Understanding NARW Abundance and Distribution Through Data Fusion

Rob Schick[1,2], Bokgyeong Kang[2,3], Alan Gelfand[2], Erin Schliep[4],
Tina Yack[2], Chris Clark[5], Susan Parks[6],
Stormy Mayo[7], Christy Hudak[7], Ryan Schosberg[7], Brigid McKenna[7]

1 Southall Environmental Associates, Inc. 2 Duke University 3 Dongguk University 4 NC State University 5 Cornell University 6 7 Center for Coastal Studies Provincetown

Funding

- US Office of Naval Research
  - N000142312562
  - N000142412501
- NOAA Fisheries
  - NA20NMF0080246
- SERDP
  - RC20-1097

Outline

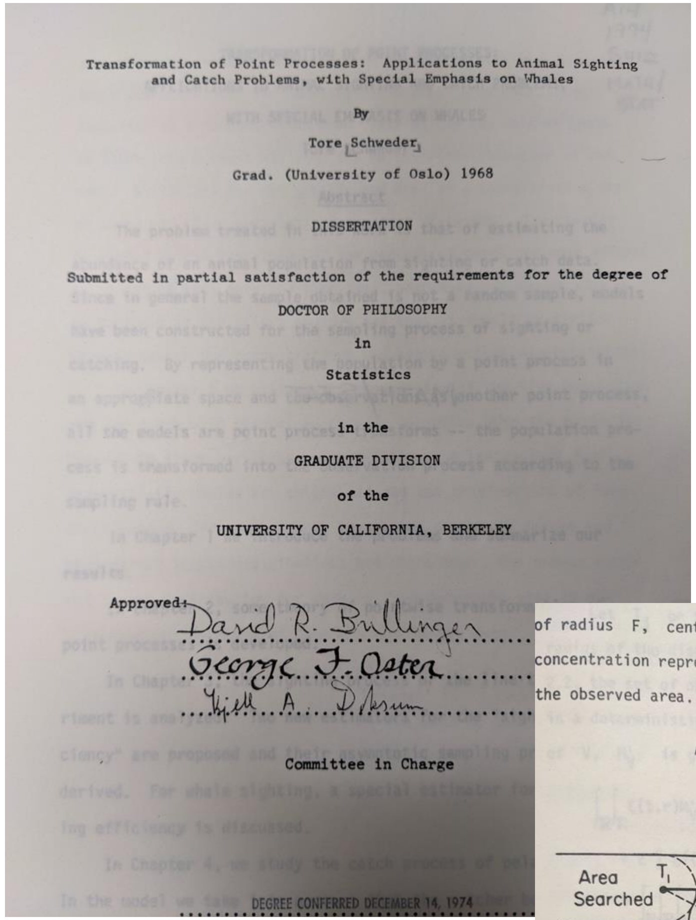
- Introduction to point processes and data fusion
- (Simulation Study)
- Application: NARW in Cape Cod Bay
  - Modeling framework
  - Key assumptions
- Results
- Extensions

NARW

- For right whales, we are fusing two data sources (aerial sightings and PAM) that relate to a spatial point pattern

Point Processes for MM Data Aren't New

- Schweder 1974
- Hedley and Buckland 2004
- Waagepetersen and Schweder 2006
- Johnson et al. 2013
- Yuan et al. 2017



Transformation of Point Processes: Applications to Animal Sighting and Catch Problems, with Special Emphasis on Whales

By

Tore Schweder

Grad. (University of Oslo) 1968

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

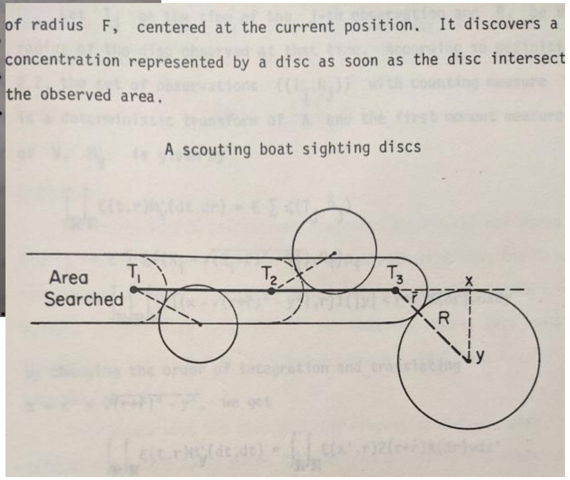
Approved: David R. Brillinger

George F. Oster

Kjell A. Doksum

Committee in Charge

DEGREE CONFERRED DECEMBER 14, 1974



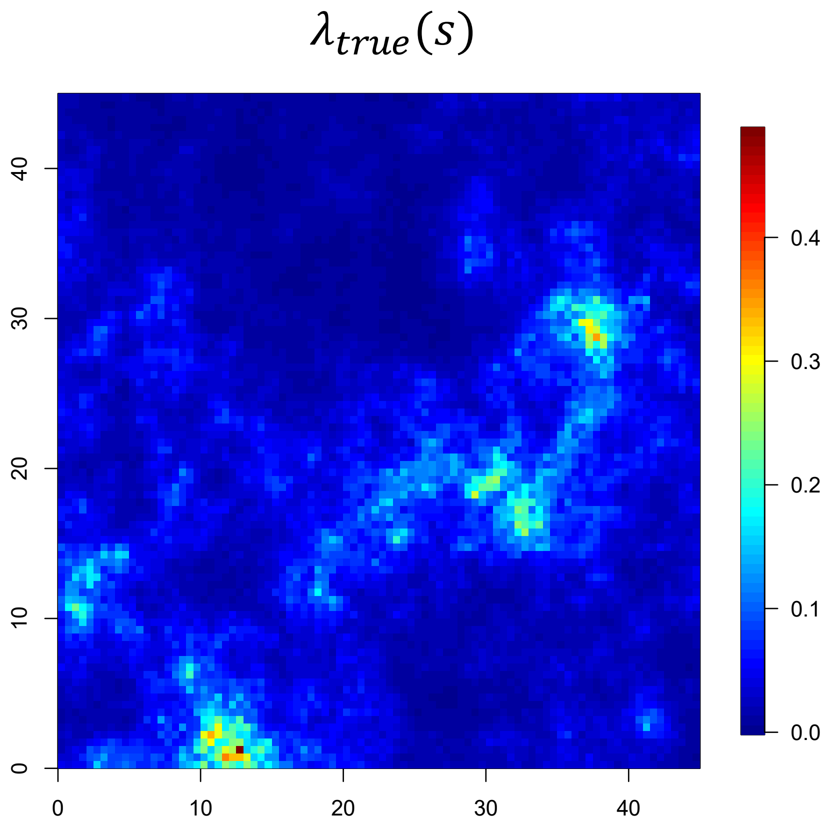
of radius F, centered at the current position. It discovers a concentration represented by a disc as soon as the disc intersect the observed area.

