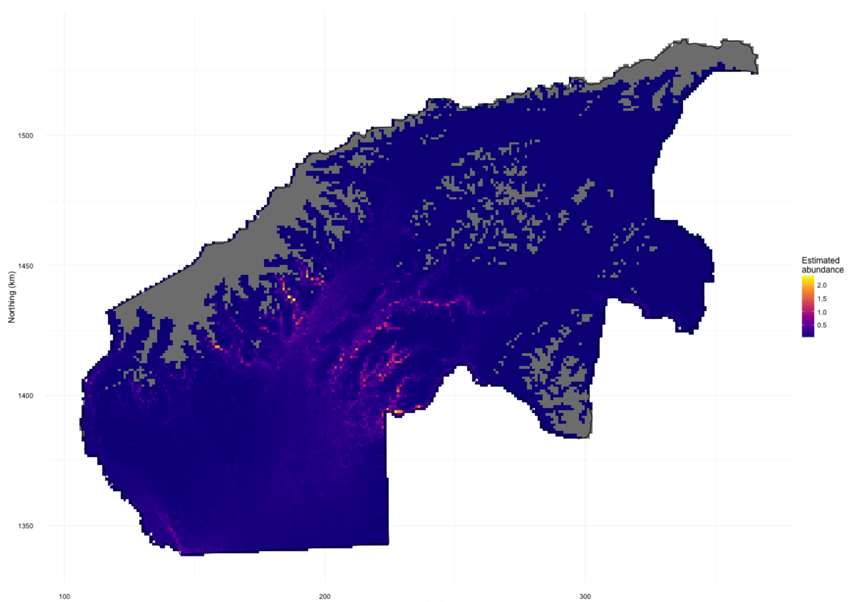
Lecture 1: distance sampling & density surface models

University of
St Andrews

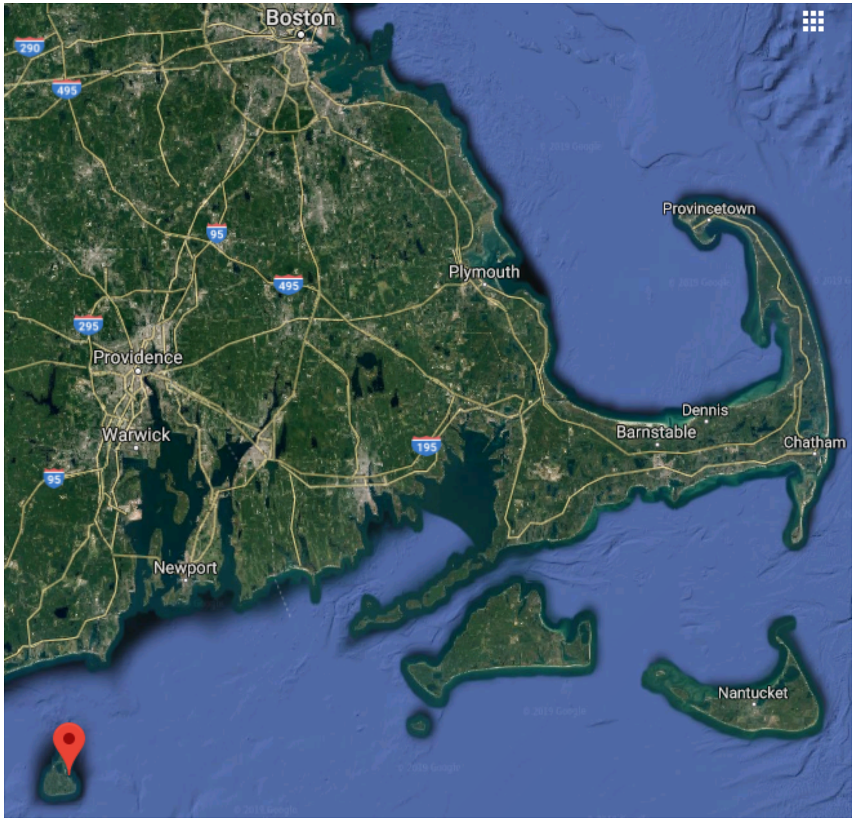
Why model abundance spatially?

Maps

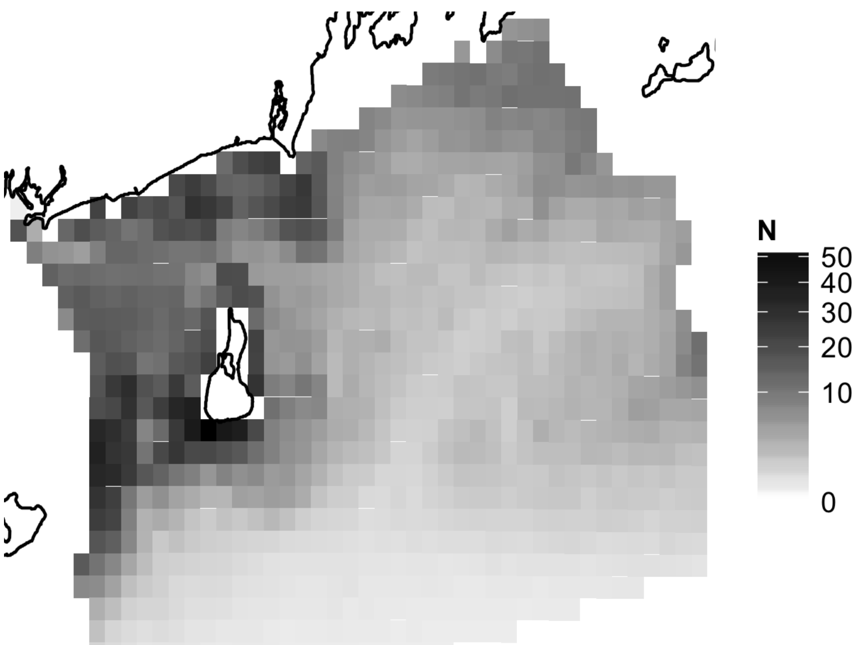


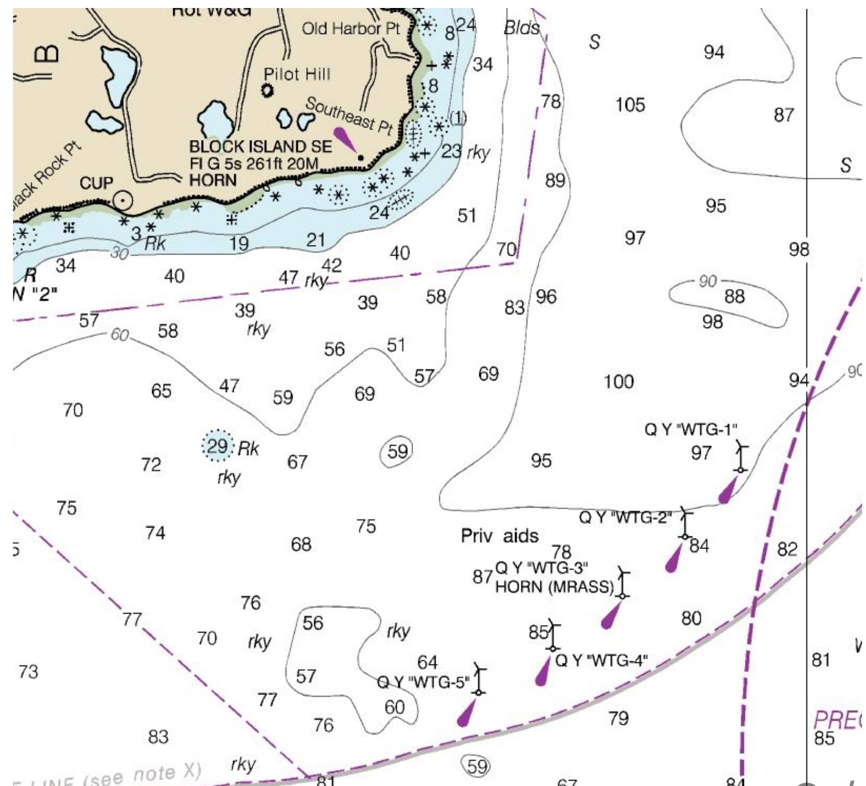
- Black bears in Alaska
- Heterogeneous spatial distribution

