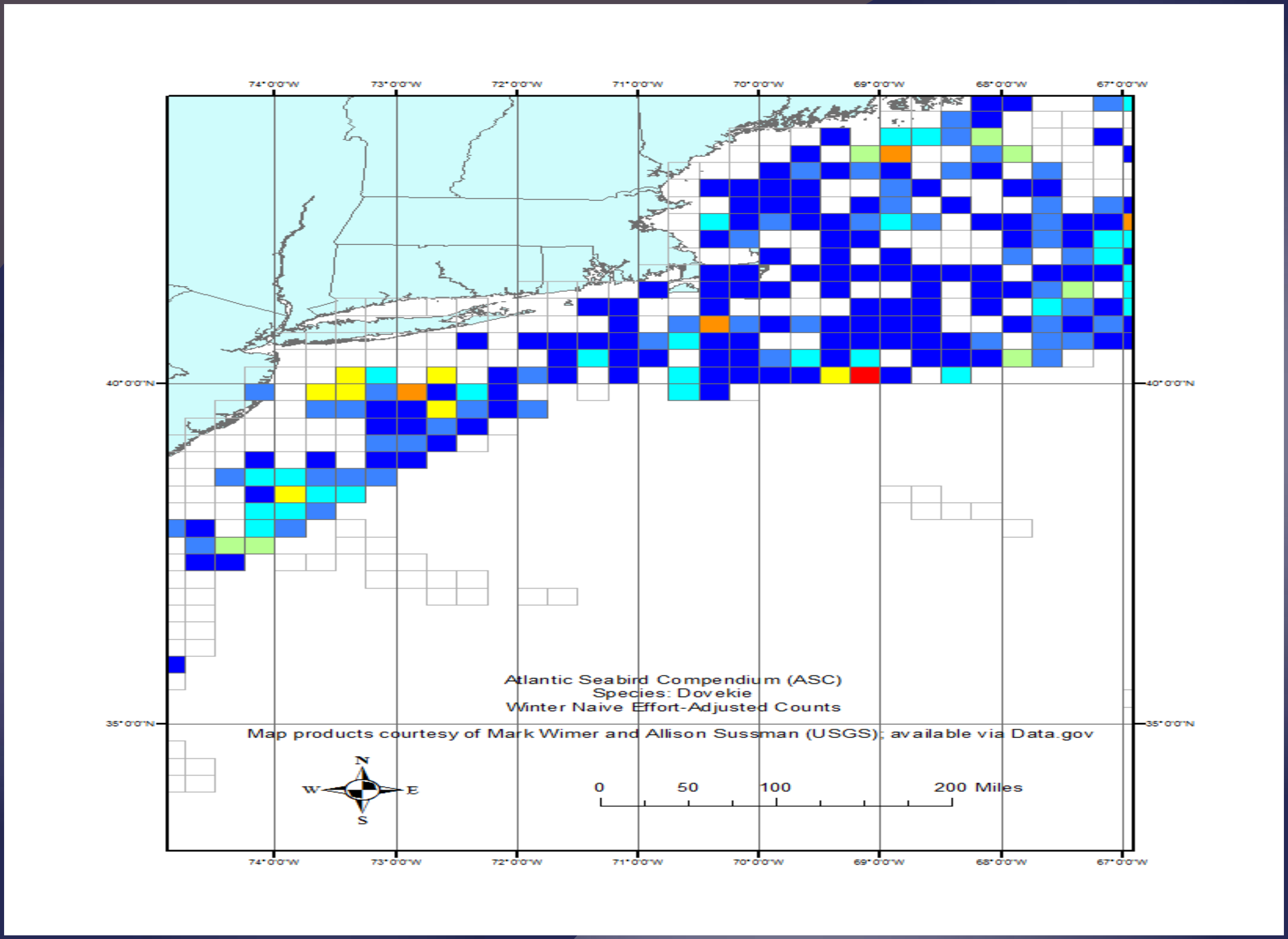


Figure 1. Example summarized historical seabird survey data, illustrating the characteristic statistical noisiness of seabird data. Determining which of the apparent "hotspots" (or "coldspots") are statistically significant is impossible without knowing the number of independent surveys that were conducted at each location. The purpose of this study is to develop guidelines for determining when a grid cell has been adequately sampled so that the relative abundance index (e.g, effort adjusted counts, as shown here) can be reliably compared to other well-sampled grid cells.

N surveys = 1

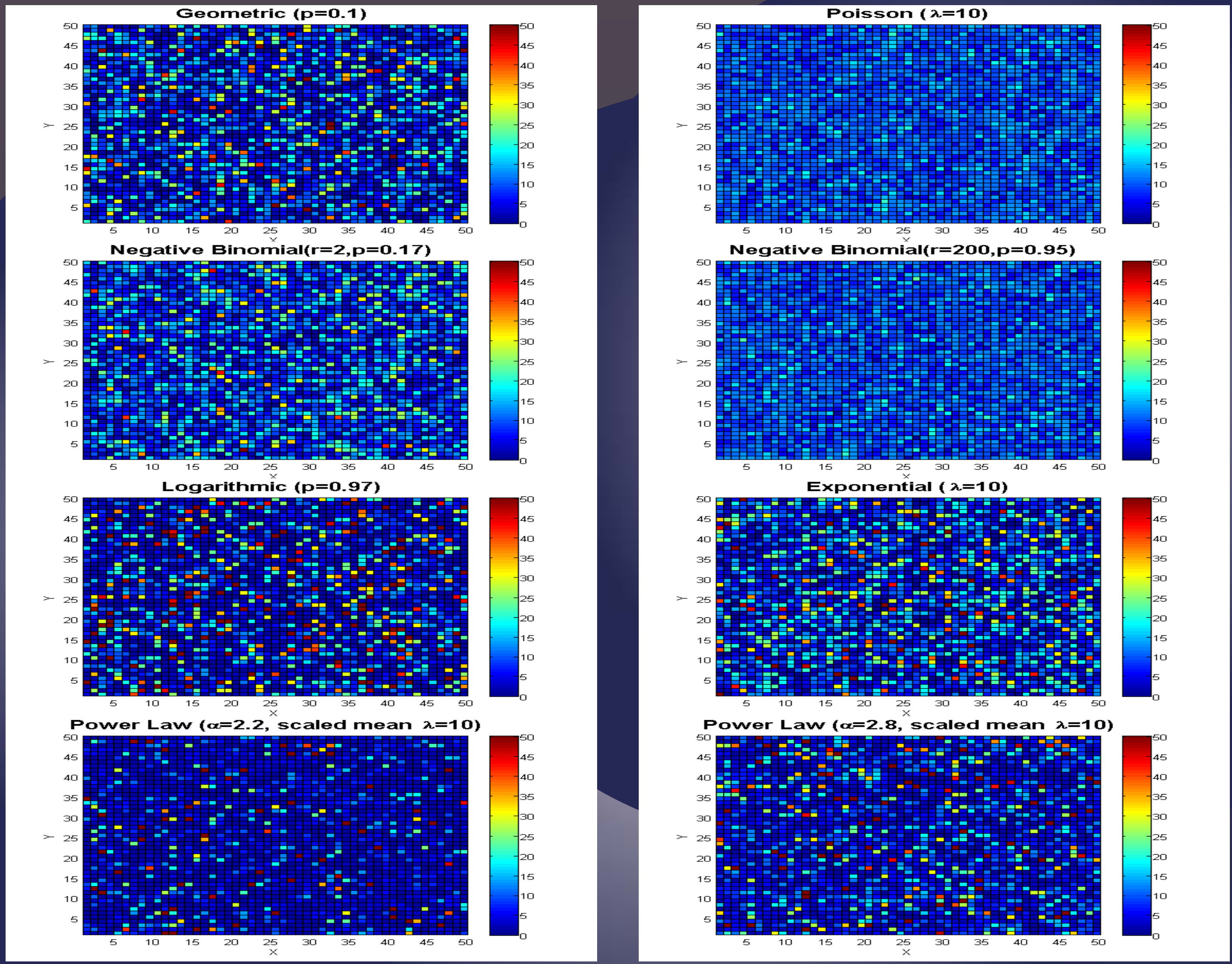


Figure 2a. Simulated seabird count maps with each of the candidate distributions (some distributions are shown with several possible parameter values, indicated in the panel title). To create each map, 2500 independent random draws were made from the indicated distribution and arranged on a 50x50 lattice. Note the apparent (false) hotspots and coldspots. All cells were drawn from a distribution with the same population mean value (λ=10) so all observed variation is purely due to statistical noise. Color scales are identical from panel to panel, and are scaled linearly.

N surveys = 3

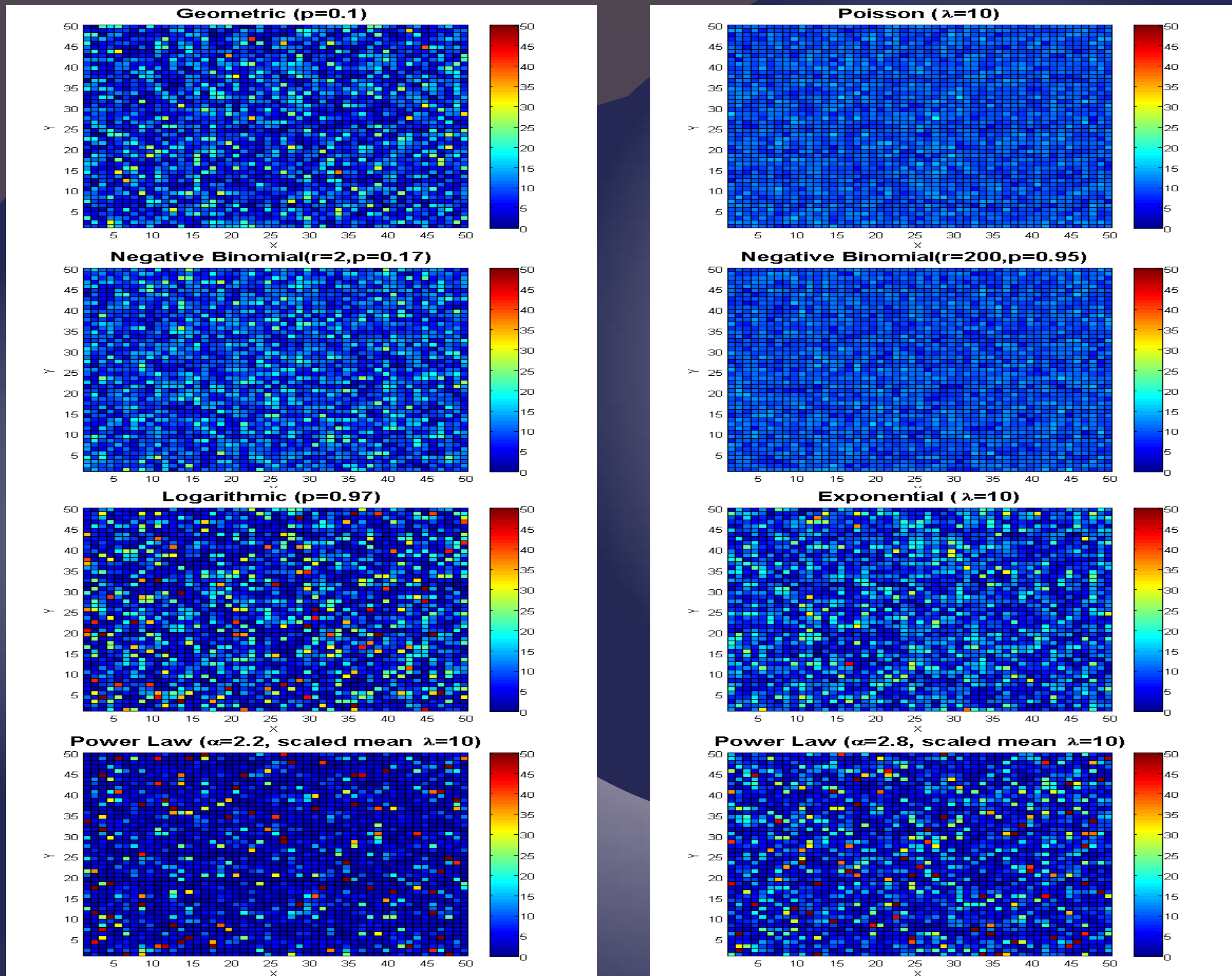


Figure 2b. Same as figure 2a, but with each point representing the average of 3 simulated surveys. Both surveys were simulated at random (i.e. first survey does not match figure 2a)