A scouting boat sighting discs

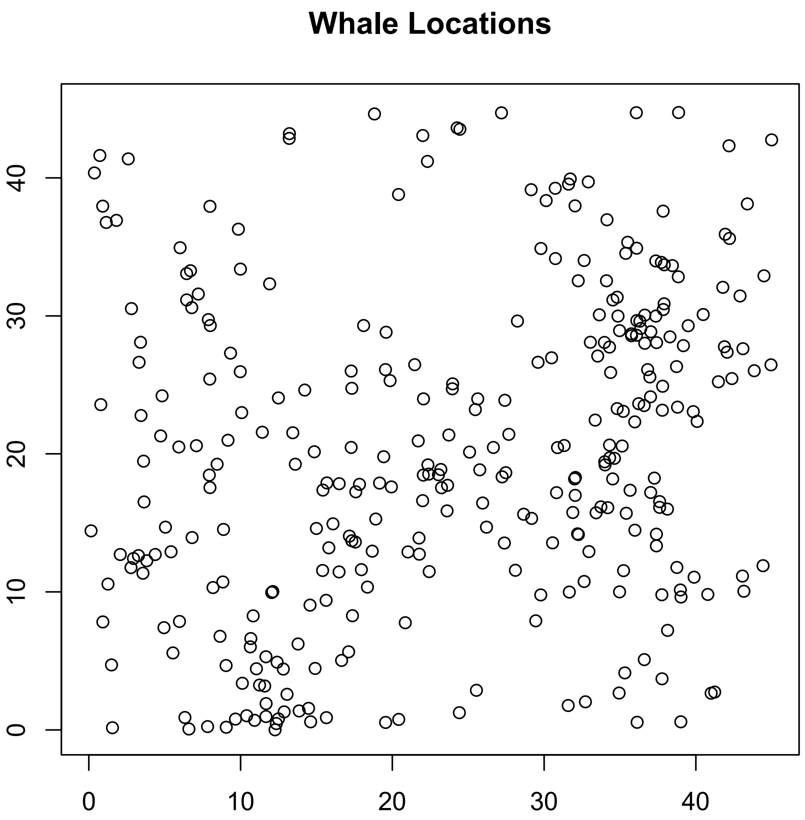
NARW

- For right whales, we are fusing two data sources (aerial sightings and PAM) that relate to a spatial point pattern
- Have to first consider *how* the data relate to this point pattern
- We assume a single unobservable
  - True point pattern (S), i.e., the locations of whales &
  - True intensity surface ($\lambda(\boldsymbol{s})$), for which the point pattern S arises
- Each data source provides a partial realization of the full point pattern
  - Each is a thinned version of S with its own thinning mechanism

True Intensity Surface & Point Pattern

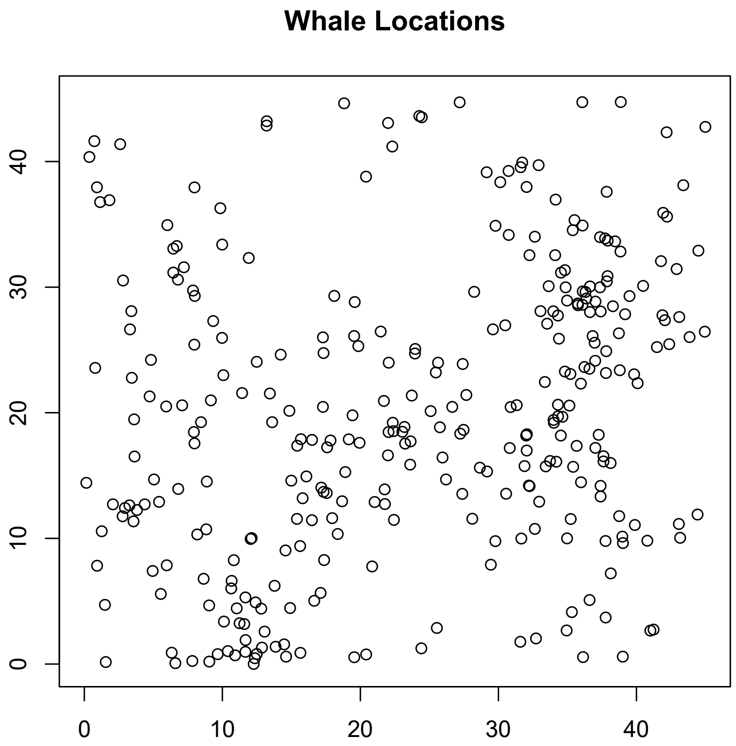


Truth



319
NARWs

Thinning/Degradation to Observations, aka "Approximation of Reality"

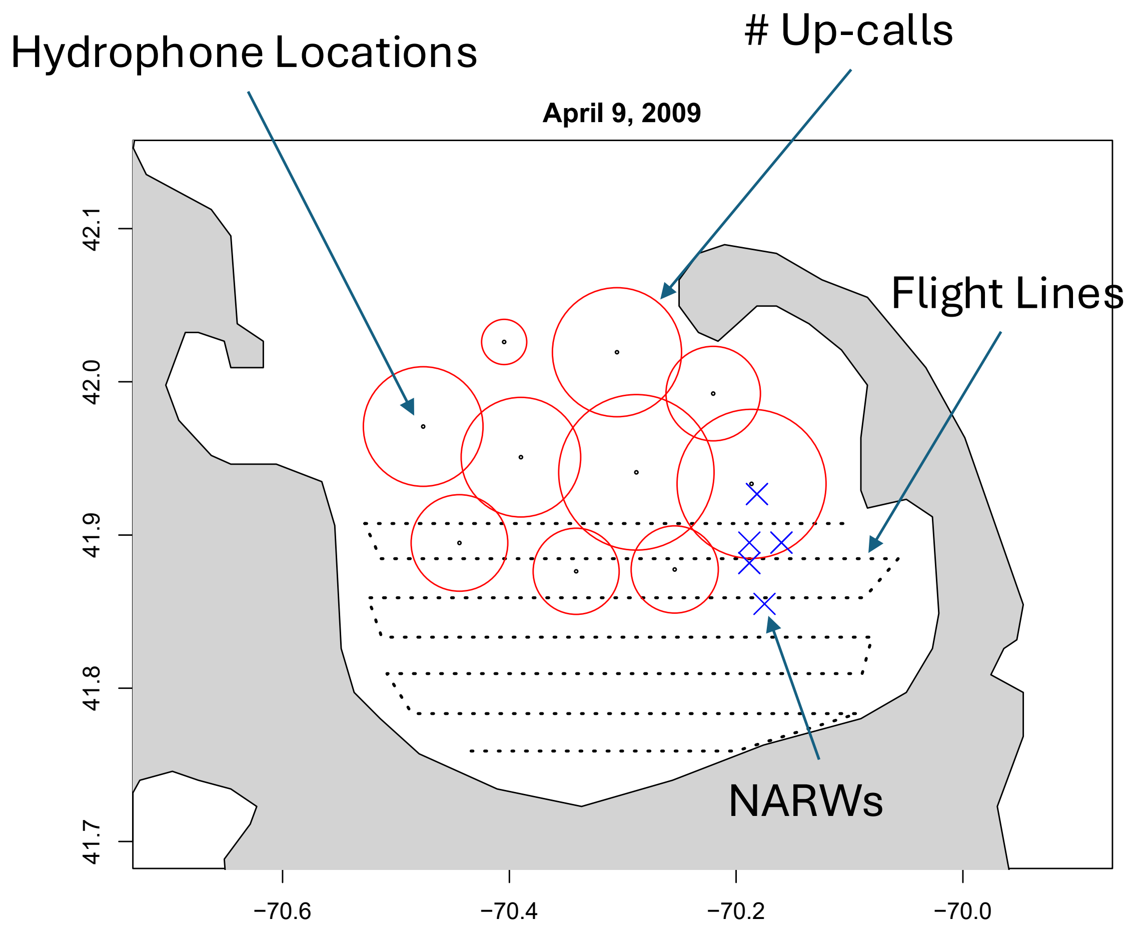


"Truth"