Spatial decision making



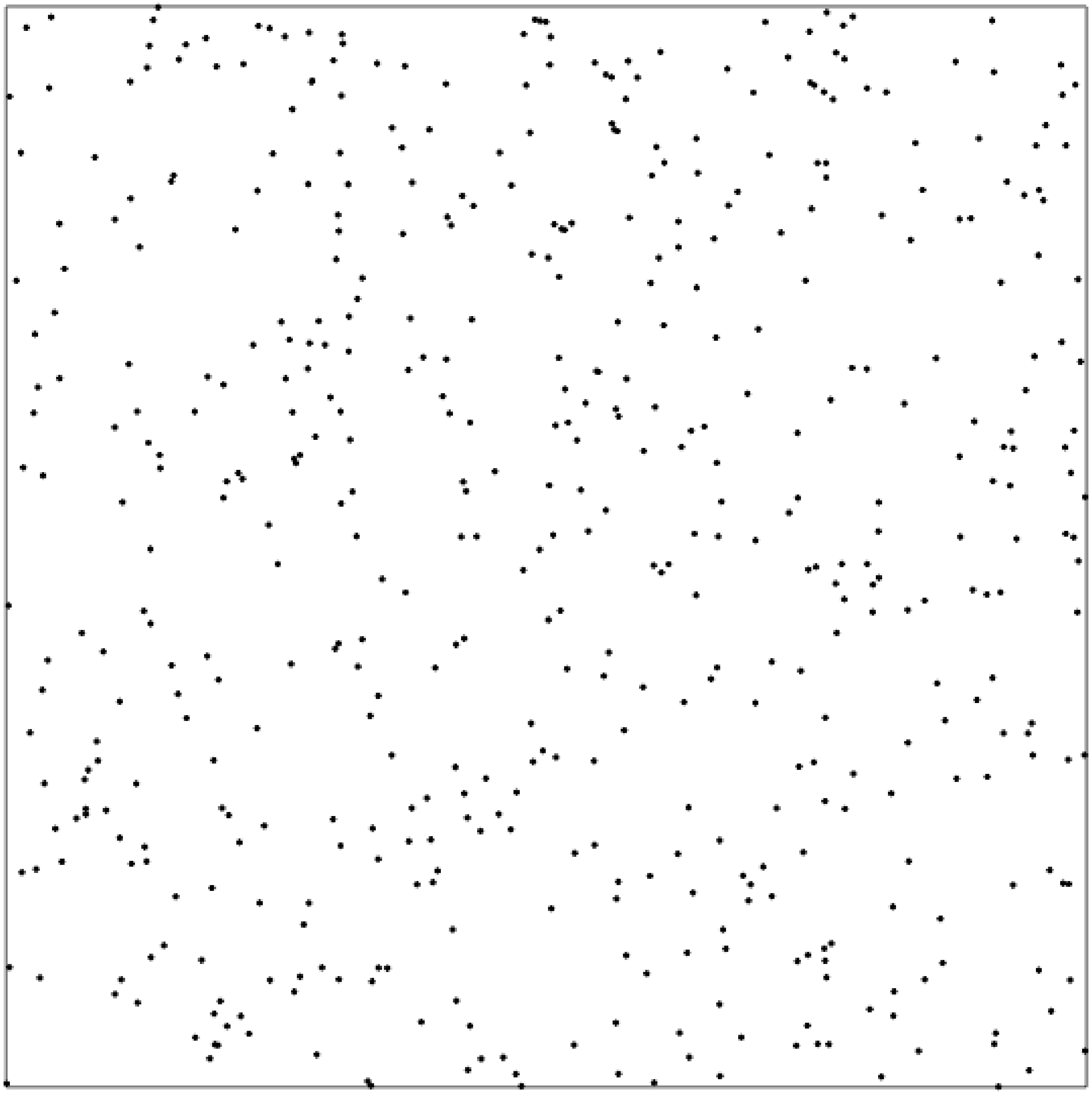
- Block Island, Rhode Island
- First offshore wind in the USA
- Spatial impact assessment