N surveys = 10

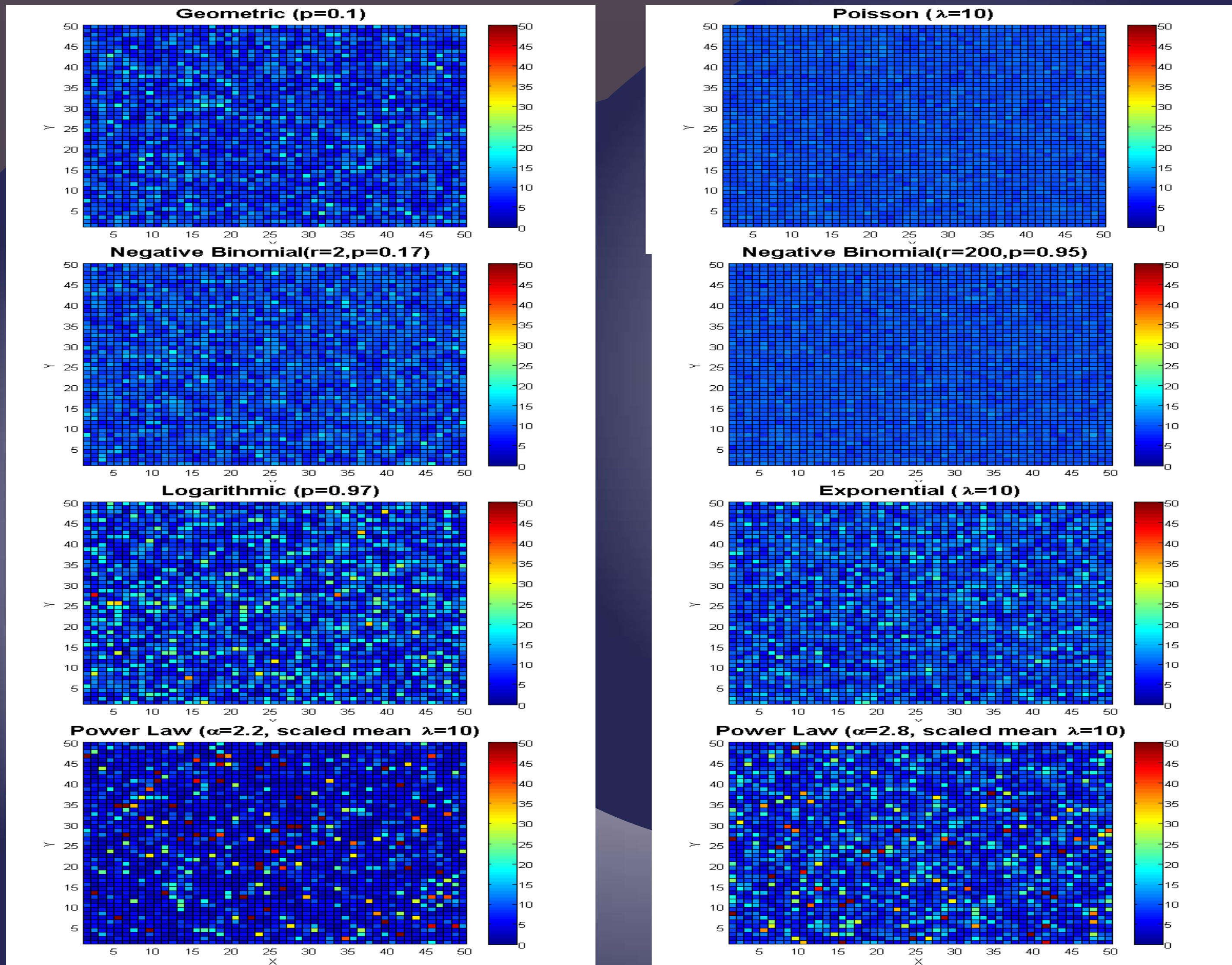


Figure 2c. Same as figure 2a, but with each point representing the average of 10 simulated surveys. Both surveys were simulated at random (i.e. first surveys do not match figures 2a or 2b)

N surveys = 100

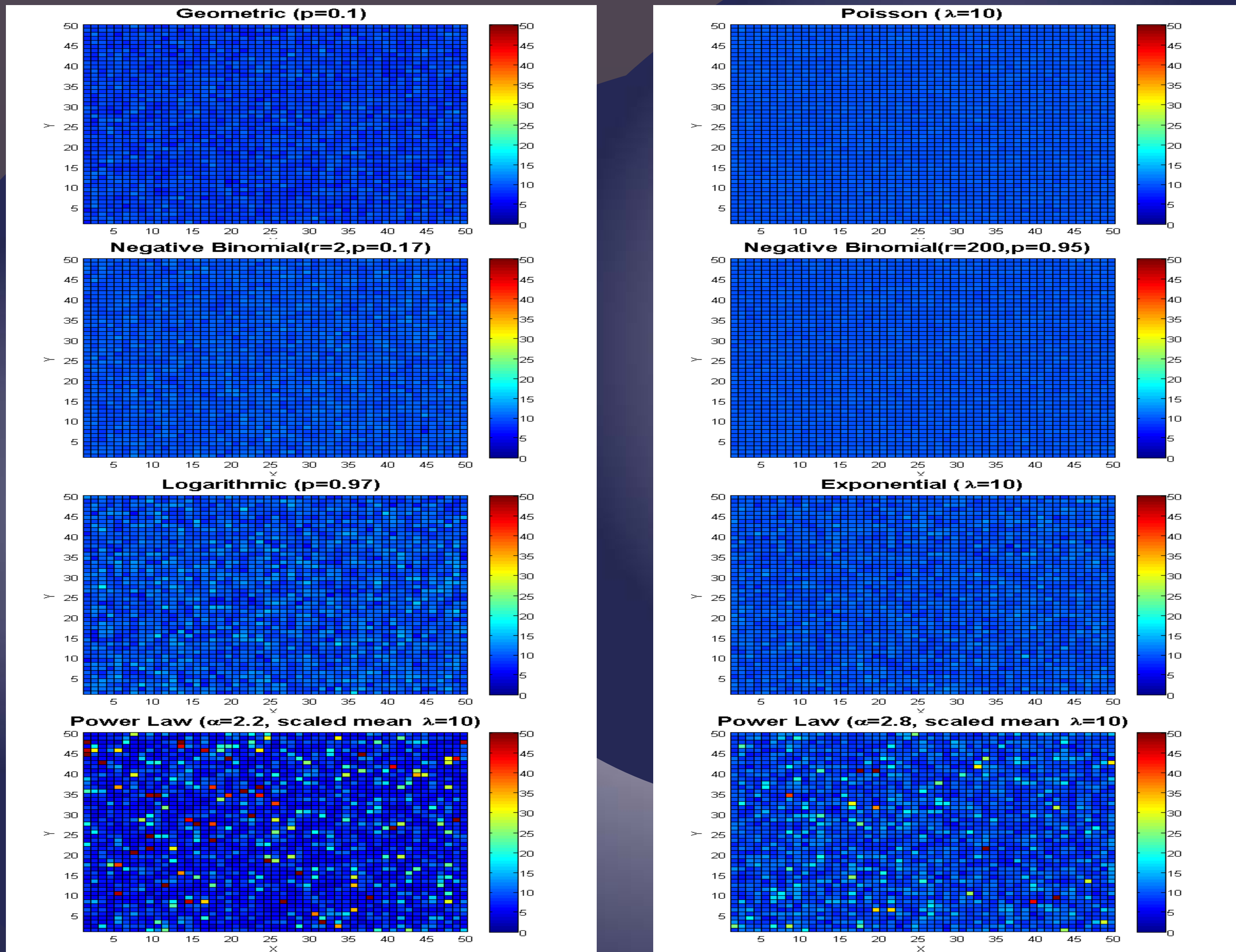
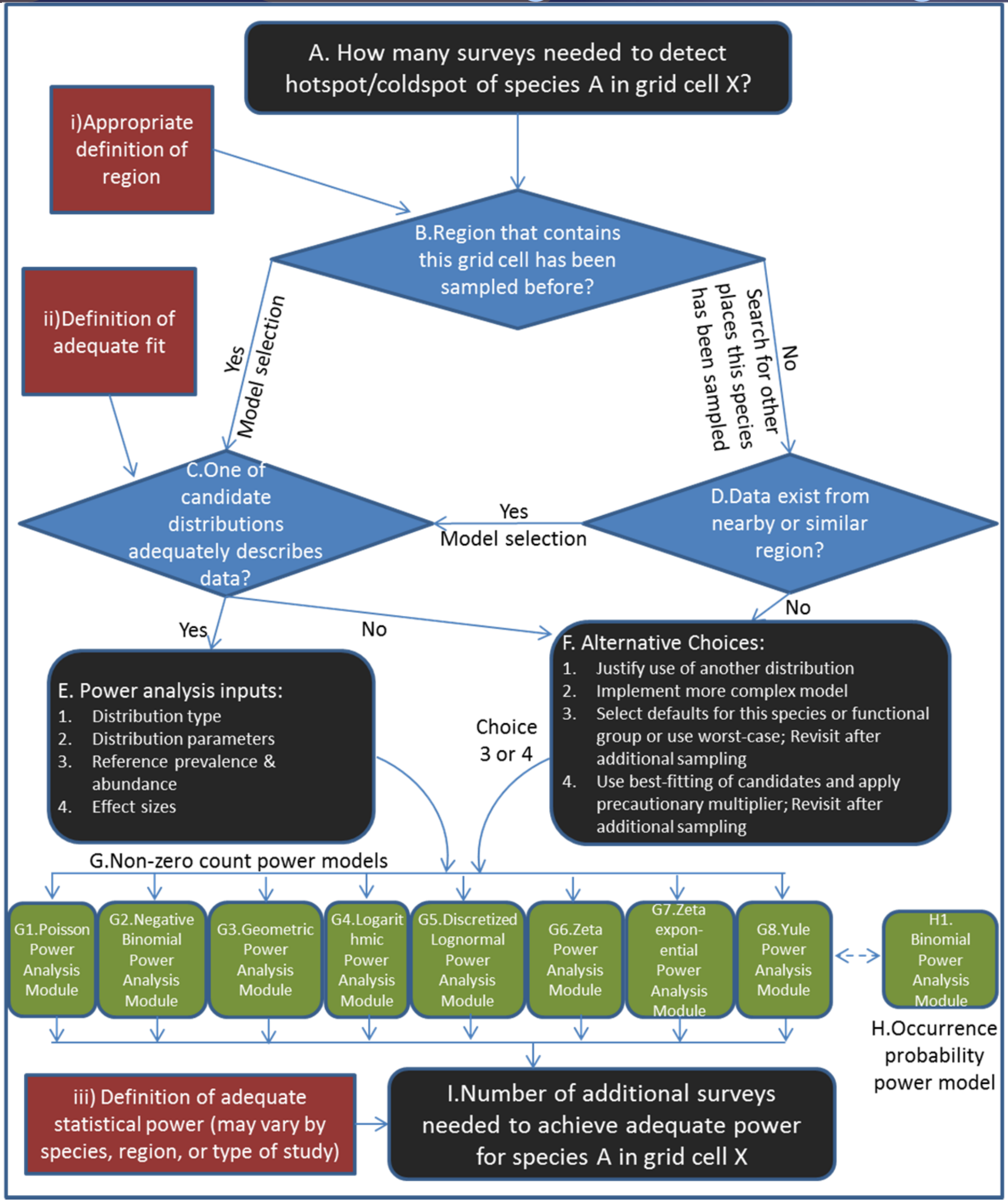
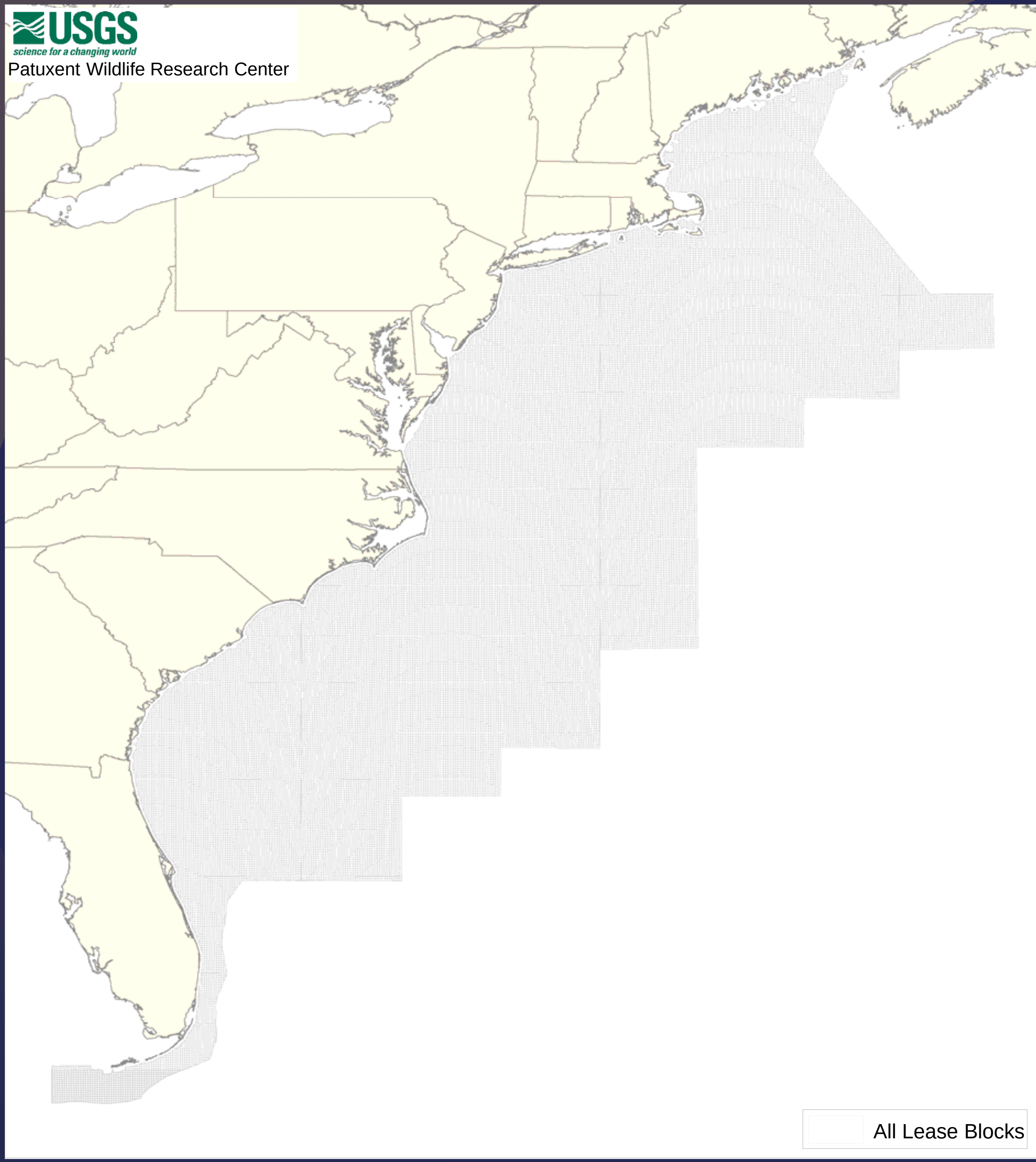