Observed

NARW in Cape Cod Bay

Real CCB Data



Key Assumptions

- Underlying spatial process that is…
- Fixed in time—snapshot
- Detection (thinning) functions are known
- Ancillary DTAG data help with:
  - Availability
  - Call rate
- Could fix these
- Could assign priors

Full Likelihood

N_k are the **detected** calls

$$\mathcal{L} = \prod_{\ell=1}^{L} \prod_{\mathbf{s} \in \mathcal{S}_\ell} \lambda_{dist_\ell}(\mathbf{s}) \exp^{-\int_D \lambda_{dist_\ell}(\mathbf{s}) d\mathbf{s}} \times \prod_{k=1}^{K} \frac{\lambda_{pam_k}^{N_k} e^{-\lambda_{pam_k}}}{N_k!}$$

$$\times \prod_{j=1}^{J} \frac{\Gamma(\nu)}{\Gamma(\pi\nu)\Gamma((1-\pi)\nu)} z_j^{\pi\nu-1}(1-z_j)^{(1-\pi)\nu-1}$$

$$\times \prod_{i=1}^{I} \frac{\left(\frac{c}{\tau^2}\right)^{c^2/\tau^2}}{\Gamma\left(\frac{c^2}{\tau^2}\right)} y_i^{c^2/\tau^2-1} e^{-c/\tau^2 y_i}$$

S_l are the locations of NARW **observed** from plane

z_j are **ancillary** surfacings from DTAG

y_i are **ancillary** calls from DTAG

Assumptions on Call Rate and Abundance

The number of calls detected by hydrophone k is modeled as

Data, known well

$$N_k \sim Poisson(\lambda_{pam_k})$$

Parameter, assume knowledge*

where

Parameter, easy to estimate

Parameter, want to learn

$$\lambda_{pam_k} = c \int_D \lambda(\boldsymbol{s}) p_{pam_k}(\boldsymbol{s}) d\boldsymbol{s}$$

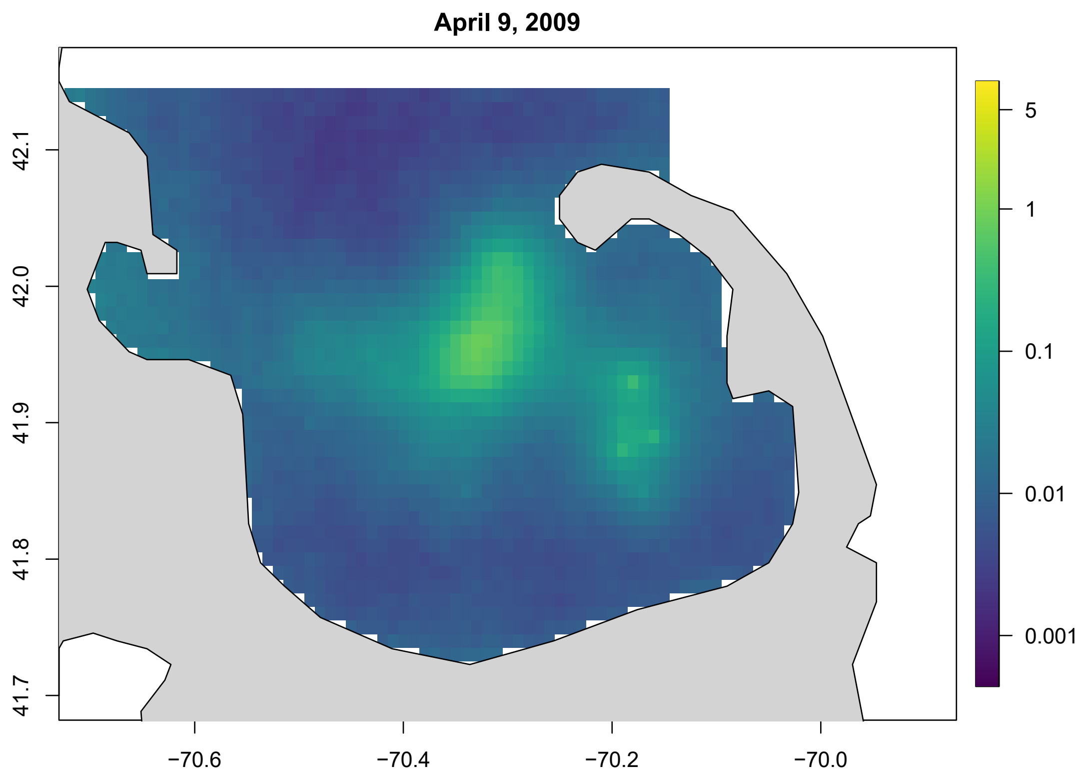
Det function, known well (Palmer et al. 2022)

- c is the average number of calls made per whale.
- $p_{pam_k}(\boldsymbol{s})$ is the detection function for hydrophone k and is a function of distance, ambient noise, and source level of the call

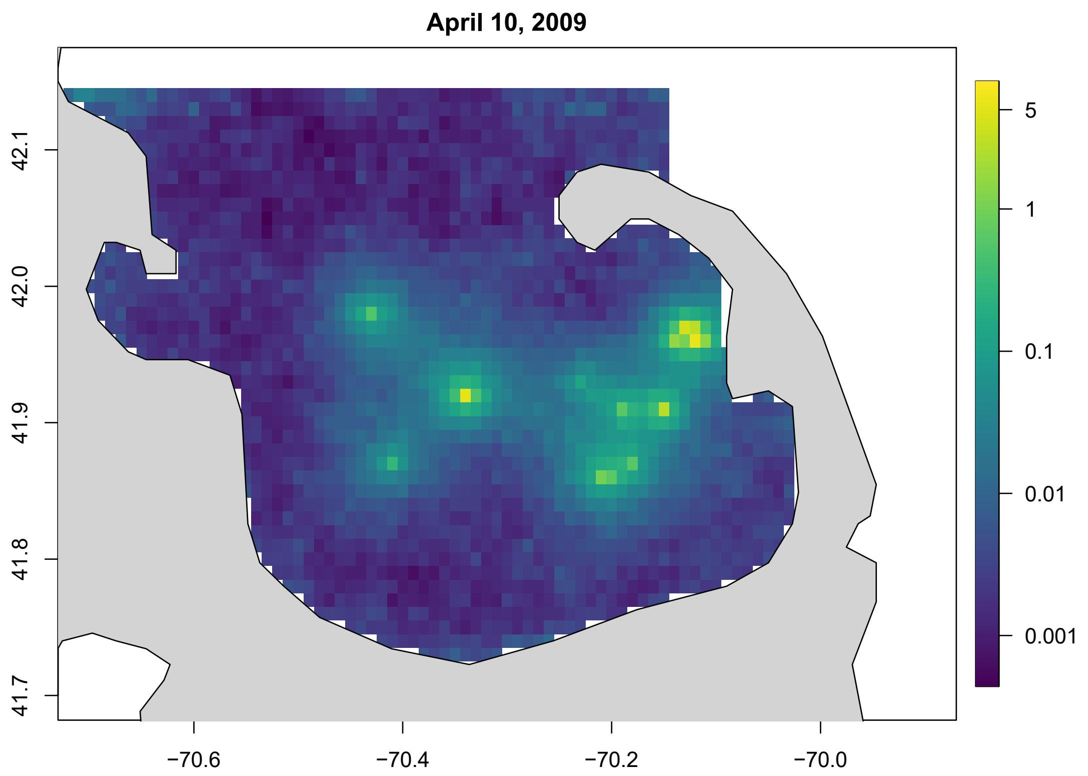
*Ideally, we'd learn $c_{s,t}$ but that's too rich a specification for these data