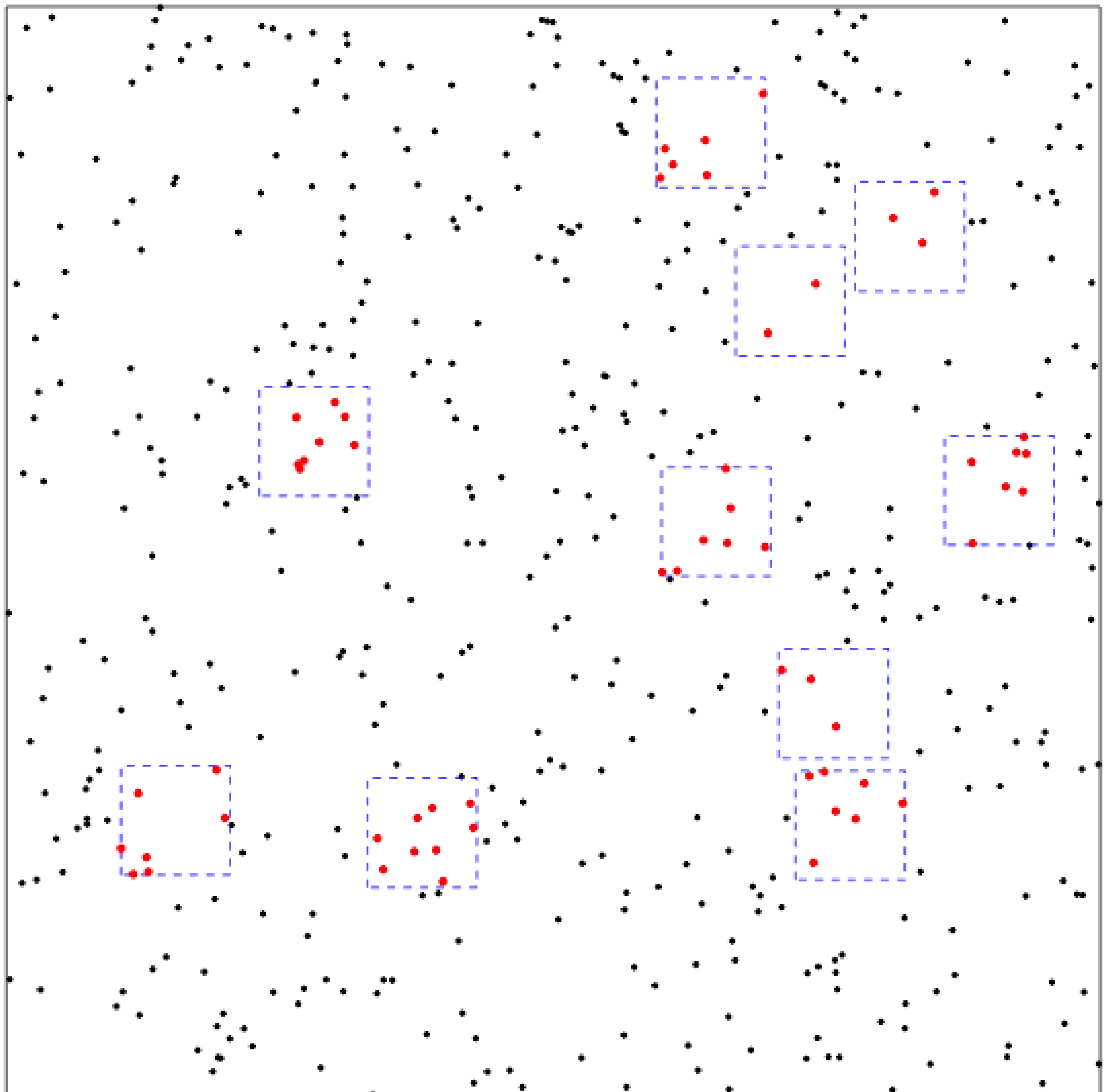


Back to regular distance sampling

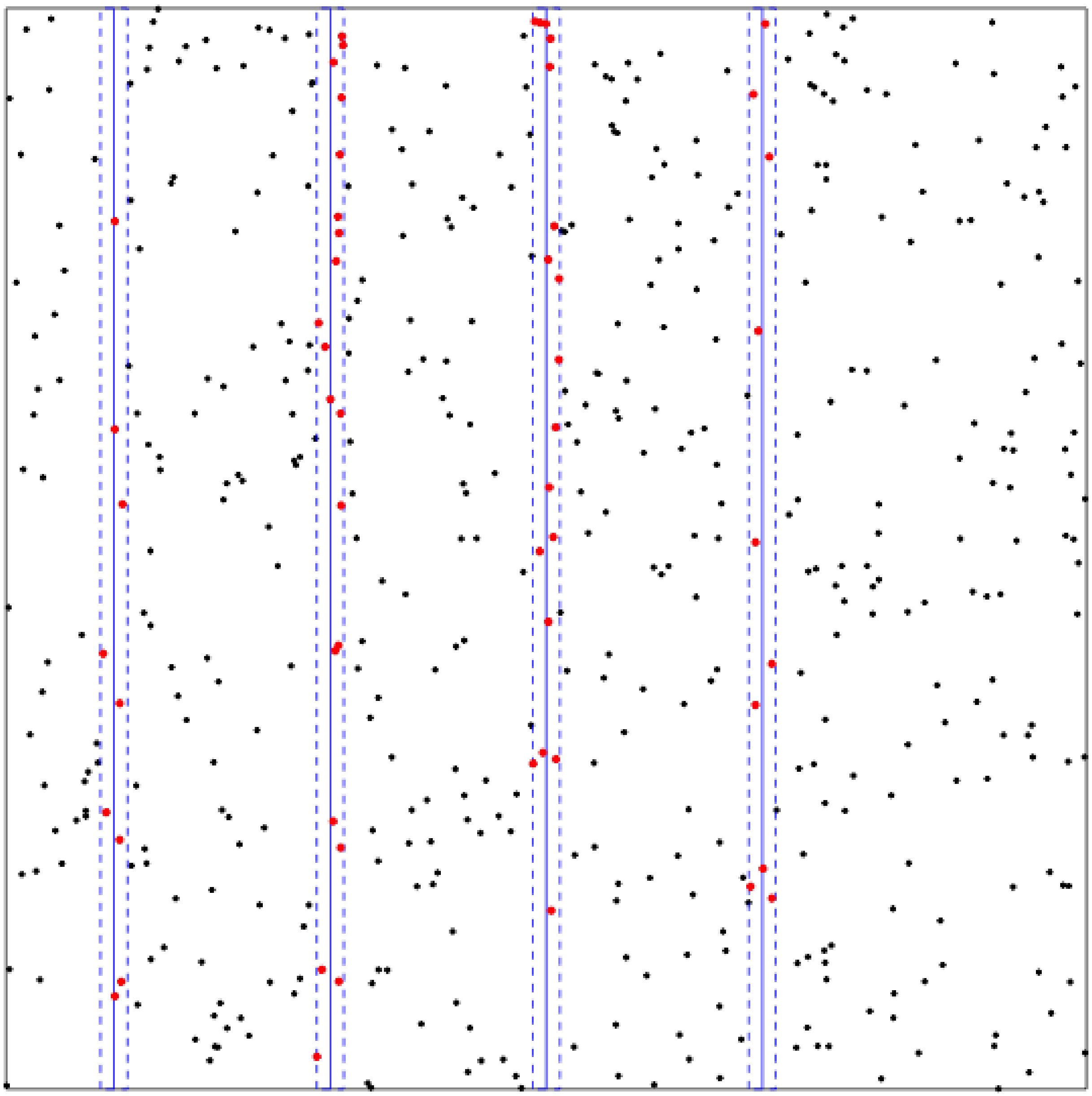
How many animals are there? (500!)



Plot sampling



Strip transect



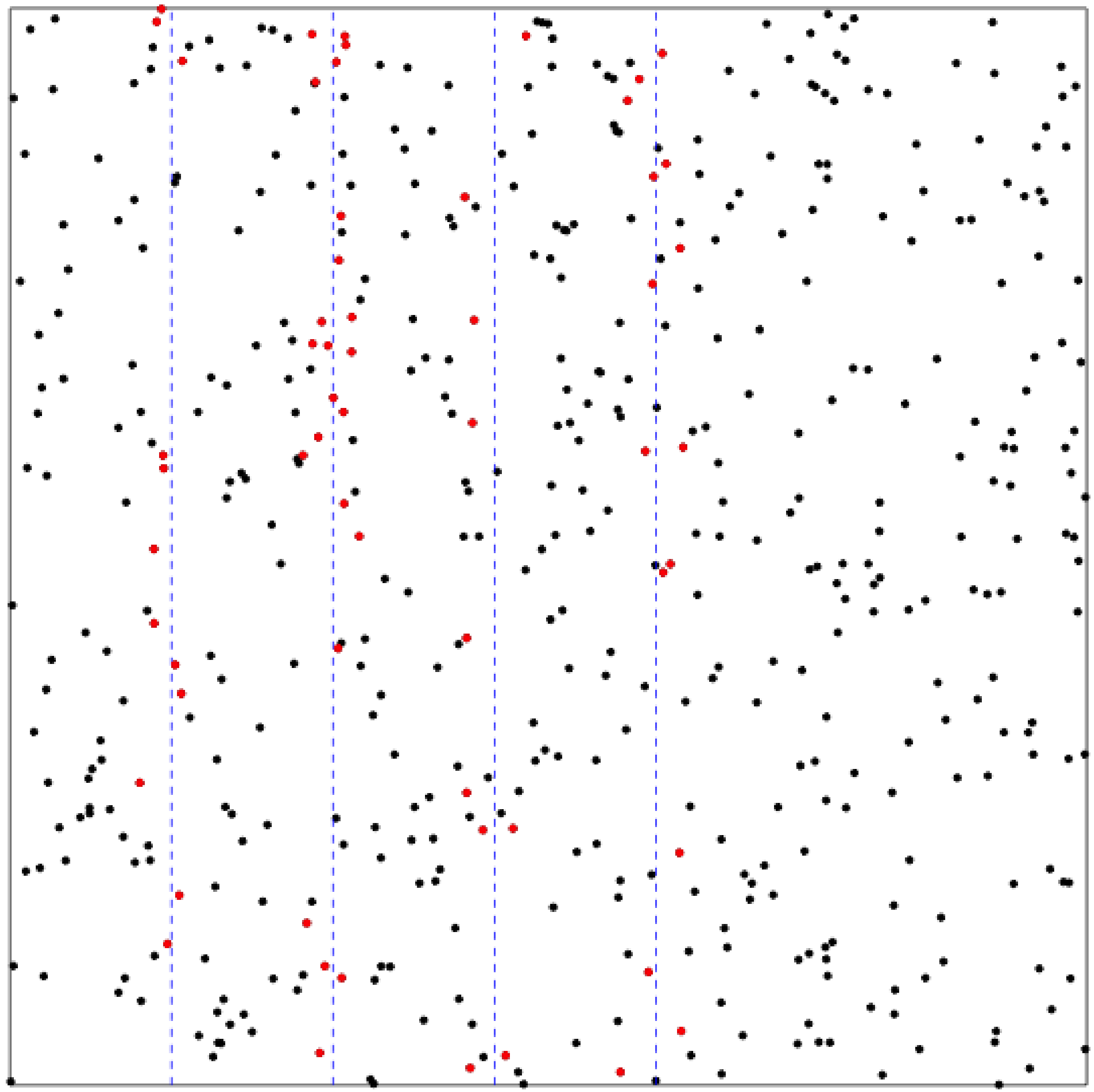
Detectability matters!