Figure 2d. Same as figure 2a, but with each point representing the average of 100 simulated surveys.
Both surveys were simulated at random (i.e. first surveys do not match figures 2a,b,c)

How many surveys?





U.S. Bureau of Ocean and Energy Management (BOEM)

- 5km x 5km lease blocks
- Along the Outer Continental Shelf of the Atlantic Ocean

The Atlantic Seabird Compendium

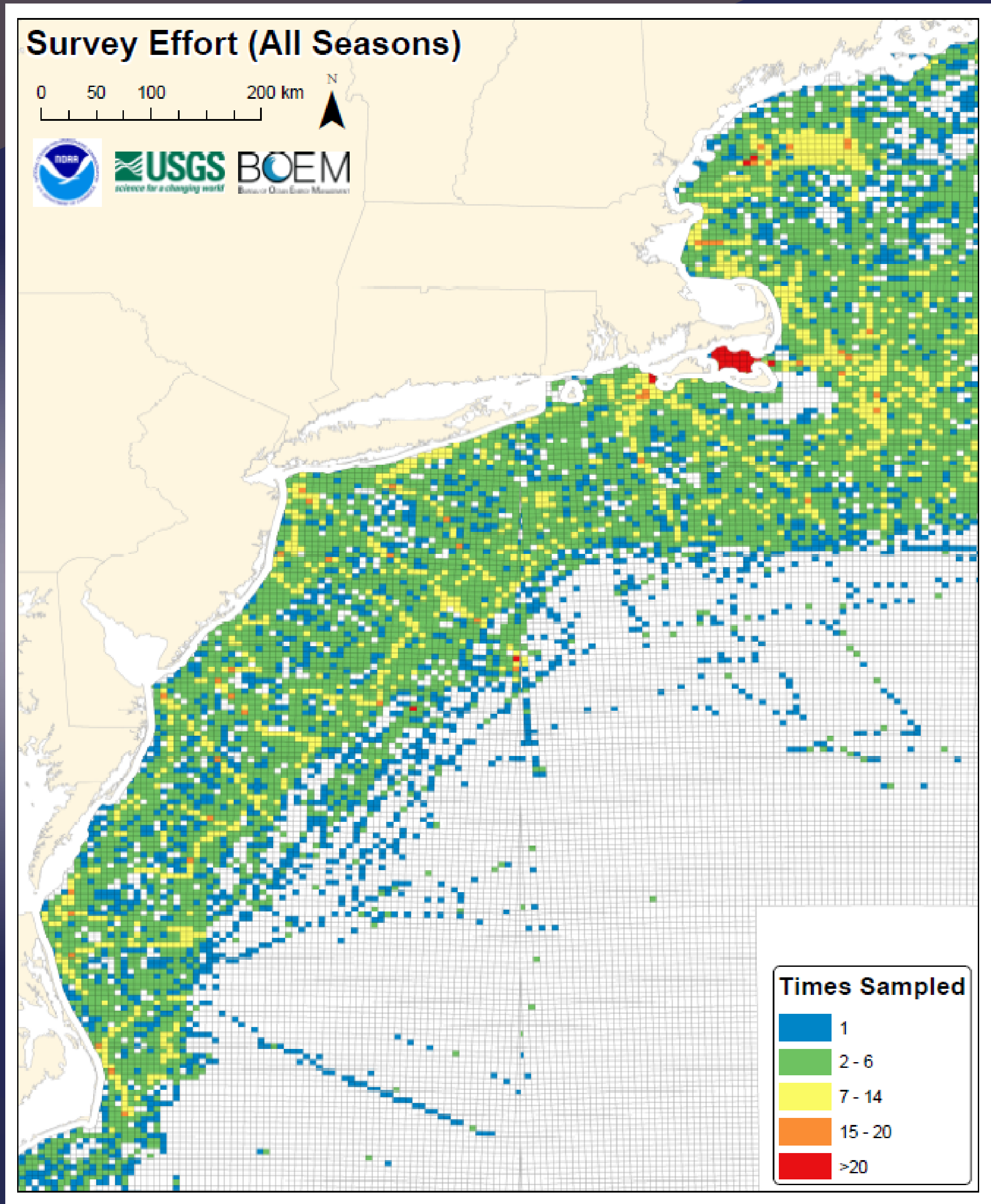
- >250,000 seabird observations from U.S. Atlantic waters
- Collected from 1978 through 2011
- Data collected using a mix of methods including non-scientific approaches

The Atlantic Seabird Compendium

- >250,000 seabird observations from U.S. Atlantic waters
- Collected from 1978 through 2011
- Data collected using a mix of methods including non-scientific approaches

We used:

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


Survey Effort (All Seasons)
0
50
100
200 km
N
USGS
science for a changing world
BOEM
Times Sampled
1
2 - 6
7 - 14
15 - 20
>20

Two part approach

1) Determine the best statistical distribution to model the count data for each species in each season

2) Conduct power analysis and significance testing on the basis of this distribution

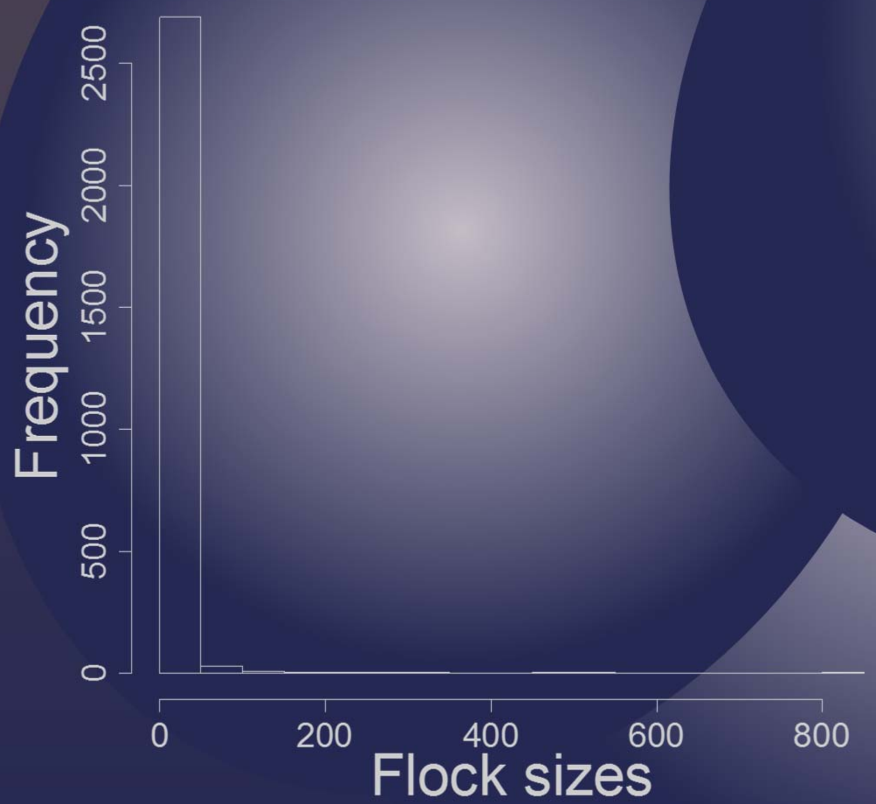
Two part approach

1) Determine the best statistical distribution to model the count data for each species in each season

2) Conduct power analysis and significance testing on the basis of this distribution