Results

Estimated NARW Abundance (Schliep et al., 2024)

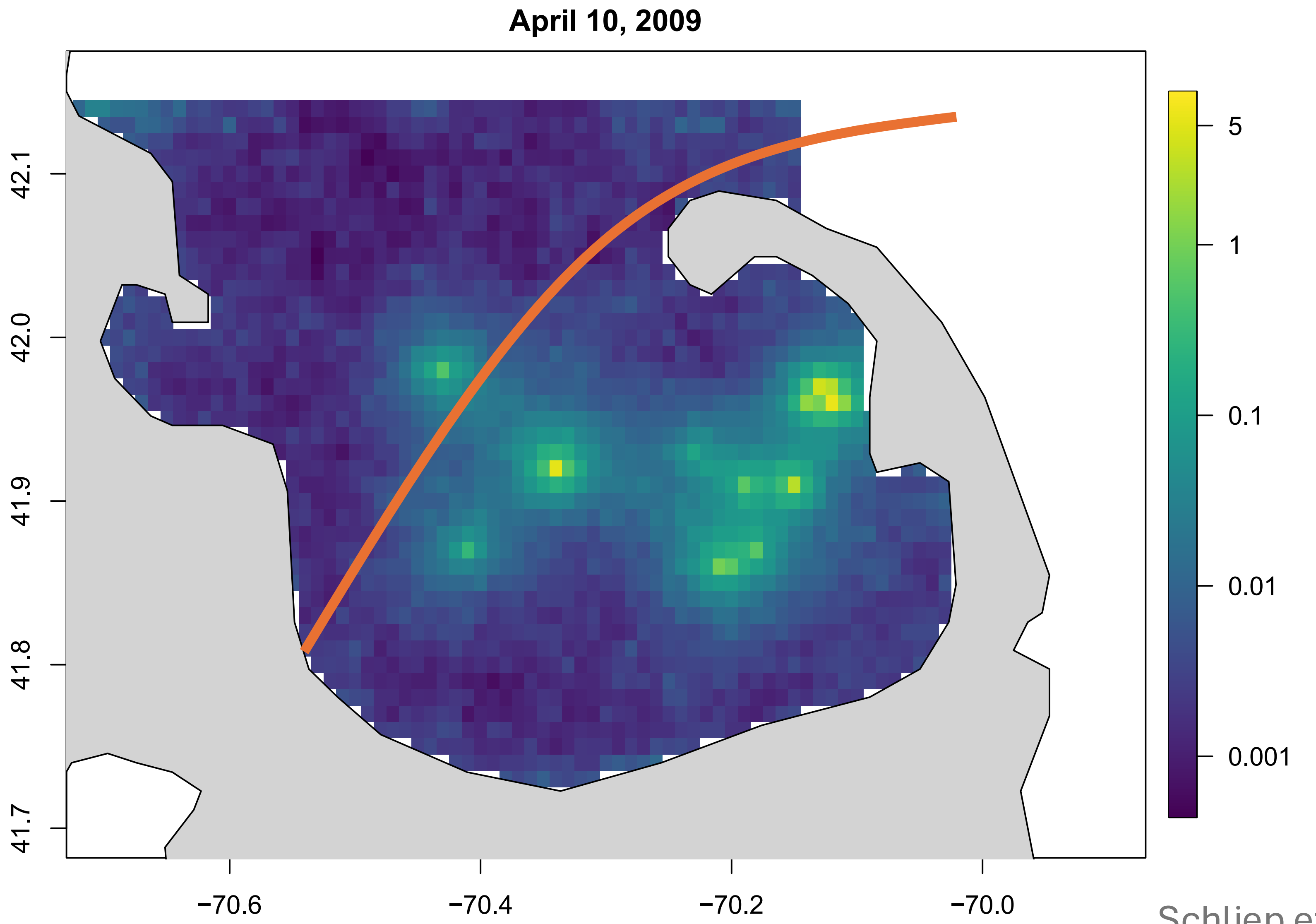


N	63.53 (14.97)



N	53.04 (8.93)

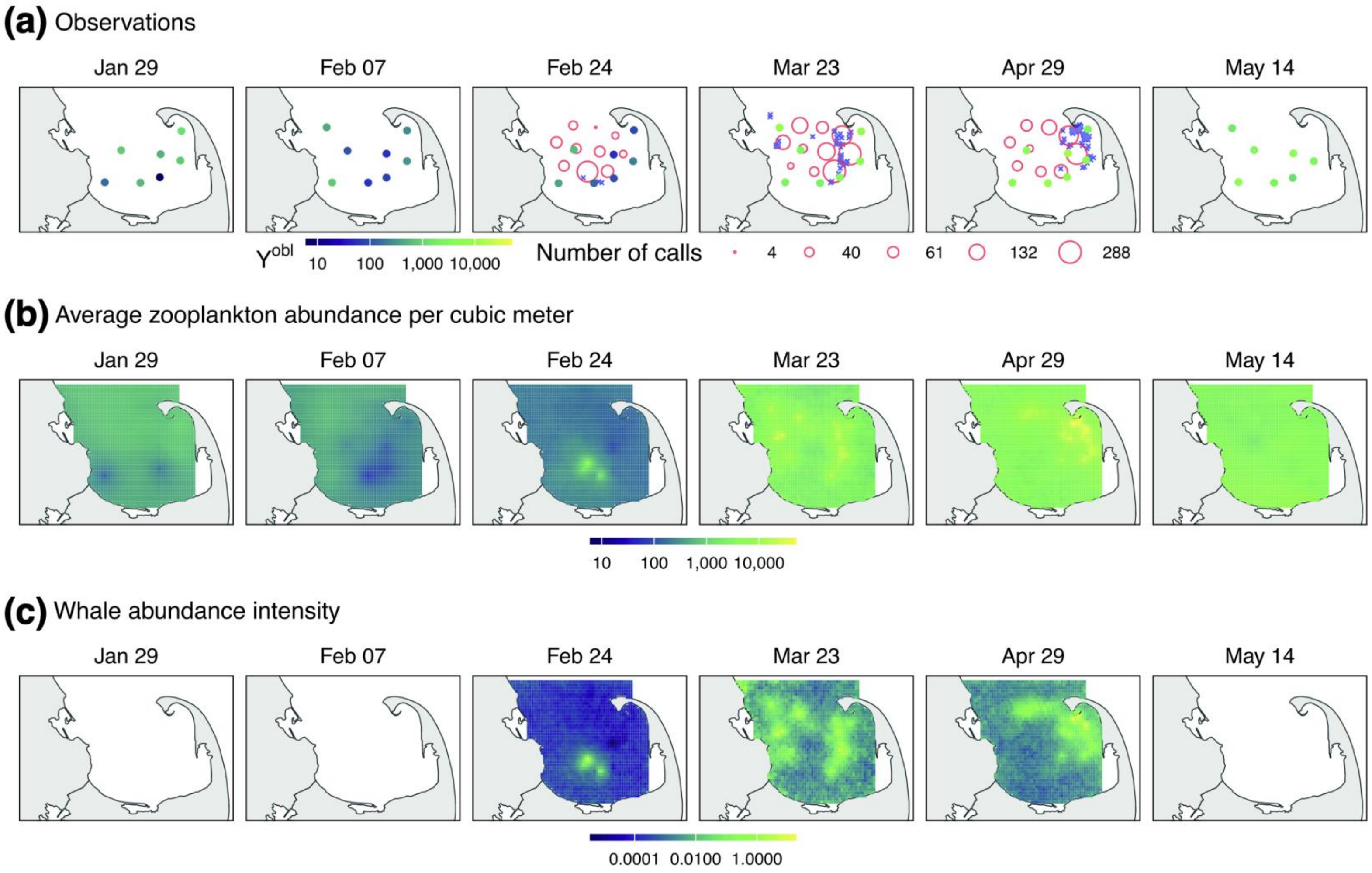
Spatial Inference—Relevance to Serious Injury