- We've assumed certain detection so far
- This rarely happens in the field
- Distance to the **object** is important
- Detectability should decrease with increasing distance

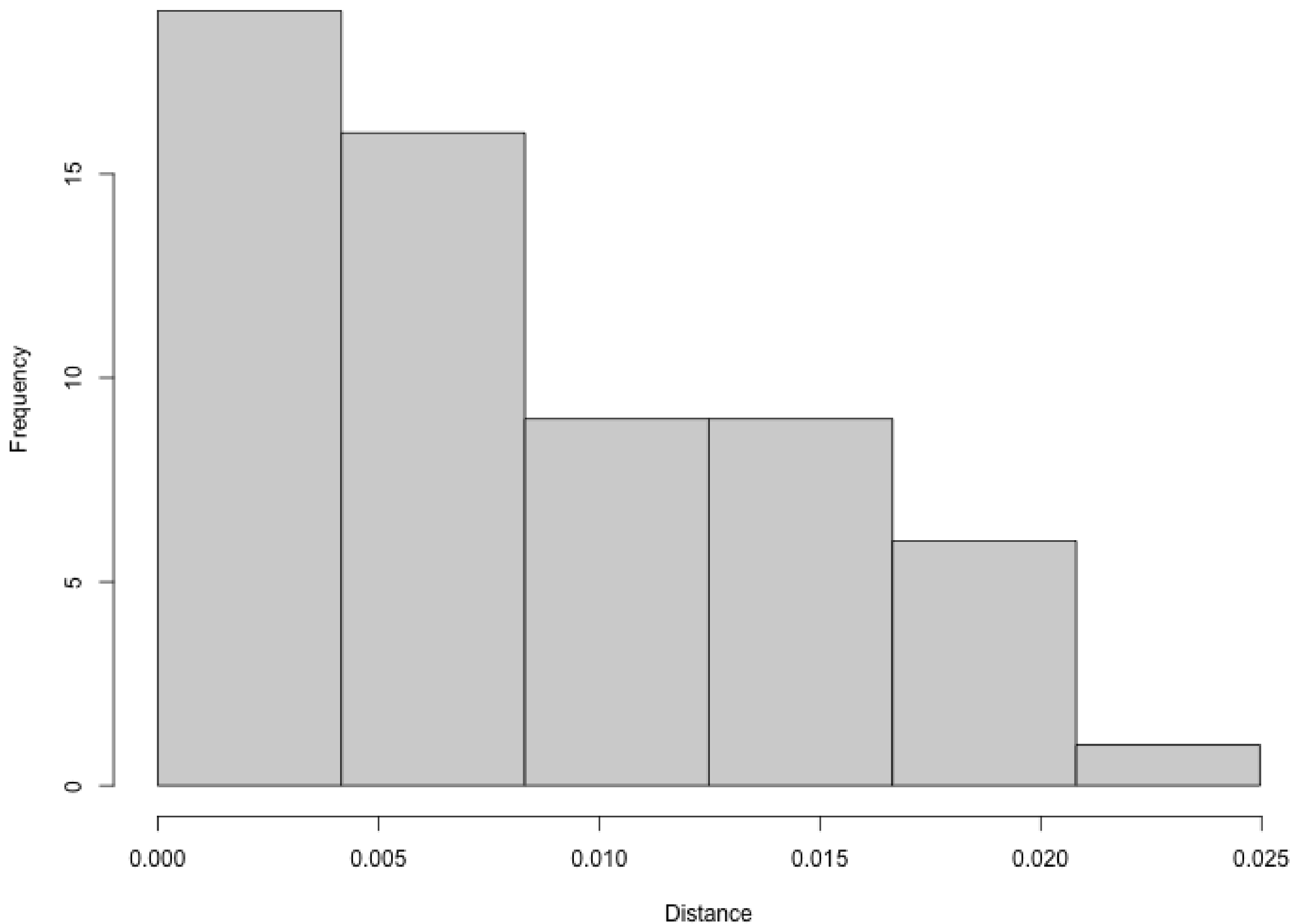
Distance and detectability

Credit Scott and Mary Flanders

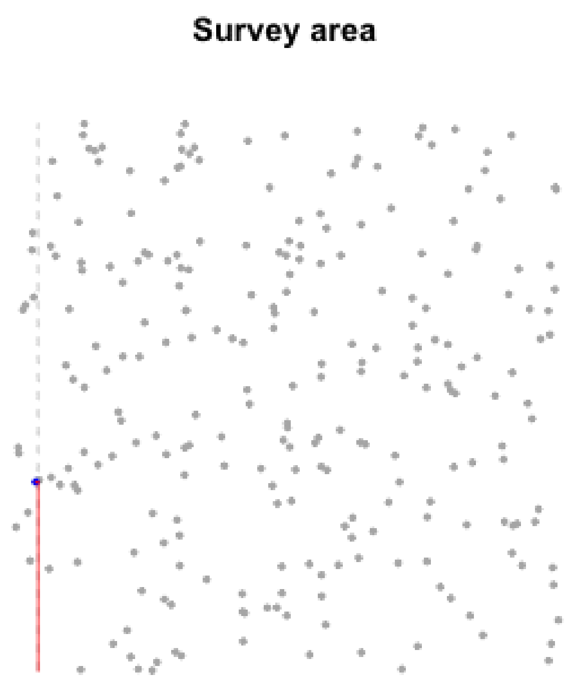
Line transect

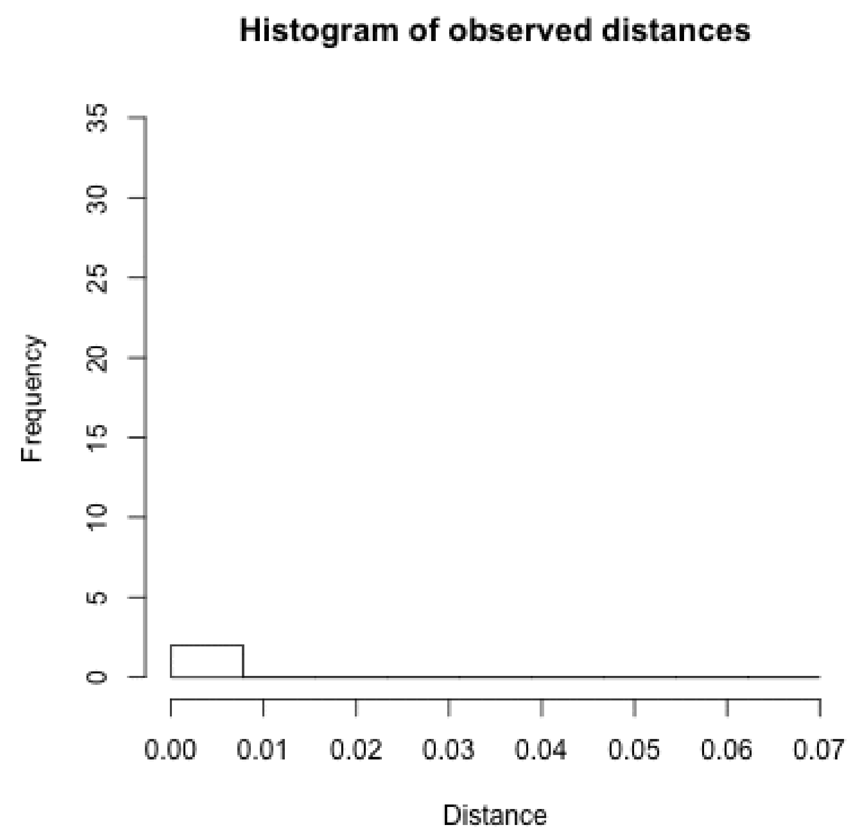


Line transects - distances

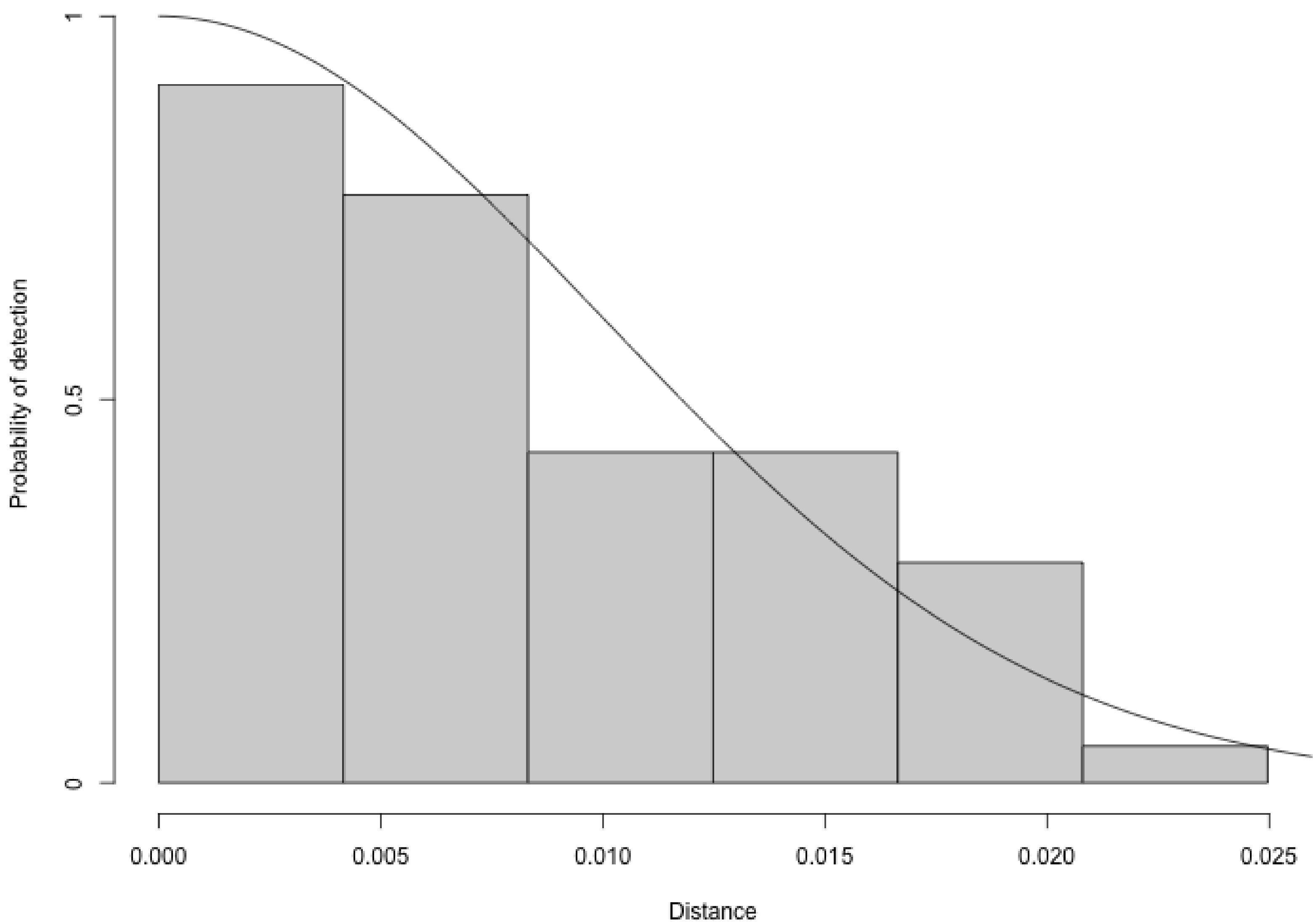


Distance sampling animation





Detection function



Distance sampling estimate

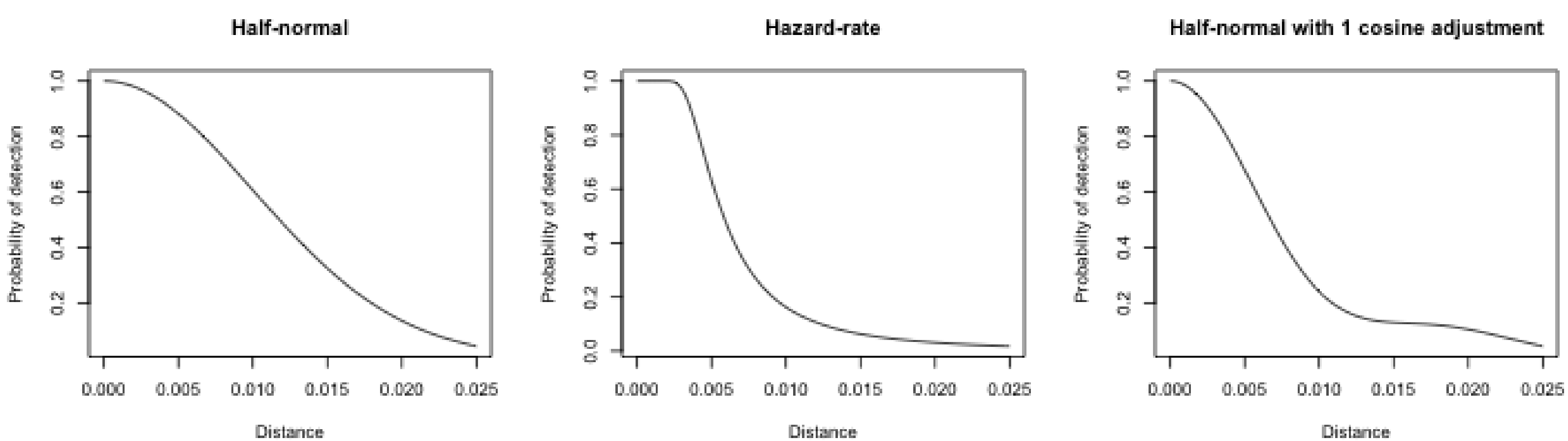
- Surveyed 5 lines (each $1 * 0.025$ units)
  - Total covered area $a = 5 * 1 * 0.02 =$ 0.2
- Probability of detection $\hat{p} = 0.5981$
- Saw $n = 60$ animals
- Inflate to $n/\hat{p} = 100.31$
- Estimated density $\hat{D} = \frac{n/\hat{p}}{a} = 502$
- Total area $A = 1$
- Estimated abundance $\hat{N} = \hat{D}A = 502$

Reminder of assumptions

1. Animals are distributed independent of lines
2. On the line, detection is certain
3. Distances are recorded correctly
4. Animals don't move before detection

What are detection functions?