Model the data

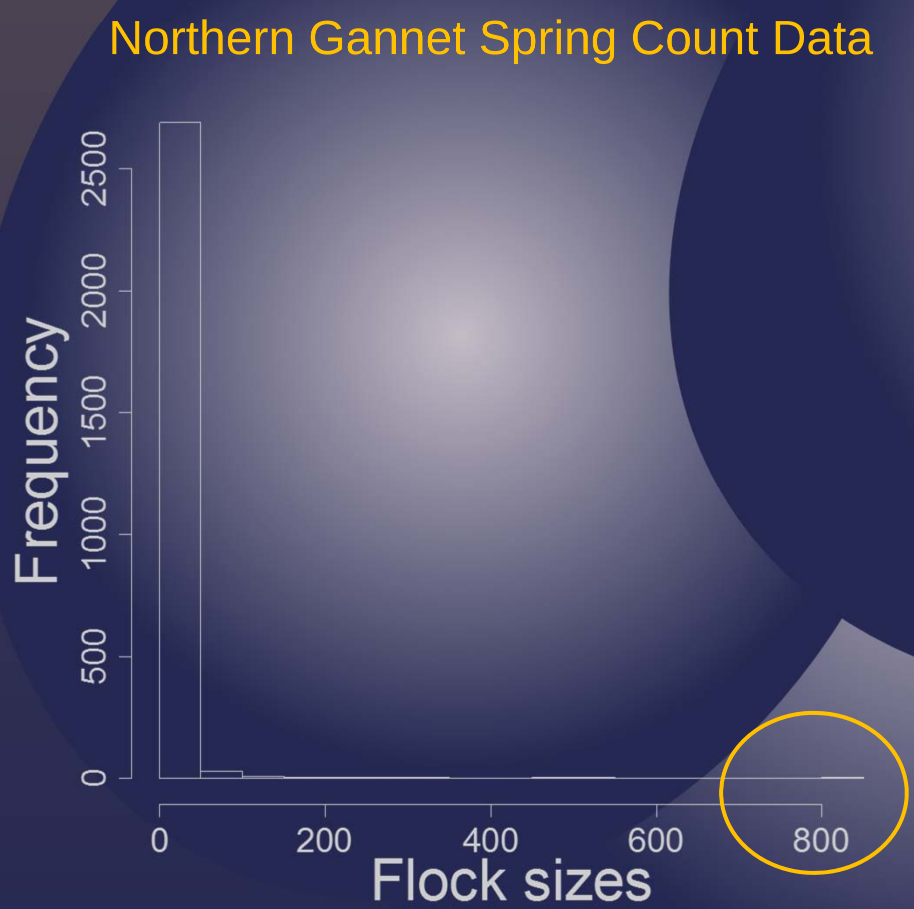
Northern Gannet Spring Count Data



Test eight statistical distributions:

Poisson
Negative binomial
Geometric
Logarithmic
Discretized lognormal
Zeta-exponential
Yule
Zeta (power law)

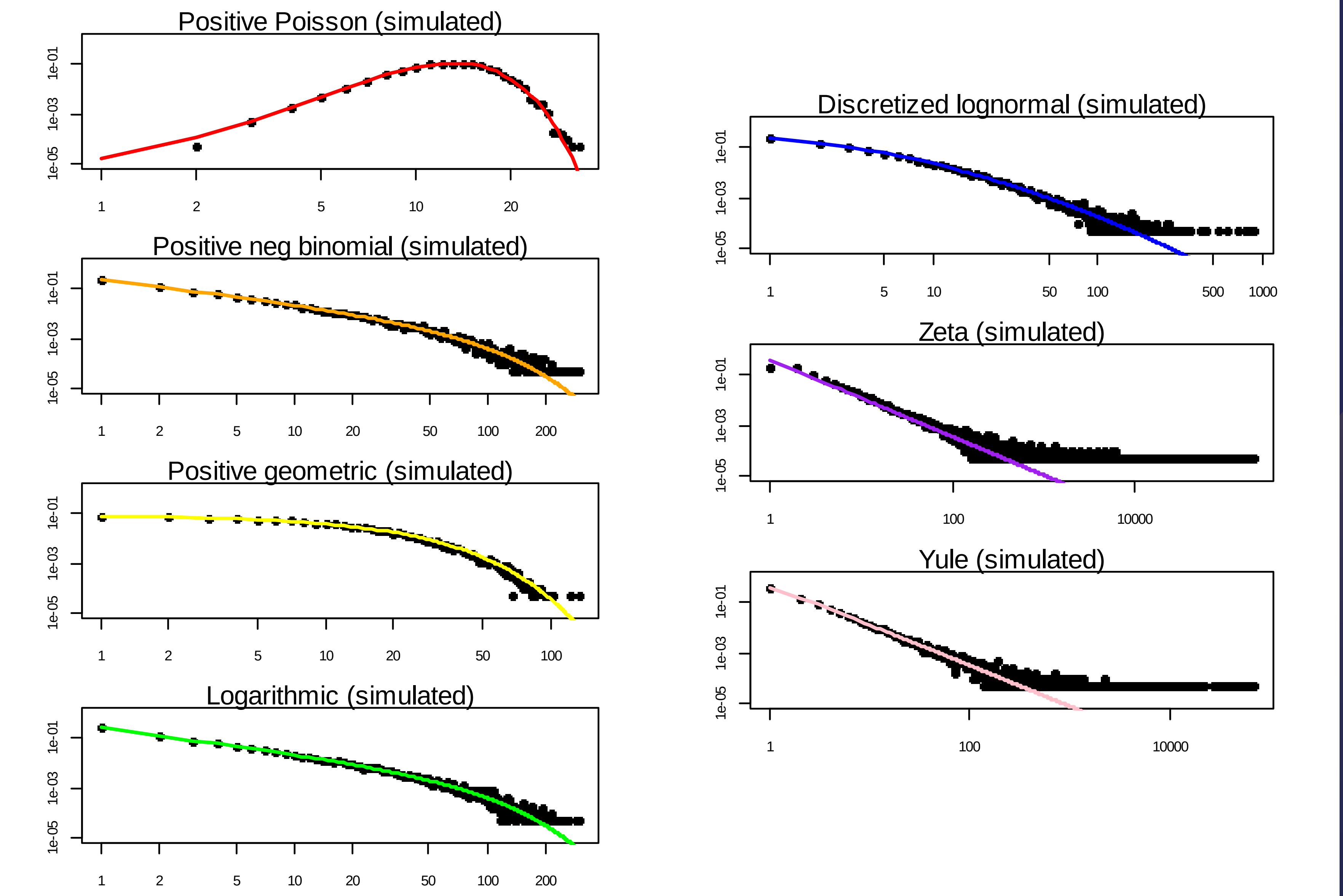
Model the data



Test eight statistical distributions:

Poisson
Negative binomial
Geometric
Logarithmic
Discretized lognormal
Zeta-exponential
Yule
Zeta (power law)

Examples of the distributions

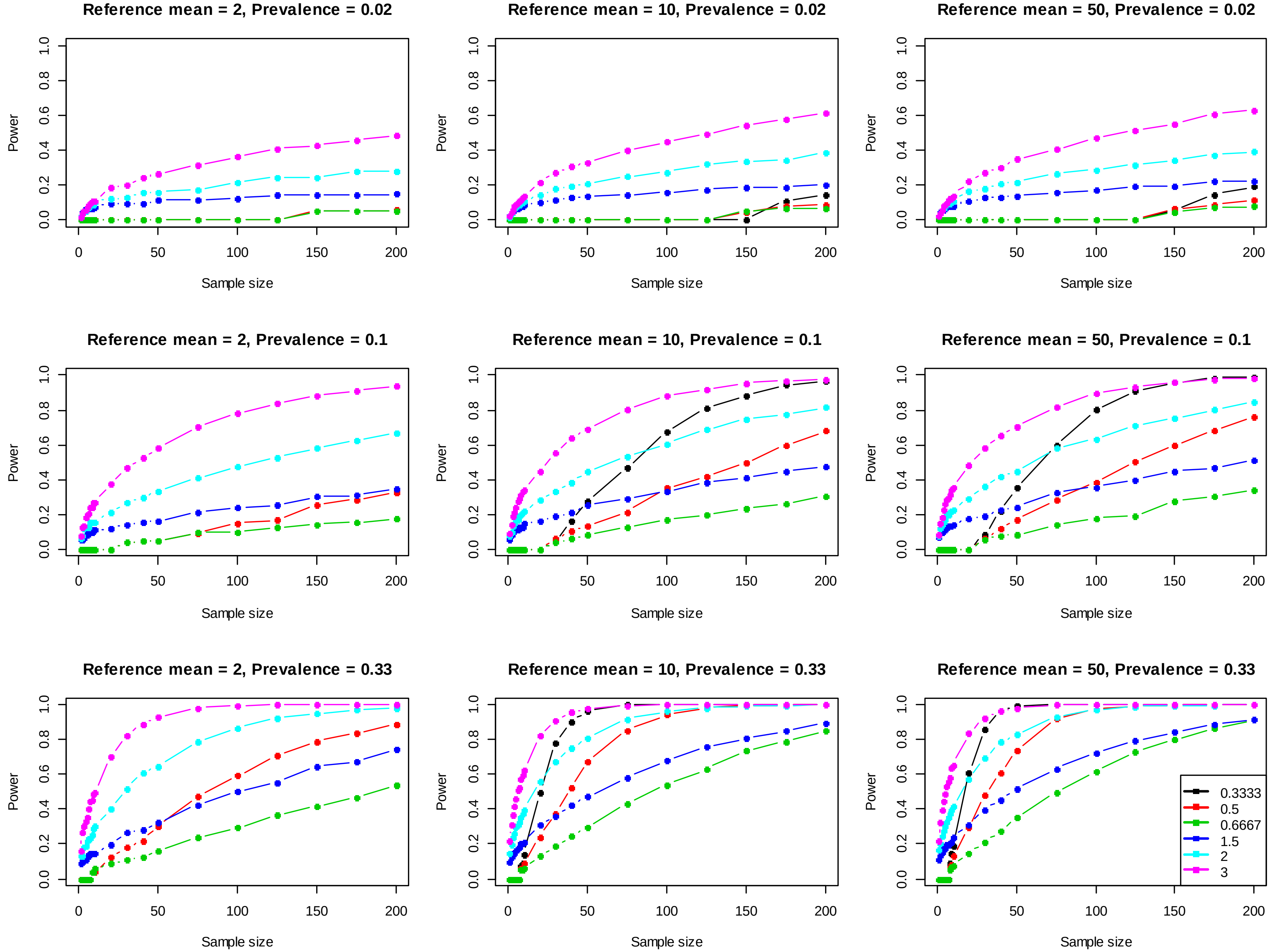


Model selection examples

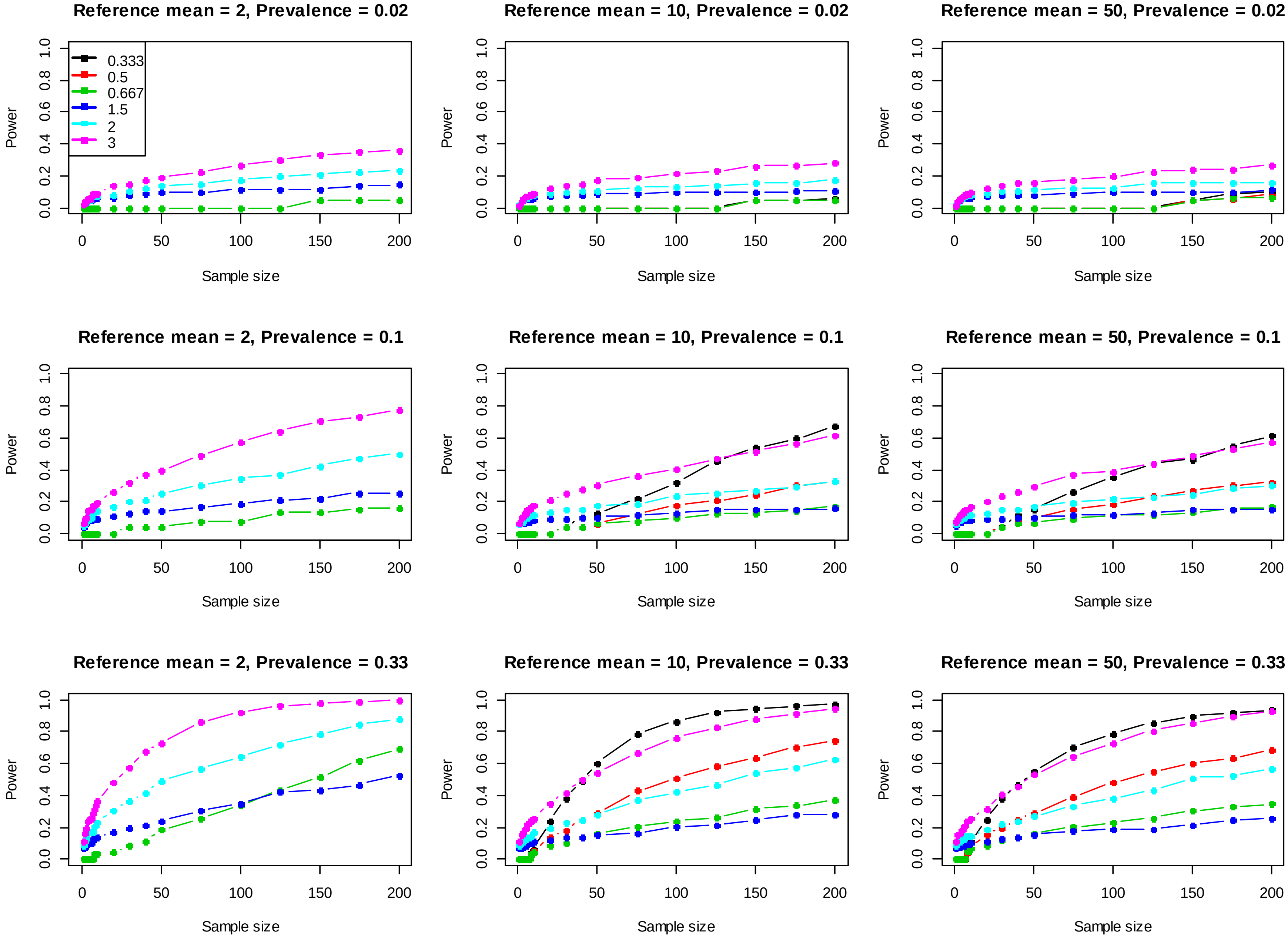
Model fitting and selection example: maximum likelihood estimates of best-fitting parameters of each candidate distribution to non-zero counts for three example species, with AICc and log-likelihood values. For each species, the models are ranked from lowest to highest AICc.