Schliep et al., (2024), AOAS

Extensions & Other Areas

Extensions—Incorporating Prey (Kang et al., 2025)



Building from Schliep et al. (2024)

Building from Castillo-Mateo et al. (2023)

Extensions, continued

- Spatial to spatio-temporal is *hard* because only continuous data (PAM) don’t identify individuals
- With localized calls, you could develop and fit a spatial-temporal point process model (we’re working on this), but…
- …latent intensity represents spatio-temporal patterns of calls, not individuals
- Spatial to spatio-temporal is *hard* without:
  - Underlying movement model
  - Linking calls to specific individuals
  - Behavioral- & individual-specific call rates

Application to Other Systems

- Anywhere with similar data structures
  - Line transect data
  - PAM Array
  - *Ancillary data on availability help to inform* π
  - *Ancillary data on call rates help to inform* c
- Southern New England – NARW (Laura Ganley, NEAq)
- Monterey Bay – Harbor porpoise (Eiren Jacobson, St Andrews)
- Gulf of Maine – NARW (Anita Murray, State of Maine)

Papers Mentioned

Castillo-Mateo, J., A. E. Gelfand, C. A. Hudak, C. A. Mayo, and R. S. Schick. "Space-time multi-level modeling for zooplankton abundance employing double data fusion and calibration." *Environmental and Ecological Statistics* 30, no. 4 (2023): 769-795

Schliep, E. M., A. E. Gelfand, C. W. Clark, C. A. Mayo, B. McKenna, S. E. Parks, T. M. Yack, and R. S. Schick. "Assessing marine mammal abundance: A novel data fusion." *The Annals of Applied Statistics* 18, no. 4 (2024): 3071-3090

Kang, B., E. M. Schliep, A. E. Gelfand, C. W. Clark, C. A. Hudak, C. A. Mayo, R. Schosberg, T. M. Yack, R. S. Schick, “Joint spatiotemporal modelling of zooplankton and whale abundance in a dynamic marine environment.” *Journal of the Royal Statistical Society Series C: Applied Statistics*, 2025; qlaf038

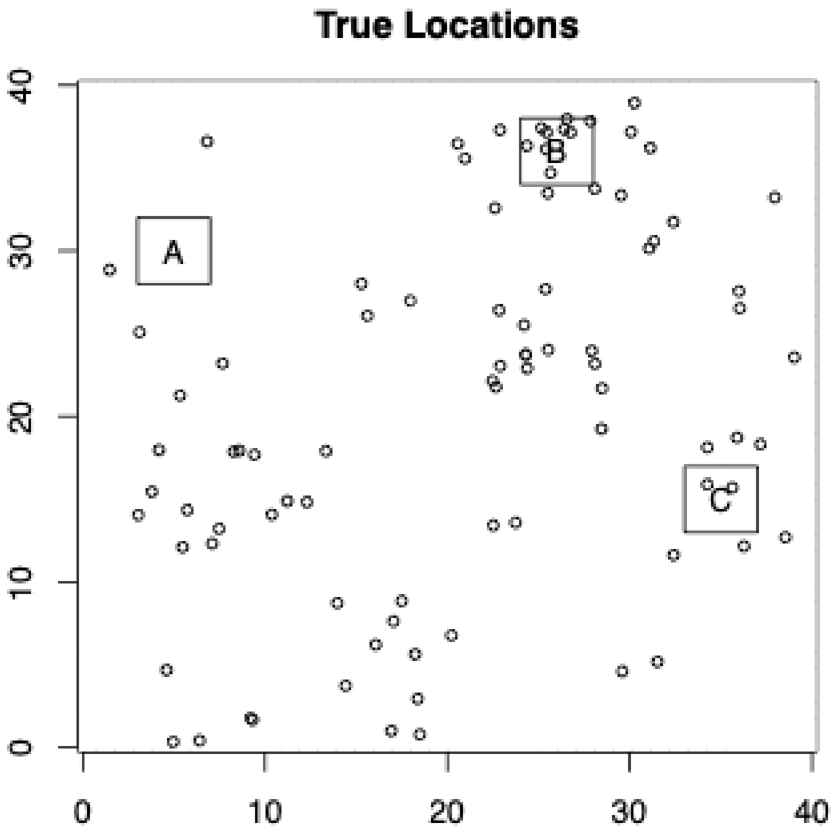
Questions?

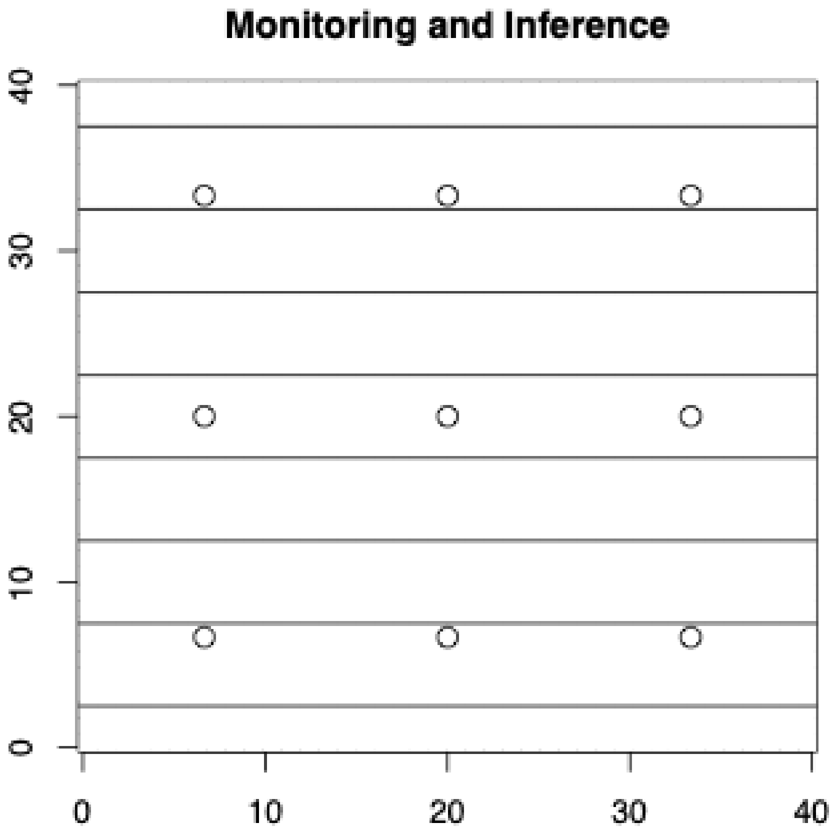
Credit: Dr. John Durban, WHOI, NOAA, SeaLife Response, Rehabilitation and Research; under NMFS Research Permit #17355.