- Model $\mathbb{P}$ (detection | animal at distance x)
- "Integrate out distance" == "area under curve" == $\hat{p}$
- Many different forms, depending on the data
- All share some characteristics



Fitting detection functions (in R!)

- Using the package `Distance`
- Function `ds()` does most of the work
- More on this in the practical!

```
library(Distance)
df_hn <- ds(distdata, truncation=6000)
```

Horvitz-Thompson-like estimators

- Once we have $\hat{p}$ how do we get $\hat{N}$?
- Rescale the (flat) density and extrapolate

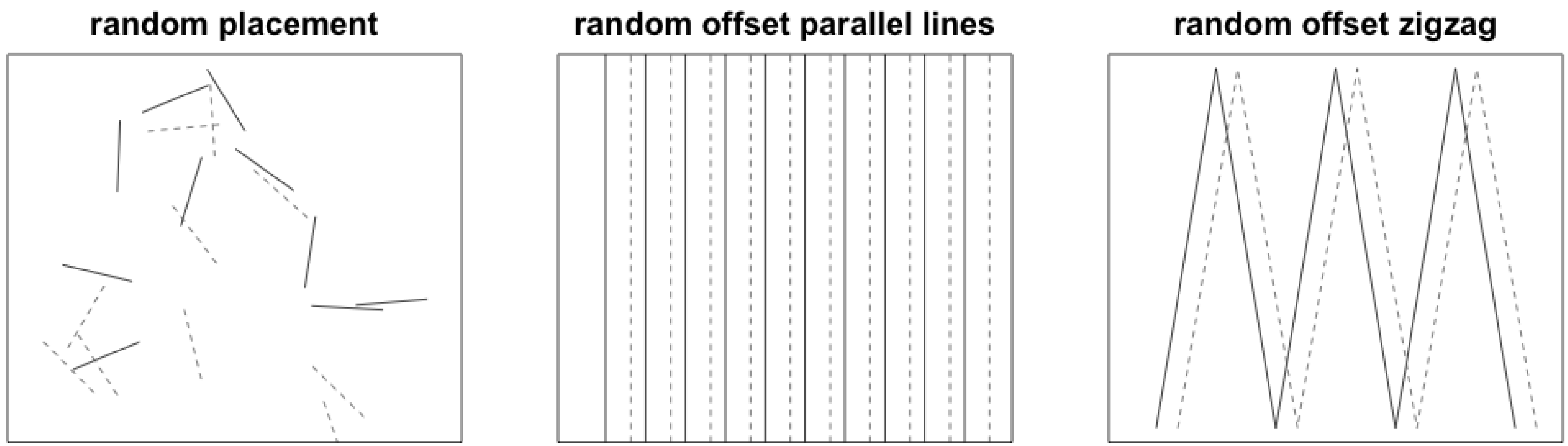
$$\hat{N} = \frac{\text{study area}}{\text{covered area}} \sum_{i=1}^{n} \frac{s_i}{\hat{p}_i}$$

- s_i are group/cluster sizes
- $\hat{p}_i$ is the detection probability (from detection function)

Hidden in this formula is a simple assumption

- Probability of sampling every point in the study area is equal
- Is this true? Sometimes.
- If (and only if) the design is randomised