		Parameter estimates		AICc Rank	AICc	Log-Likelihood
Herring Gull (Spring)						
	Discretized lognormal	μ=0.138	σ=1.857	1	20473.03	-10234.51
	Zeta exponential	a=0.422	λ=0.006	2	20644.87	-10320.43
	Yule	a=0.711		3	20699.00	-10348.50
	Zeta	a=0.599		4	20884.84	-10441.42
	Logarithmic	p=0.976		5	21214.38	-10606.19
	Negative binomial	μ=0.206	k=0.005	6	21231.33	-10613.67
	Geometric	p=0.091		7	25628.78	-12813.39
	Poisson	λ=10.961		8	157322.40	-78660.20
Northern Gannet (Spring)						
	Discretized lognormal	μ=0.367	σ=1.870	1	13042.51	-6519.253
	Yule	a=0.835		2	13114.06	-6556.027
	Zeta exponential	a=0.526	λ=0.008	3	13116.56	-6556.278
	Zeta	a=0.684		4	13230.19	-6614.093
	Logarithmic	p=0.962		5	13605.86	-6801.929
	Negative binomial	μ=0.281	k=0.012	6	13632.13	-6814.064
	Geometric	p=0.130		7	16334.22	-8166.111
	Poisson	λ=7.677		8	70701.25	-35349.62
Wilson's Storm-Petrel (Spring)						
	Discretized lognormal	μ=0.009	σ=1.683	1	7004.017	-3500.005
	Yule	a=0.836		2	7067.586	-3532.792
	Zeta exponential	a=0.539	λ=0.006	3	7090.827	-3543.409
	Zeta	a=0.680		4	7144.087	-3571.042
	Logarithmic	p=0.965		5	7386.321	-3692.159
	Negative binomial	μ=0.259	k=0.010	6	7400.027	-3698.010
	Geometric	p=0.122		7	8974.693	-4486.345
	Poisson	λ=8.190		8	48571.69	-24284.84

Full Hurdle Model – Negative Binomial – r=2
Monte Carlo test – one tailed – alpha=0.05



Full Hurdle Model – Discretized Lognormal – σ=1.6
Monte Carlo test – one tailed – alpha=0.05



Results-Model Fitting

	Spring	Summer	Fall	Winter	Total
Number species with >500 observations	12	10	15	11	48

Results-Model Fitting

	Spring	Summer	Fall	Winter	Total
Number species with >500 observations	12	10	15	11	48
Discretized lognormal					
Yule					
Negative binomial Logarithmic Zeta decay					

Results-Model Fitting

	Spring	Summer	Fall	Winter	Total
Number species with >500 observations	12	10	15	11	48
Discretized lognormal	7 (4*)	4 (3*)	8 (3*)	8 (2*)	27 (12*)
Yule	1*	3*	1*	1	1 (5*)
Negative binomial Logarithmic Zeta decay			3*		0 (3*)

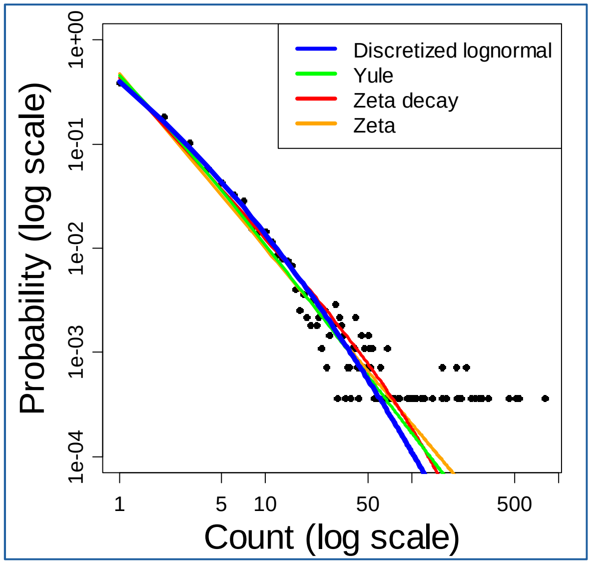
*Not significantly better for α = 0.05

Discretized Lognormal Distribution

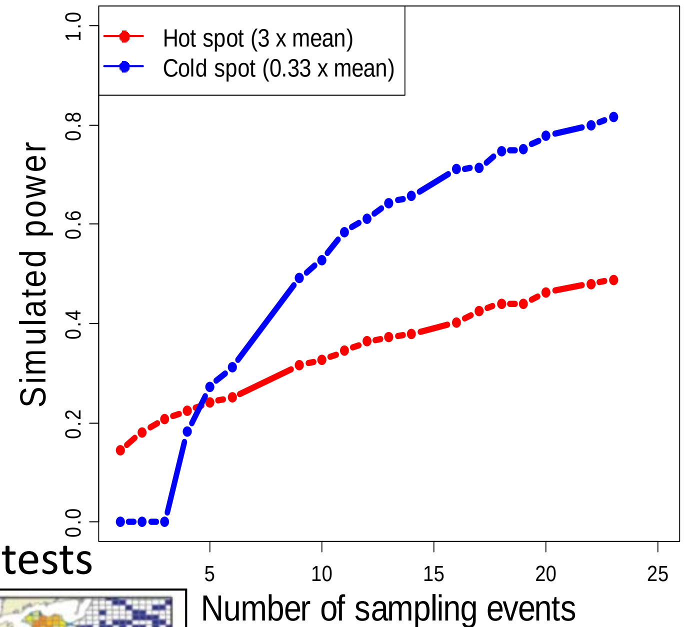
- Criteria:
  - Positive
  - Non-zero values
  - Highly skewed
  - Multiplicative effects