Simulation Study

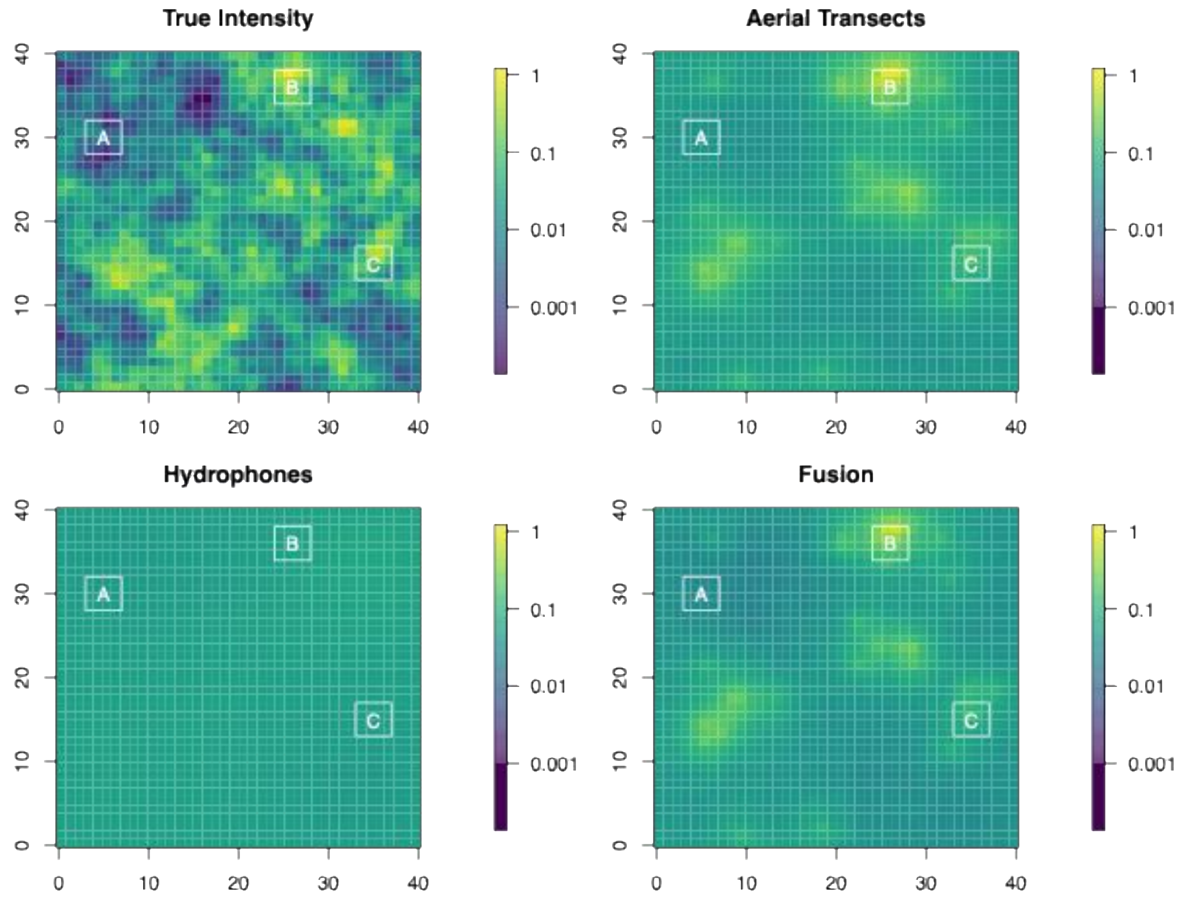
Inspired by CCB

Simulated Point Pattern





Estimating True Intensity Surface



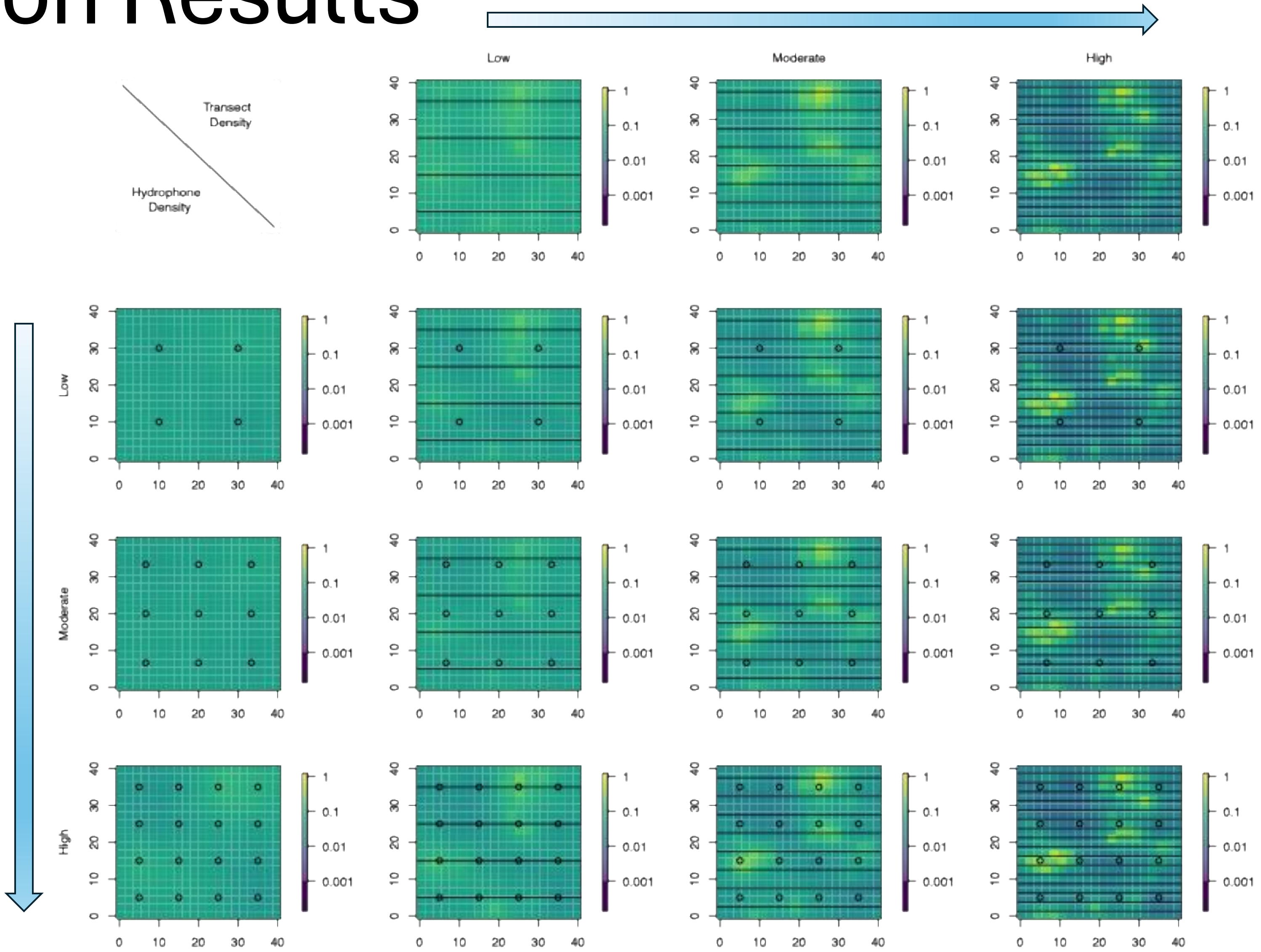
Summary

RMSE of the log posterior intensity and log posterior density

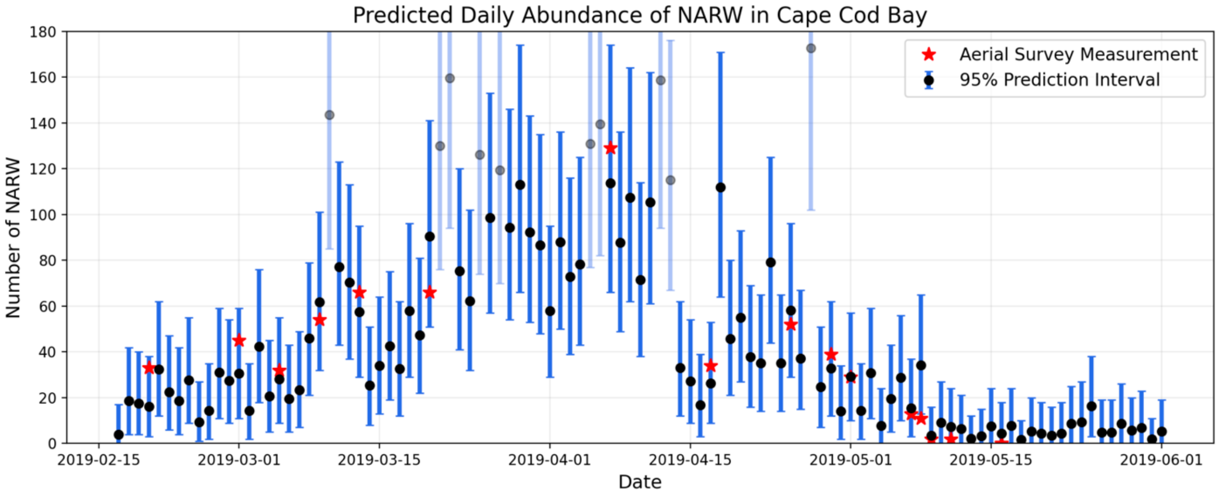
	Total	A	B	C	Log posterior density
Aerial Transect	1.90	2.91	1.21	1.29	-348.97 (12.56)
Hydrophones	1.89	3.08	1.60	1.41	-374.14 (19.67)
Fusion	1.65	2.40	0.97	1.19	-348.29 (14.59)

Simulation Results

More Hydrophones



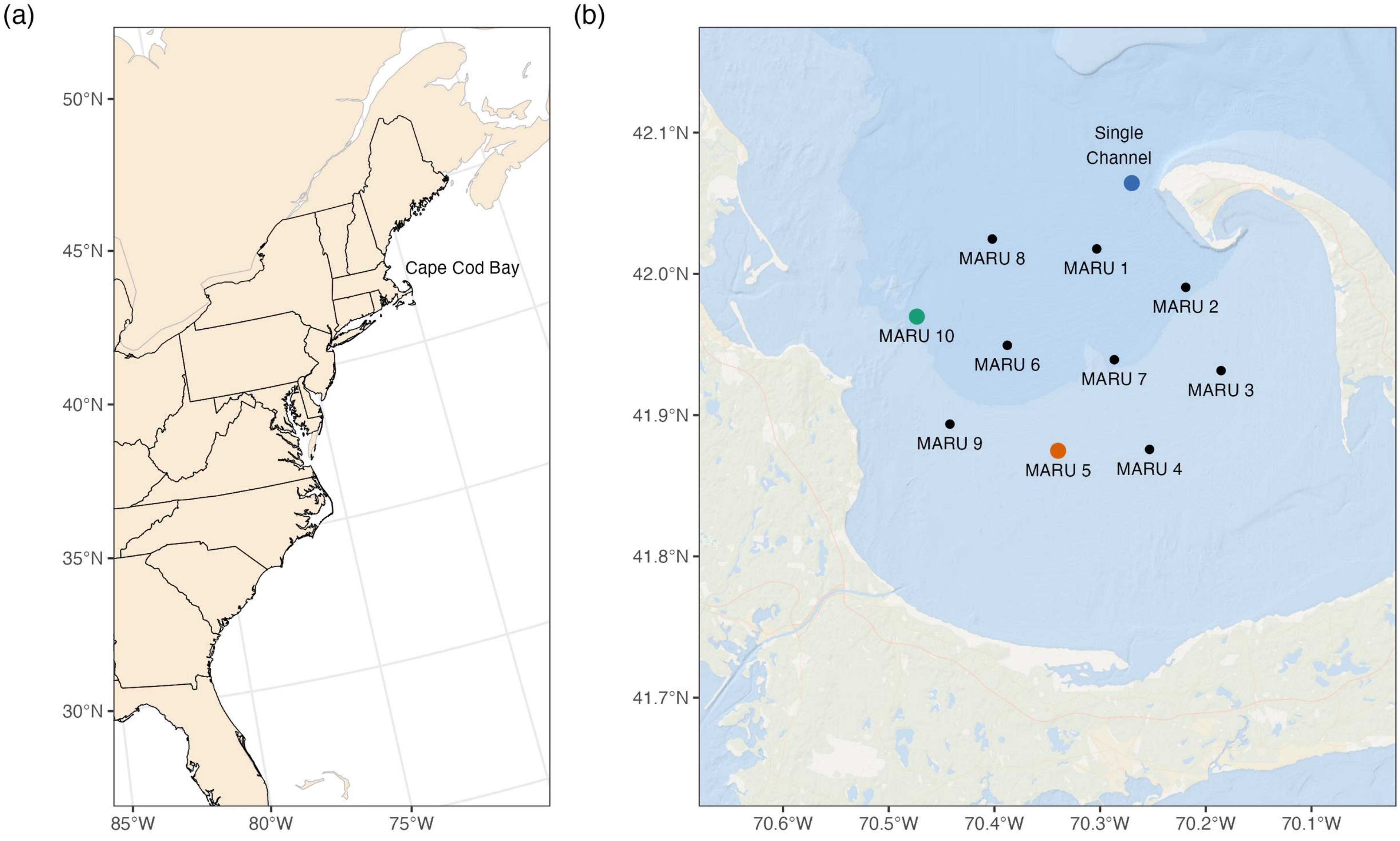
Other Methods



Garcia et al. (2025), ESR 56:101-115

- Calibration between aerial-acoustic overlap days
- No process model
- Whales ~ intercept + calls

Extra Slides


(a)
50°N
45°N
40°N
35°N
30°N
85°W
80°W
75°W
Cape Cod Bay
(b)
42.1°N
42.0°N
41.9°N
41.8°N
41.7°N
70.6°W
70.5°W
70.4°W
70.3°W
70.2°W
70.1°W
Single Channel
MARU 8
MARU 1
MARU 2
MARU 10
MARU 6
MARU 7
MARU 3
MARU 9
MARU 5
MARU 4

Thinning Equations

- Distance Sampling

Let $d(\boldsymbol{s}, \ell)$ denote the distance between location $\boldsymbol{s}$ and transect ℓ.

$$p_{dist_\ell}(\boldsymbol{s}) = \begin{cases} 1 & d(\boldsymbol{s}, \ell) \leq 0.75km \\ \exp(-(d(\boldsymbol{s}, \ell) - 0.75)^2) & d(\boldsymbol{s}, \ell) > 0.75km \end{cases}$$

Ganley et al. (2019)