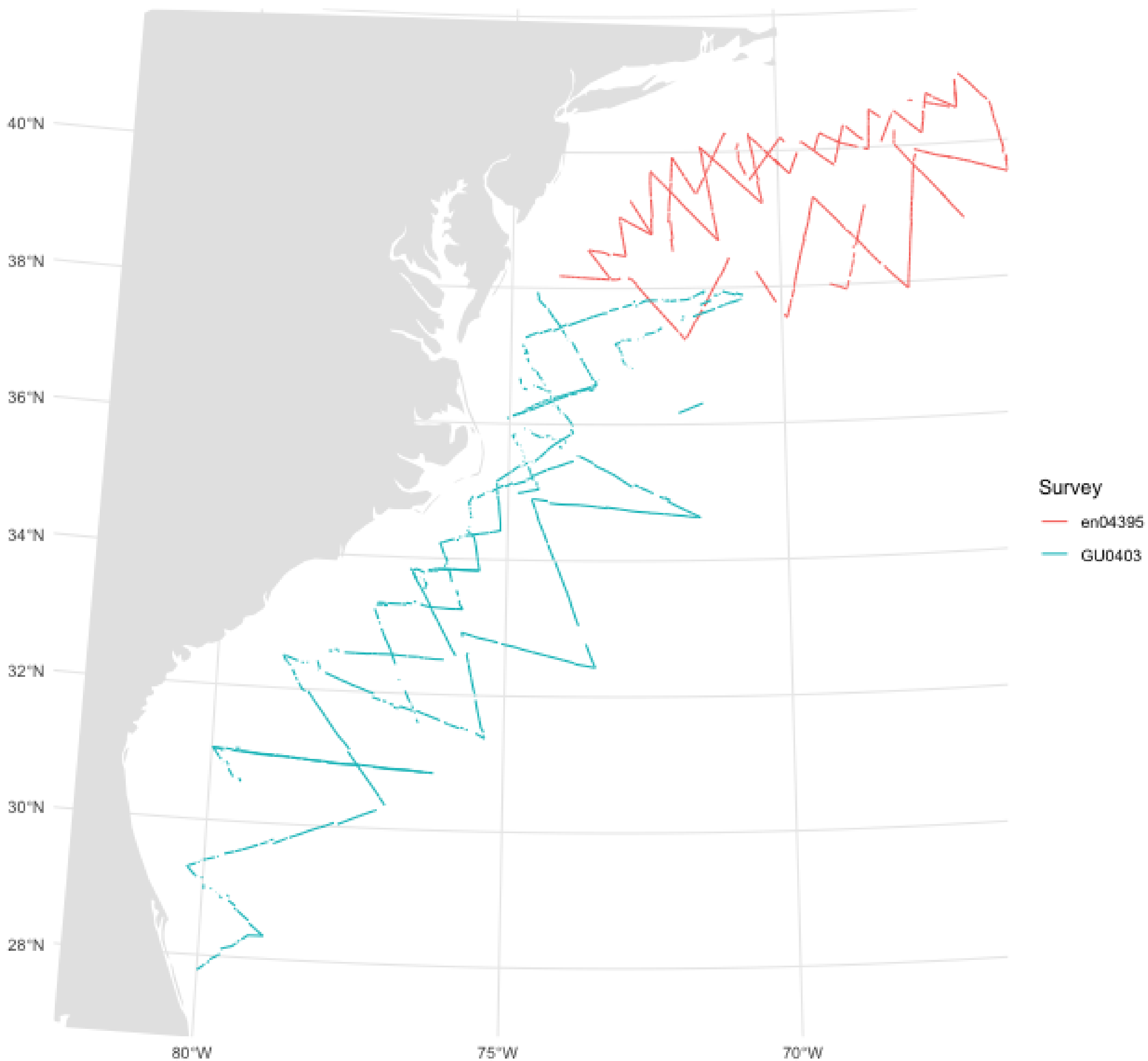
Many faces of randomisation



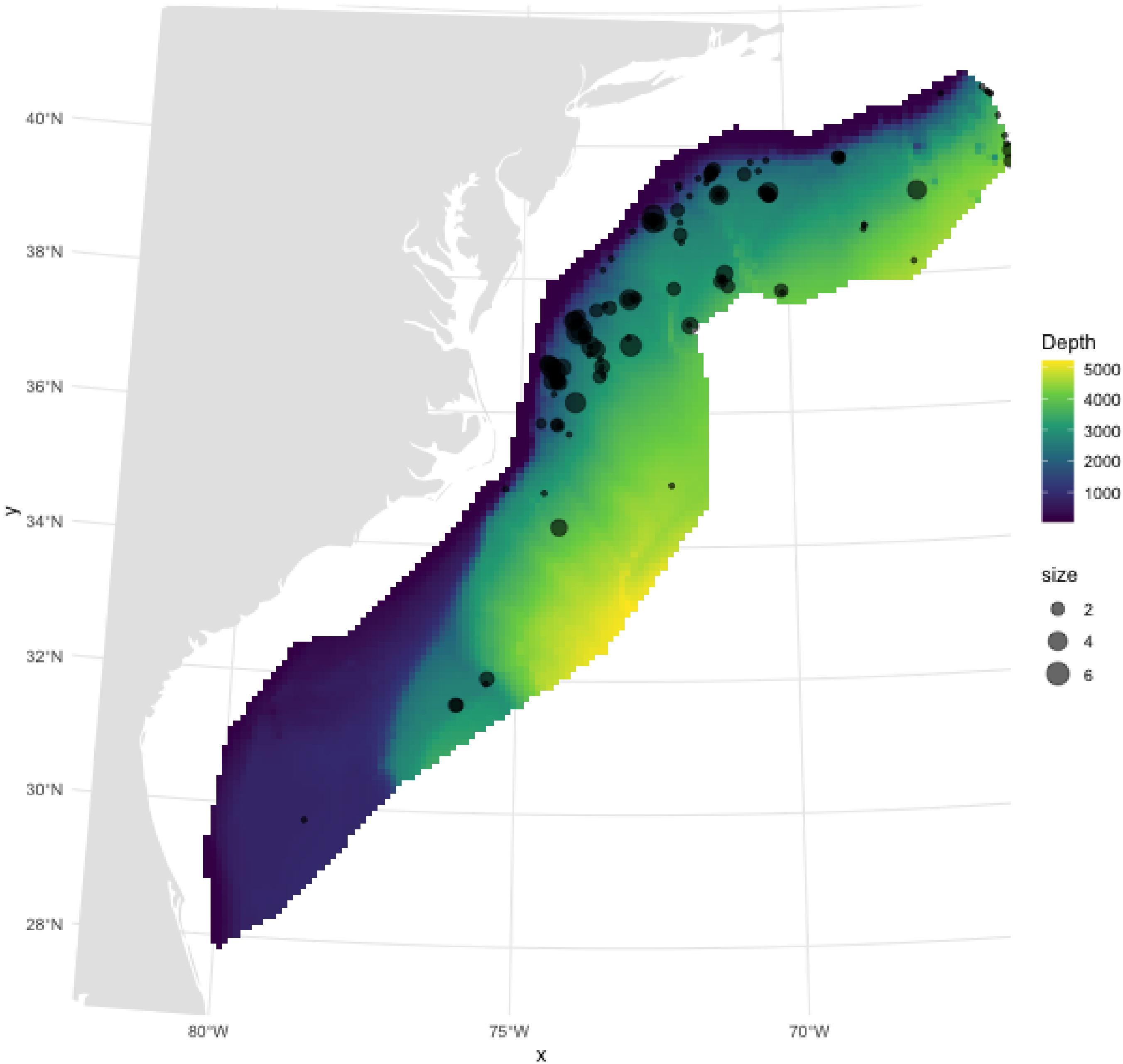
Randomisation & coverage probability

- H-T equation above assumes even coverage
  - (or you can estimate)