Model fit → Power Analysis

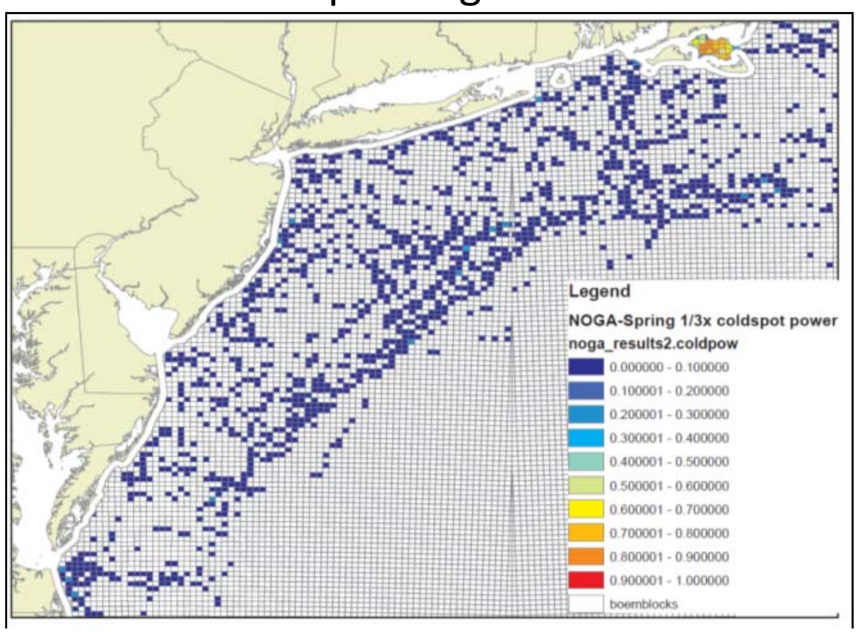
Model selection



Power curves



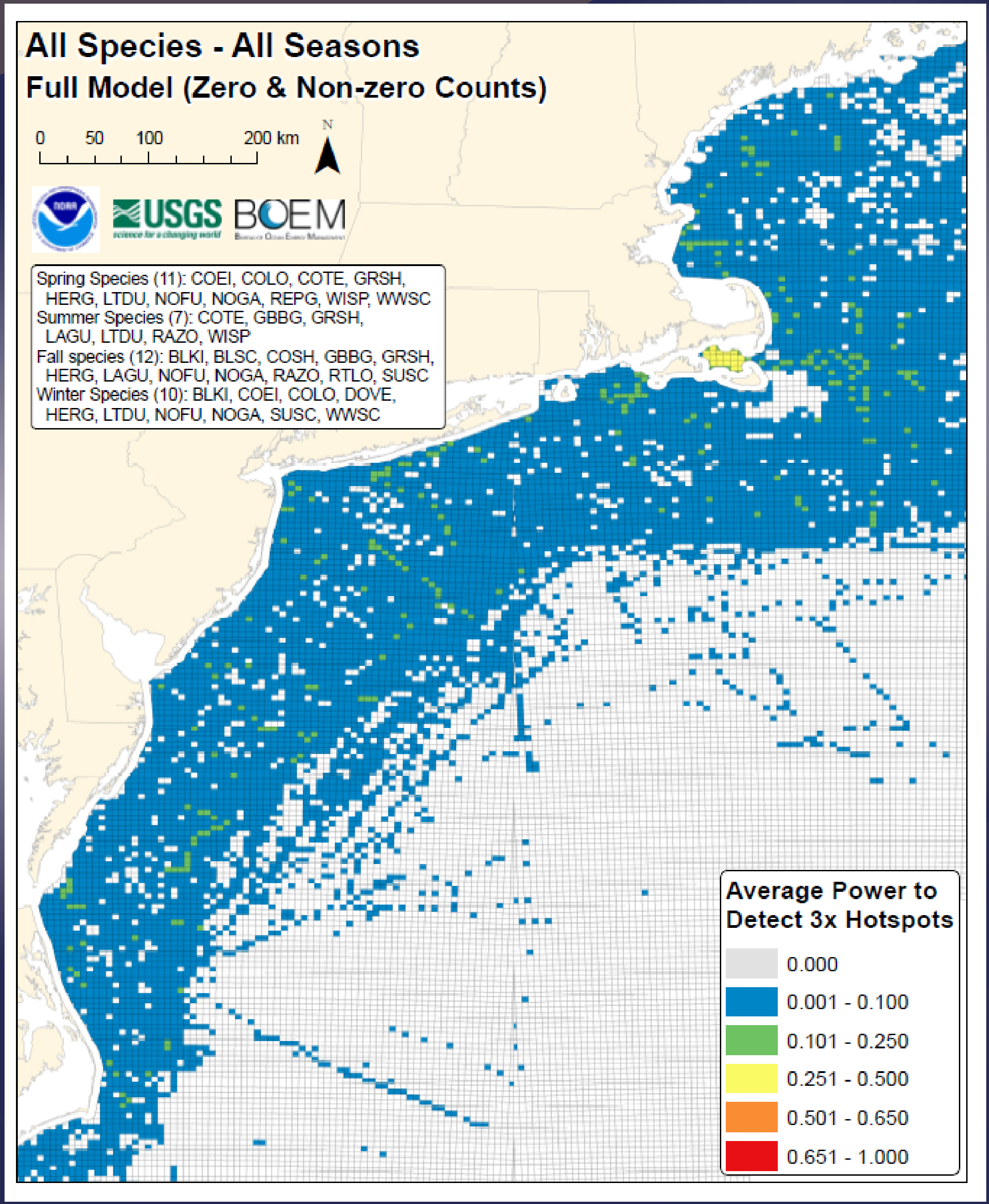
Power Maps & Significance tests



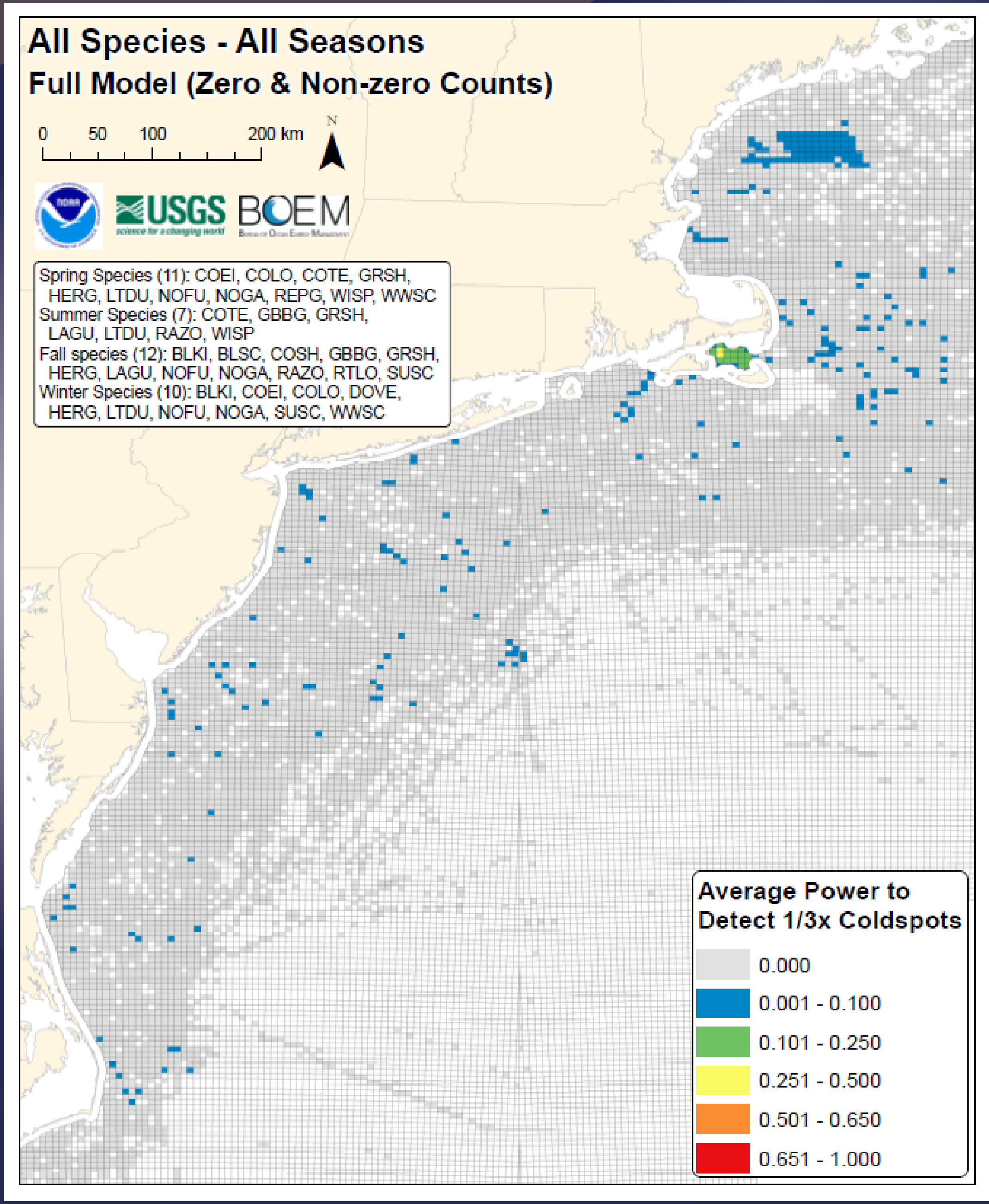
Products

- Interim report (Jan 2012)
- Mid-Term Technical Report (July 2012)
- Presented at 4th International Wildlife Management Conference in South Africa (July 2012)
- Tech memo: Kinlan, B.P., E.F. Zipkin, A.F. O'Connell, and C. Caldow. 2012. Statistical analyses to support guidelines for marine avian sampling: final report. U.S. Department of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs, Herndon, VA. OCS Study BOEM 2012-101. NOAA Technical Memorandum NOS NCCOS 158. xiv+77 pp.
- Journal article: Zipkin, E.F., J.B. Leirness, B.P. Kinlan, A.F. O'Connell, and E.D. Silverman. 2012. Fitting statistical distributions to sea duck count data: implications for survey design and abundance estimation. Statistical Methodology. doi:10.1016/j.stamet.2012.10.002
- Journal article: Kinlan, B.P., E.F. Zipkin, A.F. O'Connell, M. Wimer, D. Rypkema, A. Sussman, C. Caldow . 2013. Detection of "hotspots" and "coldspots" in marine avian survey data: power analysis and implications for survey design and interpretation. *In preparation for submission to* Journal of Applied Ecology.