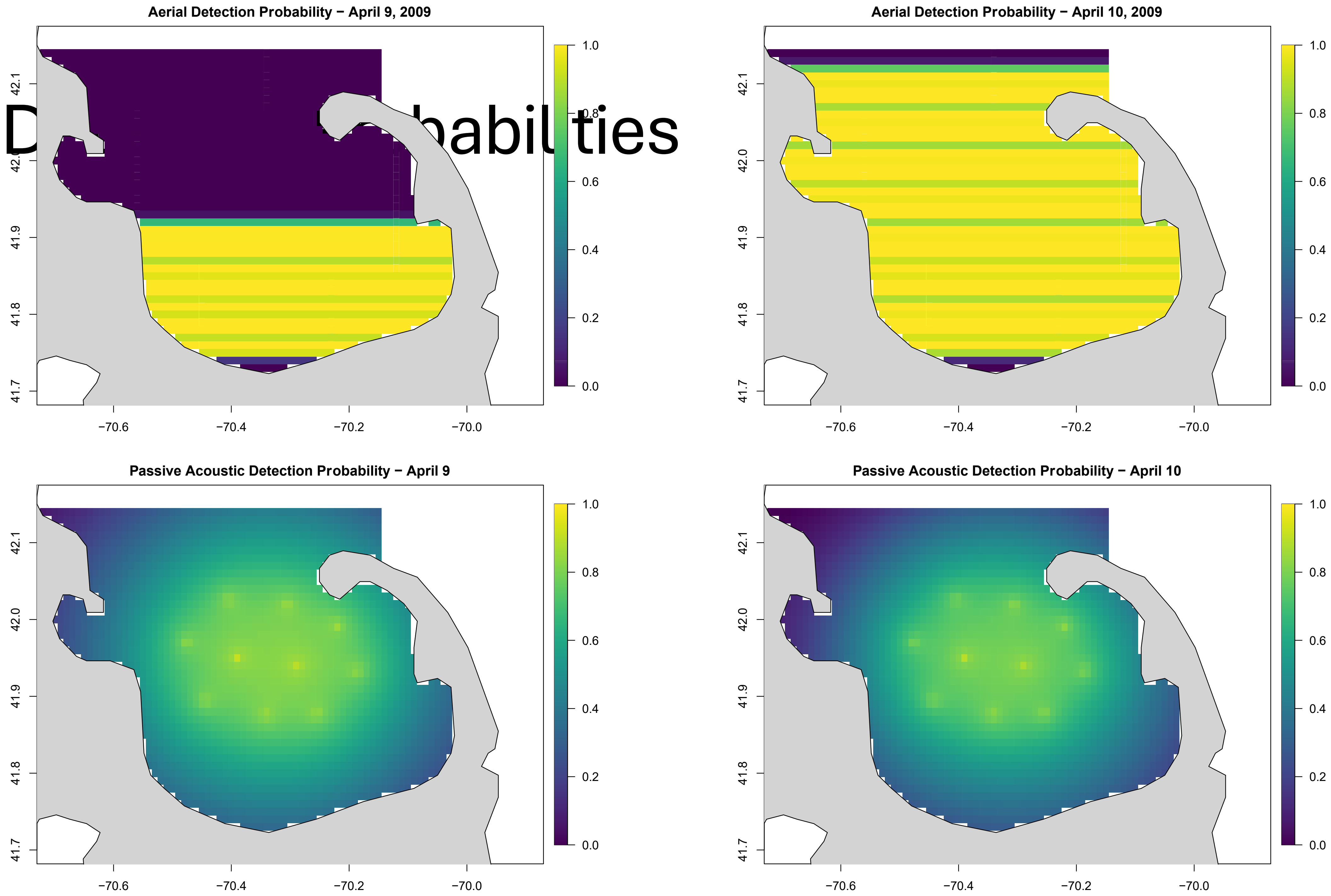
- Acoustic

Source level is assumed Uniform(141, 197).

$$p_{pam_k}(\boldsymbol{s}) = P(SL(\boldsymbol{s}) - 14.5\log_{10}(d(\boldsymbol{s}, \boldsymbol{s}_k^*)) > 104 + 26)$$

$$p_{pam_k}(\boldsymbol{s}) = \begin{cases} 0 & 14.5\log_{10}(d(\boldsymbol{s}, \boldsymbol{s}_k^*)) > 197 - 130 \\ 1 & 14.5\log_{10}(d(\boldsymbol{s}, \boldsymbol{s}_k^*)) < 141 - 130 \\ 1.20 - 0.26\log_{10}(d(\boldsymbol{s}, \mathbf{h}_k)) & else \end{cases}$$

Detection Probabilities



The full likelihood is the product of the two source likelihoods as well as the two auxiliary data sources

$$\mathcal{L} = \prod_{\ell=1}^{L} \prod_{\boldsymbol{s}\in\mathcal{S}_\ell} \lambda_{dist_\ell}(\boldsymbol{s}) \exp^{-\int_D \lambda_{dist_\ell}(\boldsymbol{s})d\boldsymbol{s}} \times \prod_{k=1}^{K} \frac{\lambda_{pam_k}^{N_k} e^{-\lambda_{pam_k}}}{N_k!}$$

$$\times \prod_{j=1}^{J} \frac{\Gamma(\nu)}{\Gamma(\pi\nu)\Gamma((1-\pi)\nu)} z_j^{\pi\nu-1}(1-z_j)^{(1-\pi)\nu-1}$$

$$\times \prod_{i=1}^{I} \frac{\left(\frac{c}{\tau^2}\right)^{c^2/\tau^2}}{\Gamma\left(\frac{c^2}{\tau^2}\right)} y_i^{c^2/\tau^2-1} e^{-c/\tau^2 y_i}$$

Sightings from Aerial Line Transects (CCS)

Visual Sighting Detection Function (Ganley et al. 2019)

Up-calls (Clark)

Acoustic Detection Function (Palmer et al. 2022)

Availability from DTAG dive data (Susan Parks and Dave Wiley)

Call rates from DTAGs (Susan Parks and Dave Wiley)

$$p_{dist,\ell}(\mathbf{s}) = \pi f(d(\mathbf{s},\ell))$$

$$\lambda_{PAM,k} = c \int_D \lambda(\mathbf{s}) p_{PAM,k}(\boldsymbol{s}) d\mathbf{s}$$

$$\mathcal{L} = \prod_{\ell=1}^{L} \prod_{\boldsymbol{s} \in S_\ell} \lambda_{dist_\ell}(\boldsymbol{s}) \exp^{-\int_D \lambda_{dist_\ell}(\boldsymbol{s}) d\boldsymbol{s}} \times \prod_{k=1}^{K} \frac{\lambda_{pam_k}^{N_k} e^{-\lambda_{pam_k}}}{N_k!}$$

$$\times \prod_{j=1}^{J} \frac{\Gamma(\nu)}{\Gamma(\pi\nu)\Gamma((1-\pi)\nu)} z_j^{\pi\nu-1}(1-z_j)^{(1-\pi)\nu-1}$$

$$\times \prod_{i=1}^{I} \frac{\left(\frac{c}{\tau^2}\right)^{c^2/\tau^2}}{\Gamma\left(\frac{c^2}{\tau^2}\right)} y_i^{c^2/\tau^2-1} e^{-c/\tau^2 y_i}$$

- Surface data: $Z_j \overset{iid}{\sim} Beta(\mu = \pi, \nu = 15)$
  - $\pi \sim Uniform(0, 1)$

- Call data: $Y_i \overset{iid}{\sim} Gamma(\mu = c, \tau^2 = 10)$
  - $c \sim Uniform(0, 100)$