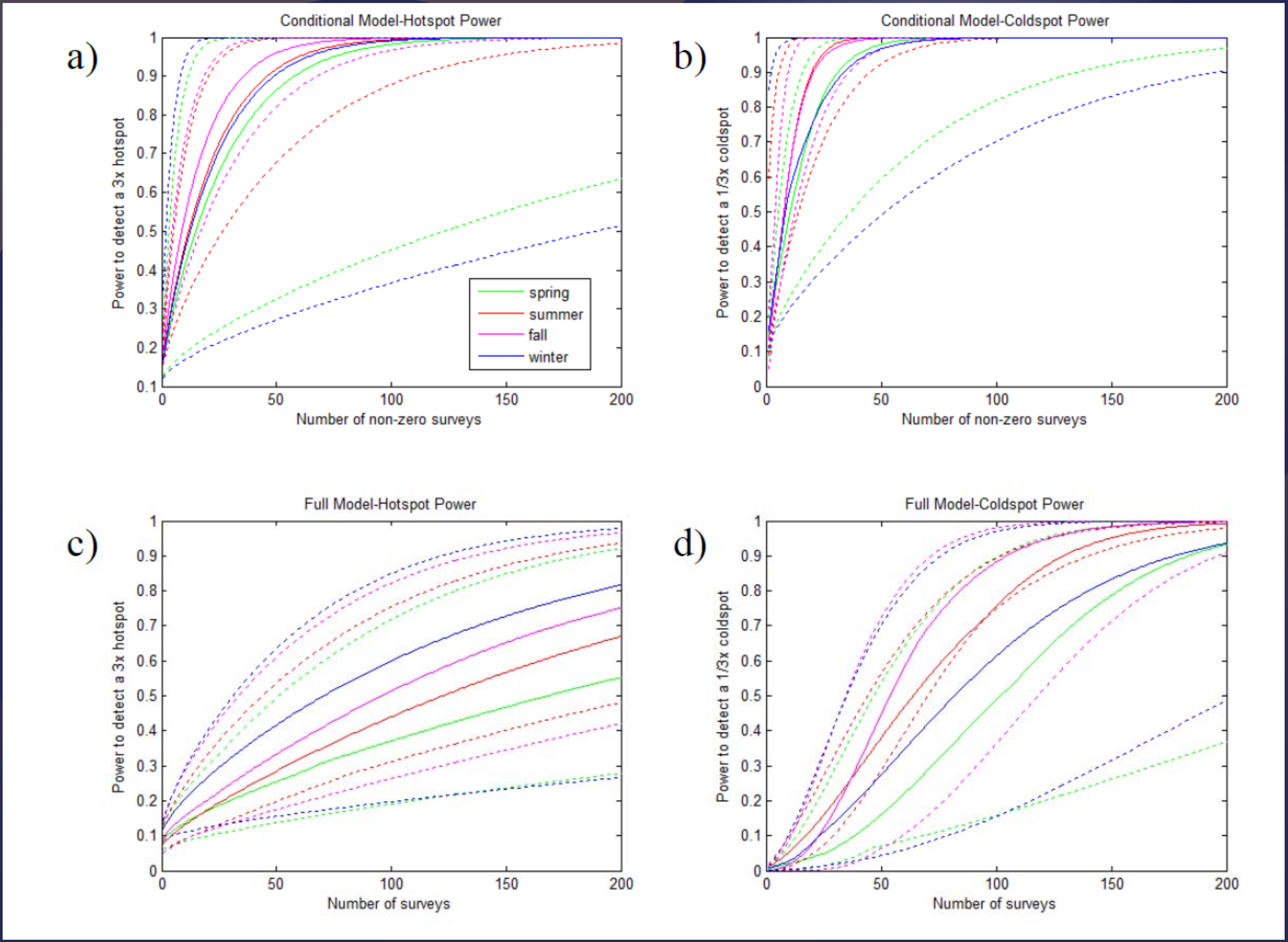
Average hotspot power



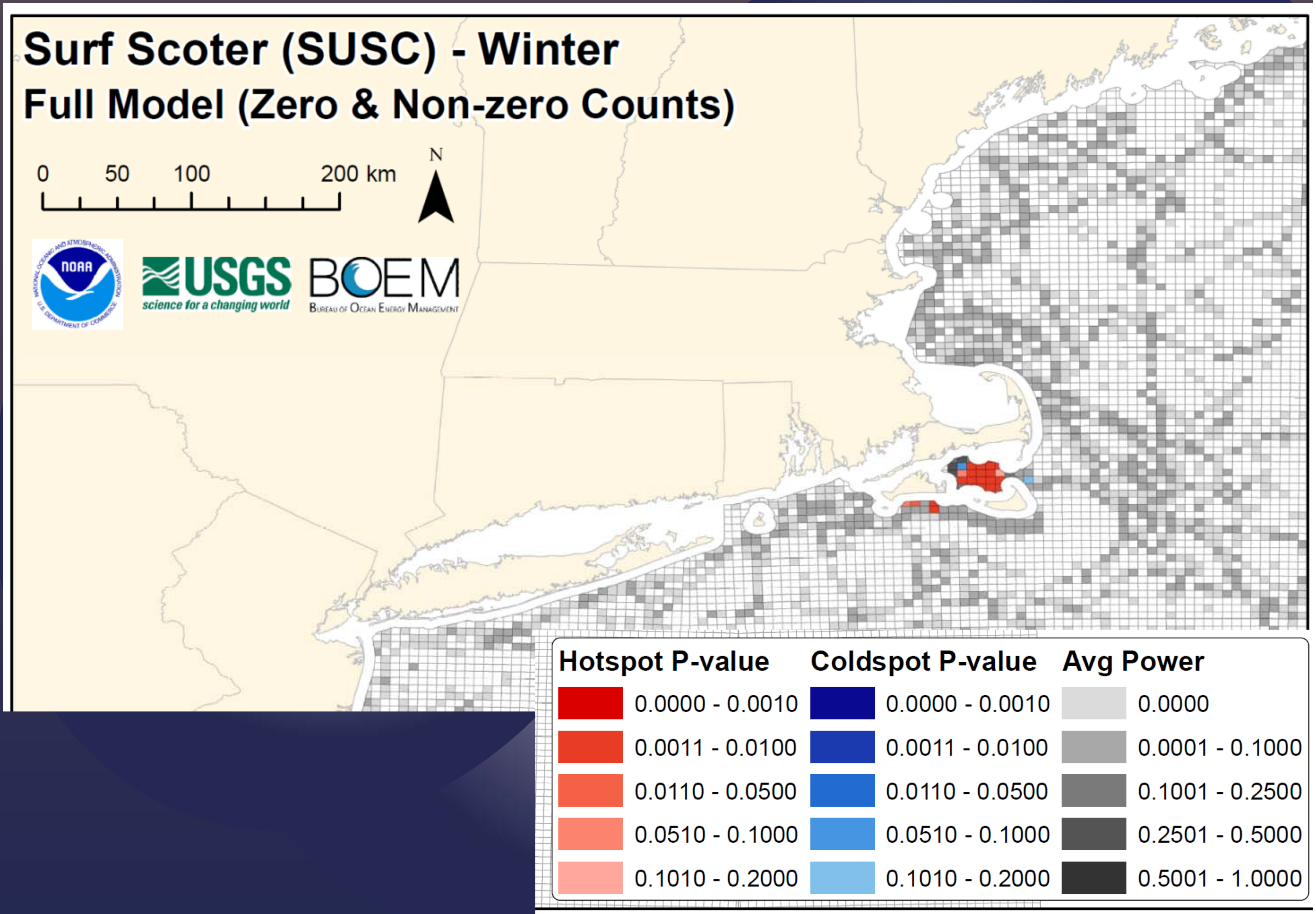
Average coldspot power



Multi-species summary of power curves



Significance tests



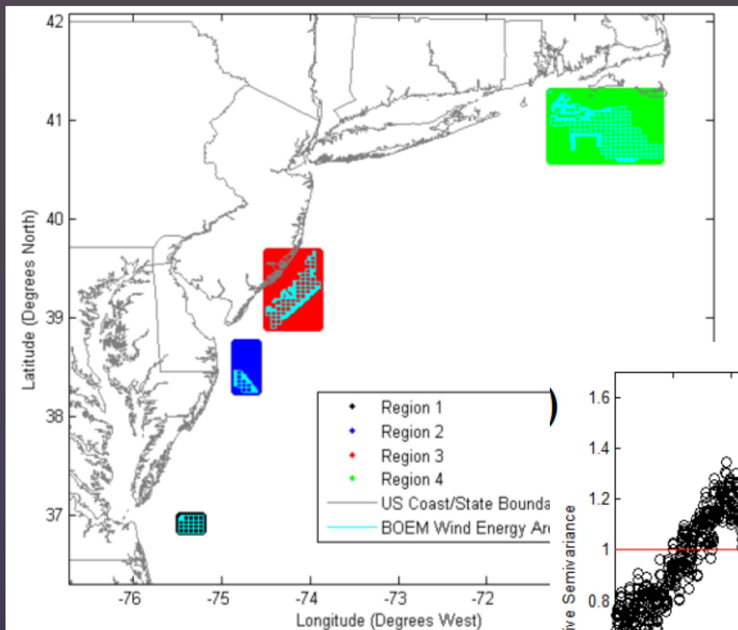
Broad summary of results

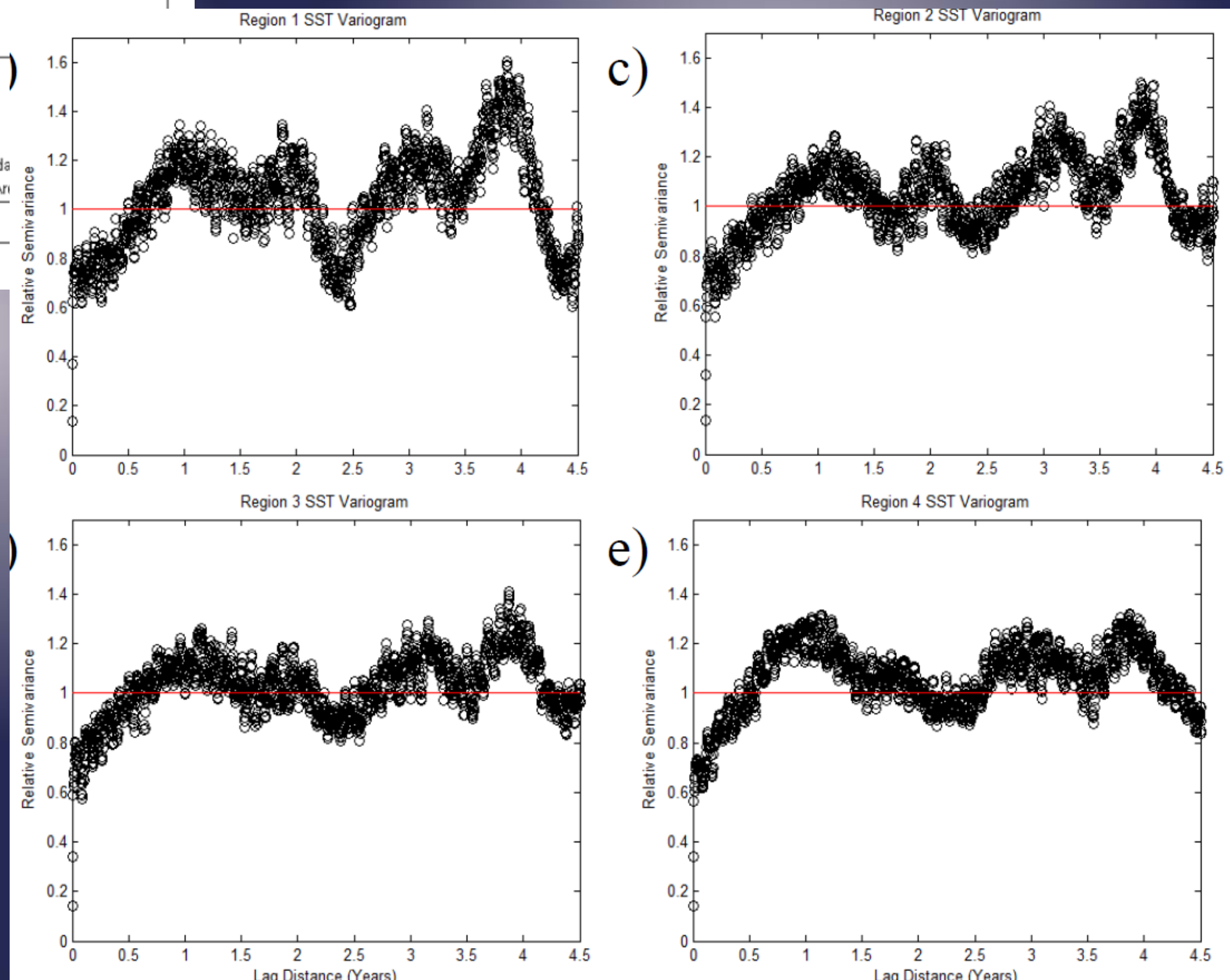
- Useful technique
- Need to do additional focal work on key species of interest
- Most areas of the Atlantic need additional sampling to have adequate power to detect hotspots/coldspots
- Maps could be used to select well-studied areas where less additional sampling required
- Rare species a challenge

Characterizing Temporal Variability

Sea Surface Temperature

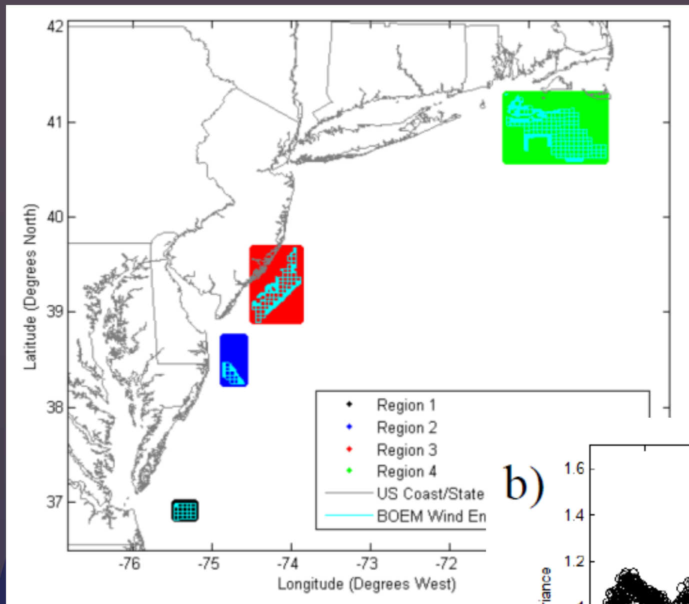
Variograms of de-seasoned SST in WEA areas

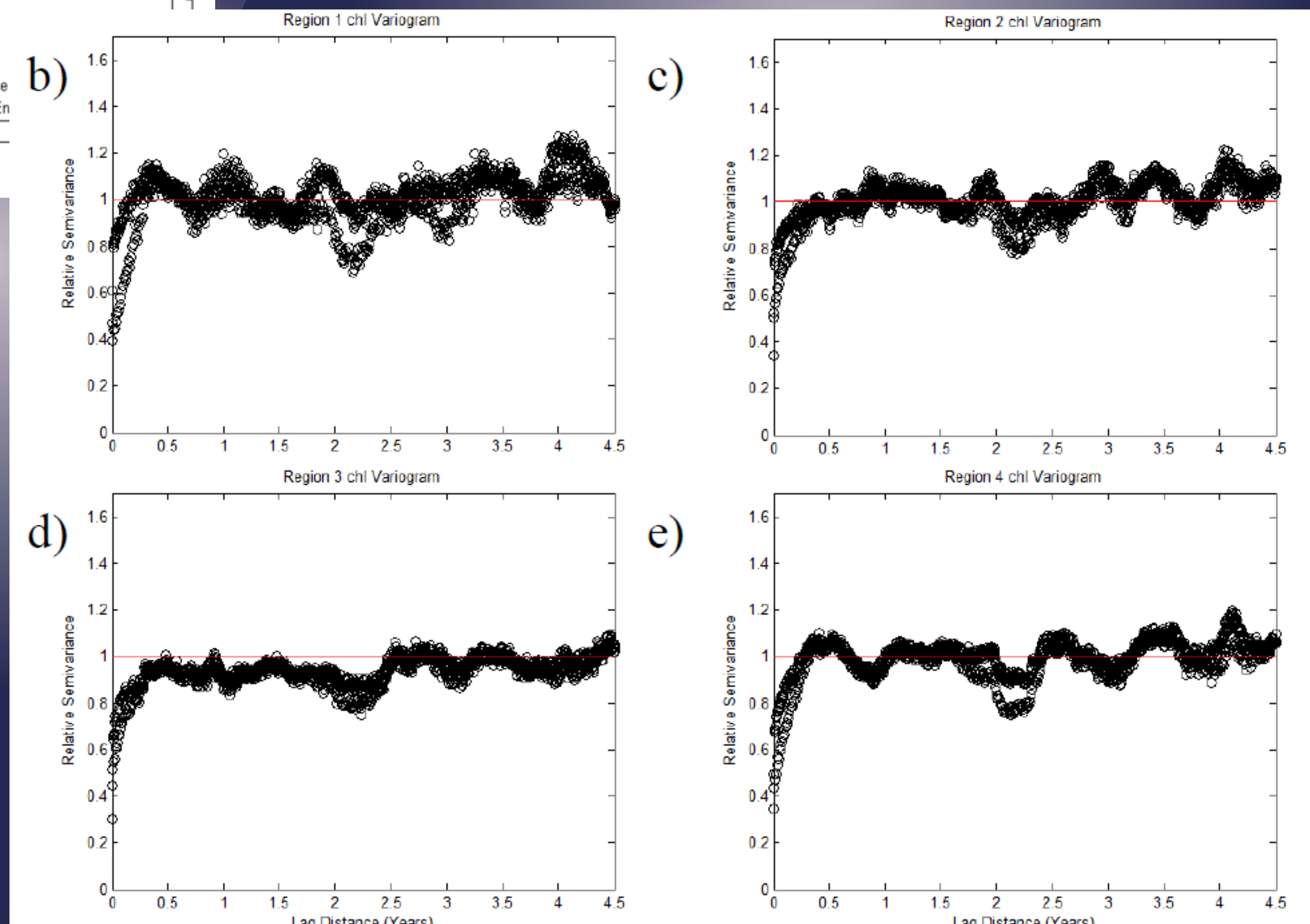




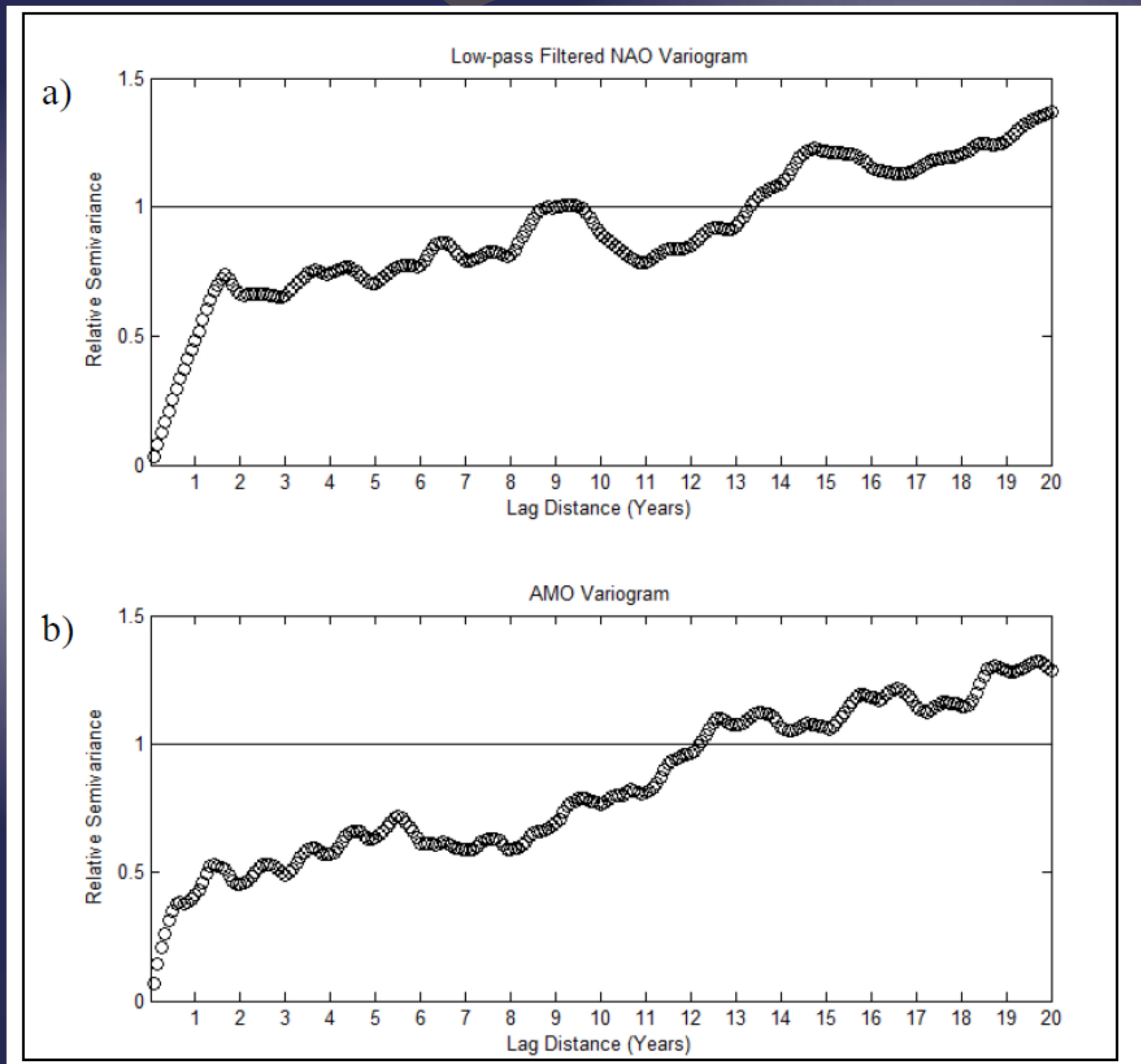
Surface Chlorophyll

Variograms of de-seasoned Log10(Chl) in WEA areas

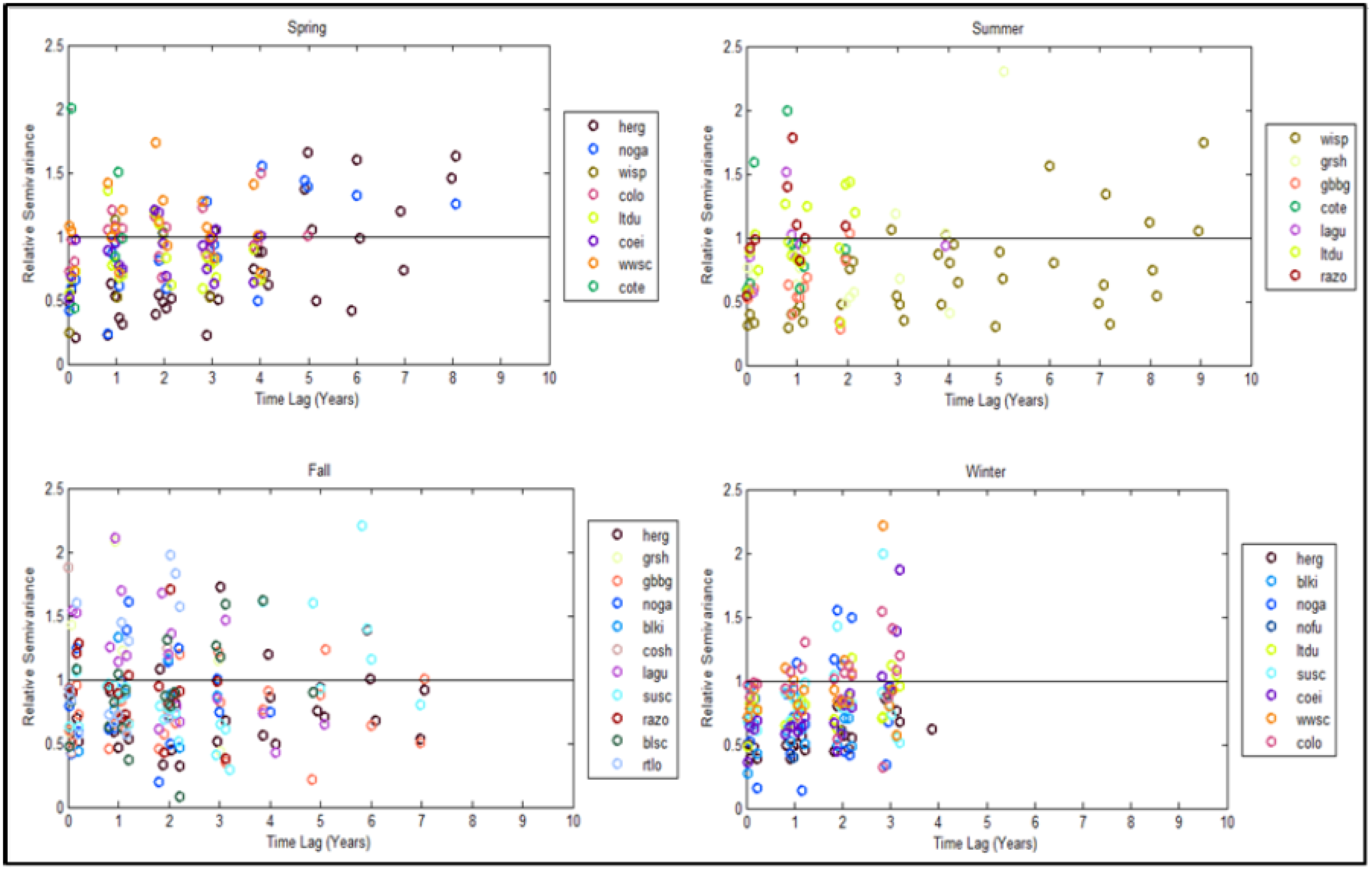




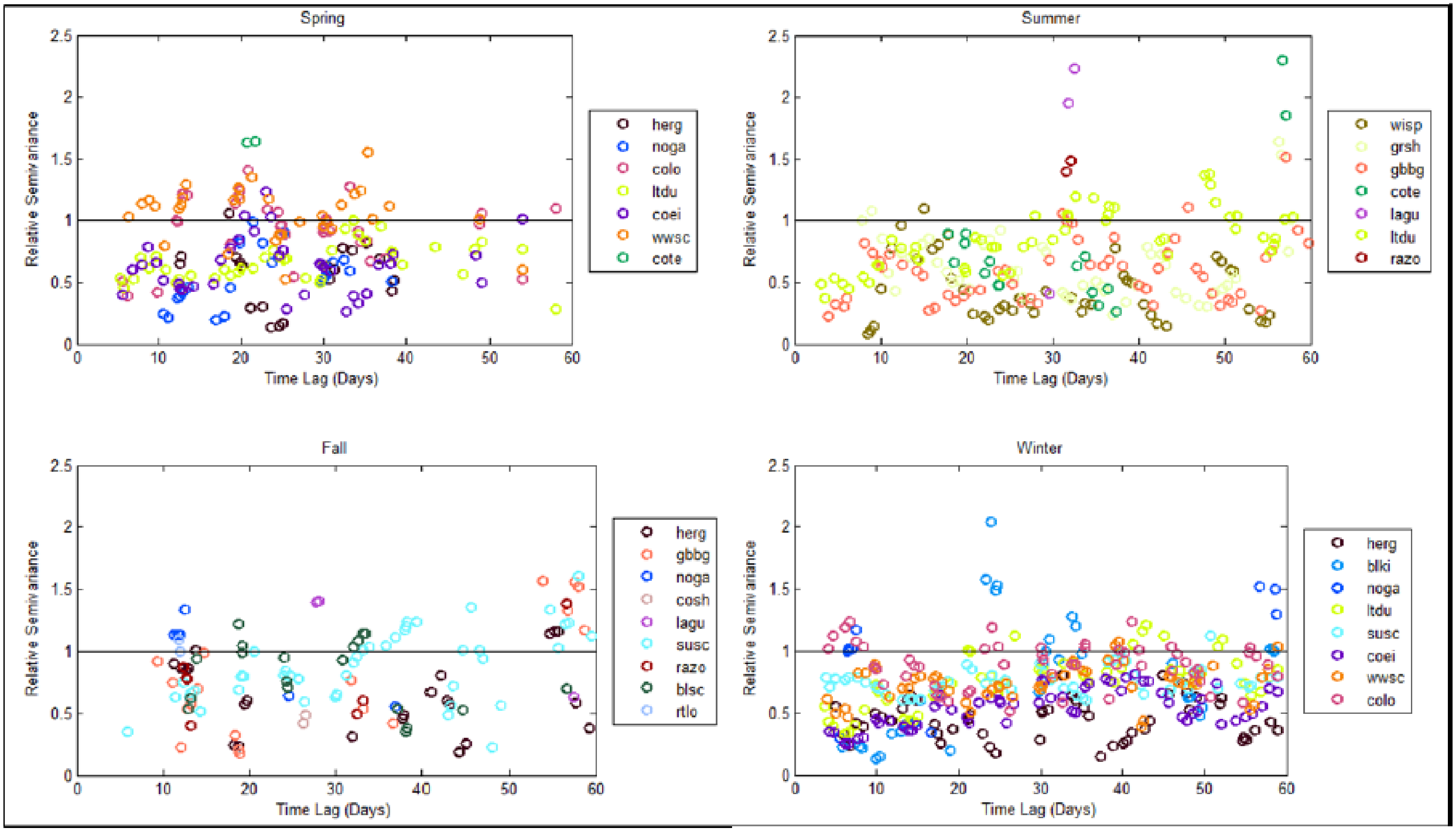
Longer term variability – Interdecadal climate indices



Temporal variability in marine bird count data within BOEM lease blocks – LONG TERM



Temporal variability in marine bird count data within BOEM lease blocks – SHORT TERM



Discussion

- Overview of final report
  - General walk-through
  - Look at and discuss results for species of interest
- Discuss issues
  - Spatial scale
  - Temporal scale/environmental variability
  - Spatial and temporal trends
  - Rare species/data poor situations
  - Comparison to other approaches
  - Detectability and other observer/platform issues
  - Next steps/practical applications

Acknowledgements

Sampling Design/Power Analysis Project

Diana Rypkema (NOAA Hollings Scholar)
Emily Silverman (USFWS)
Jeffery Leirness (USFWS)

Technical Reviewers – Jim Baldwin (USDA Forest Service), Jocelyn Brown-Saracino (DOE), David Bigger (BOEM), BOEM Renewable Energy/Avian Biology Team

Data :
Atlantic Seabird Survey Compendium

Funding:
BOEM, USGS

Brian.Kinlan@NOAA.gov 301-713-3028 x157