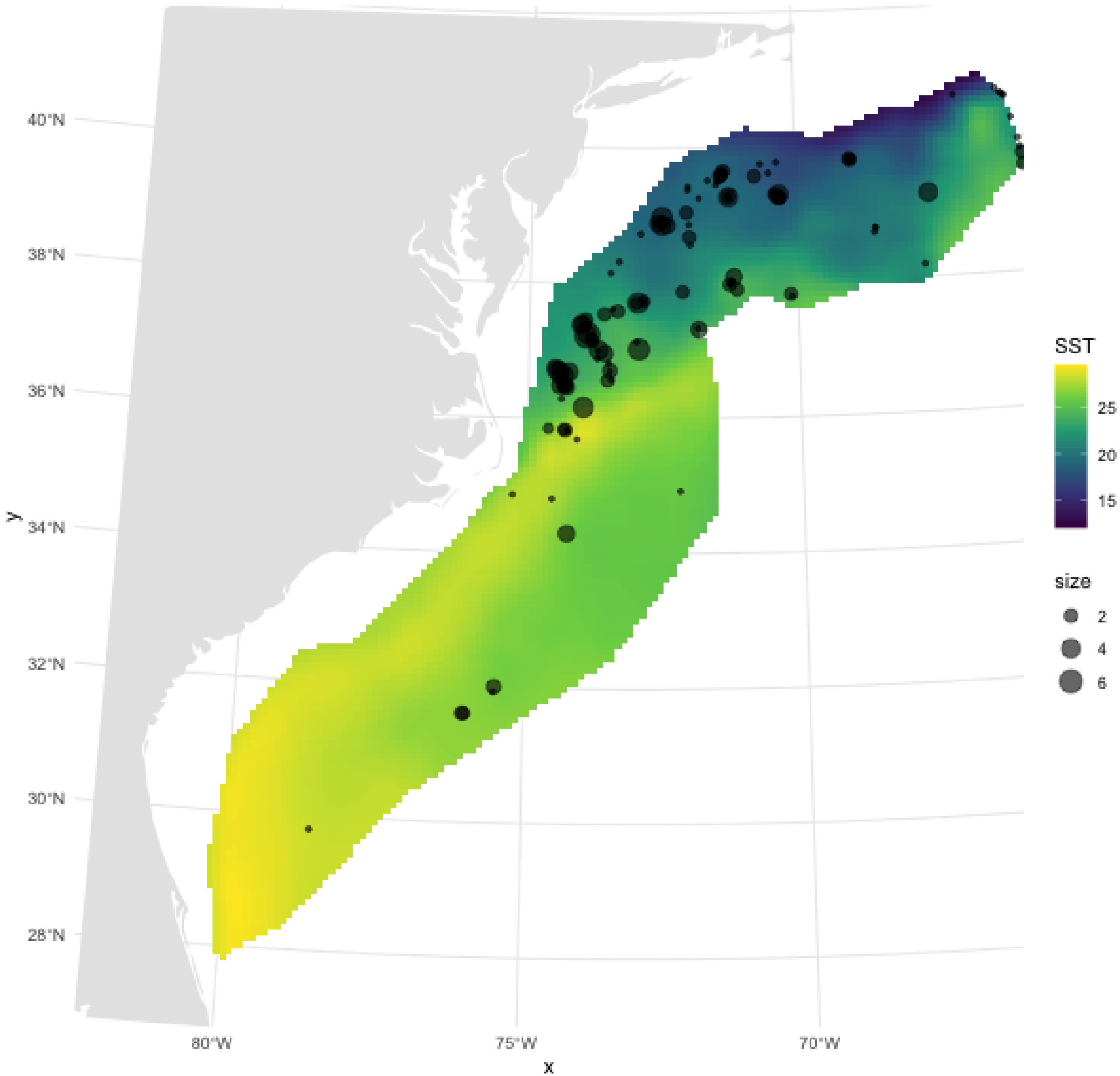
Extra information



Extra information - depth

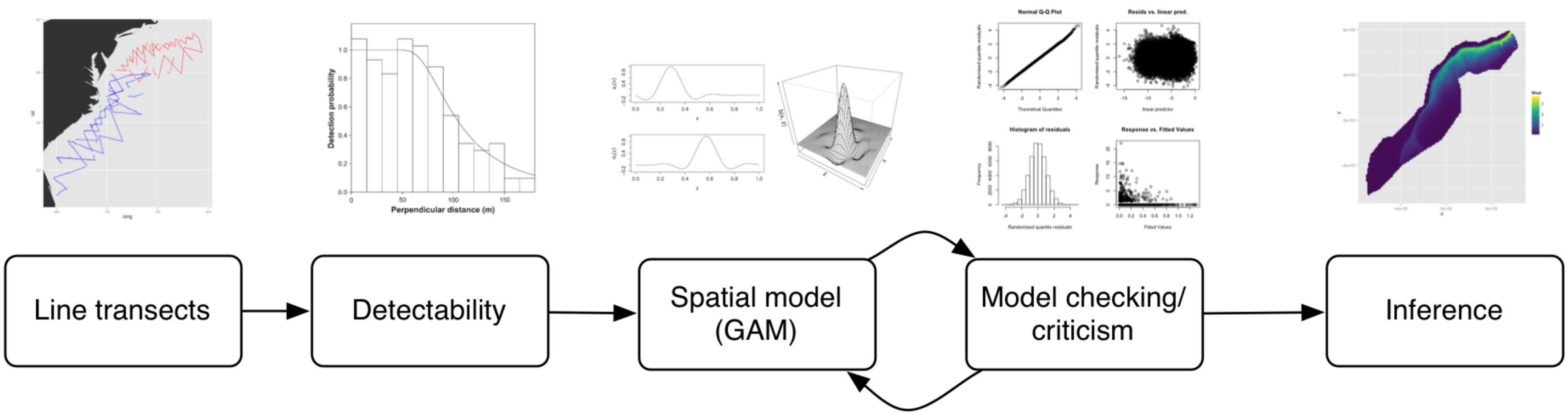


Extra information - SST



We should model that!

DSM flow diagram

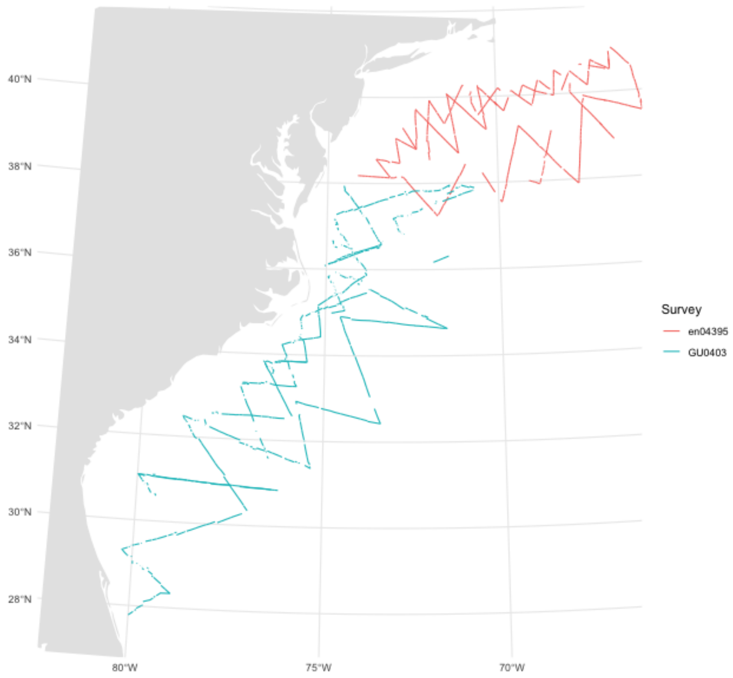


Modelling requirements

- Include detectability
- Account for effort
- Flexible/interpretable effects
- Predictions over an arbitrary area

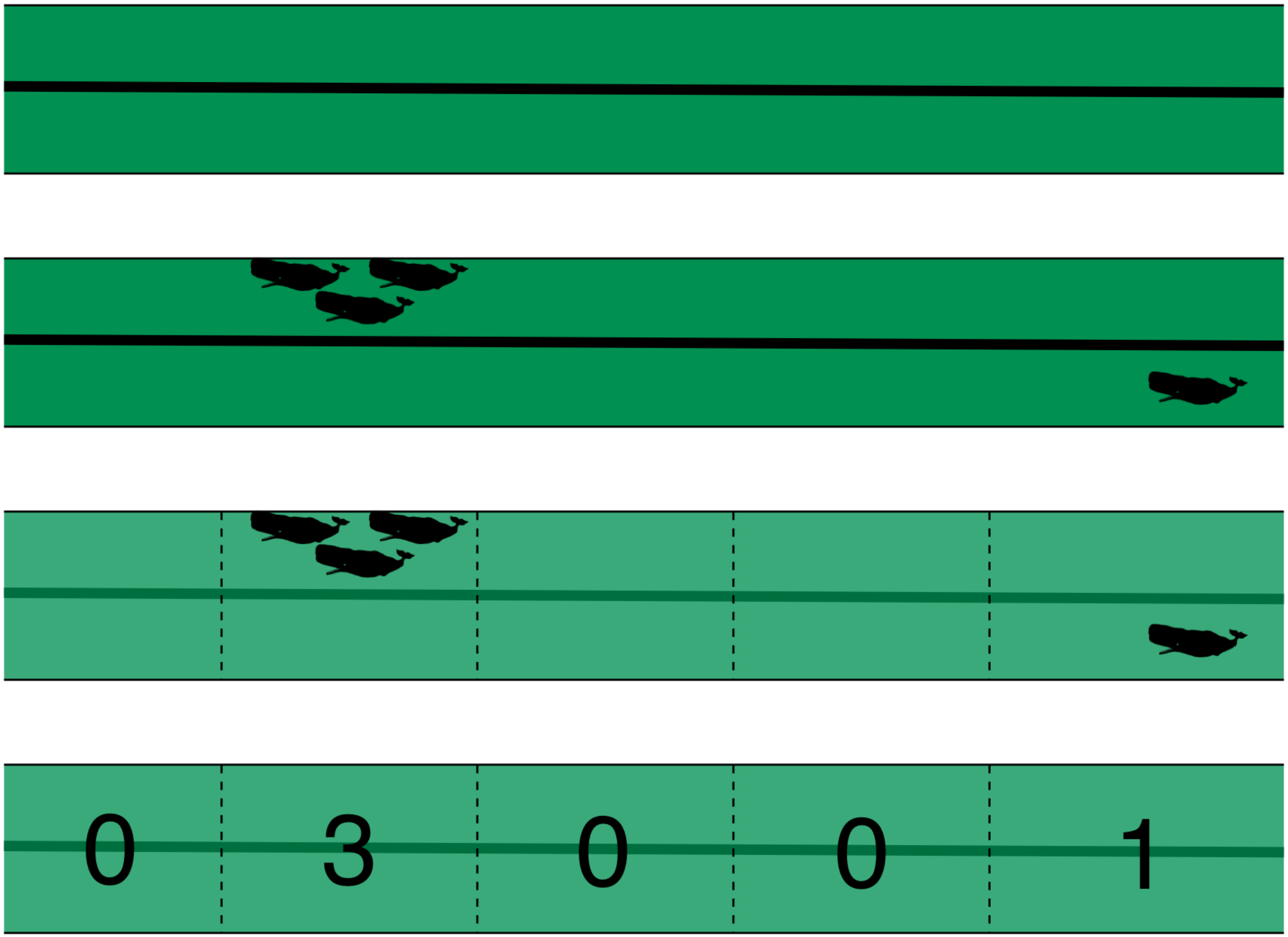
Accounting for effort

Effort



- Have transects
- Variation in counts and covars along them
- Want a sample unit w/ minimal variation
- "Segments": chunks of effort

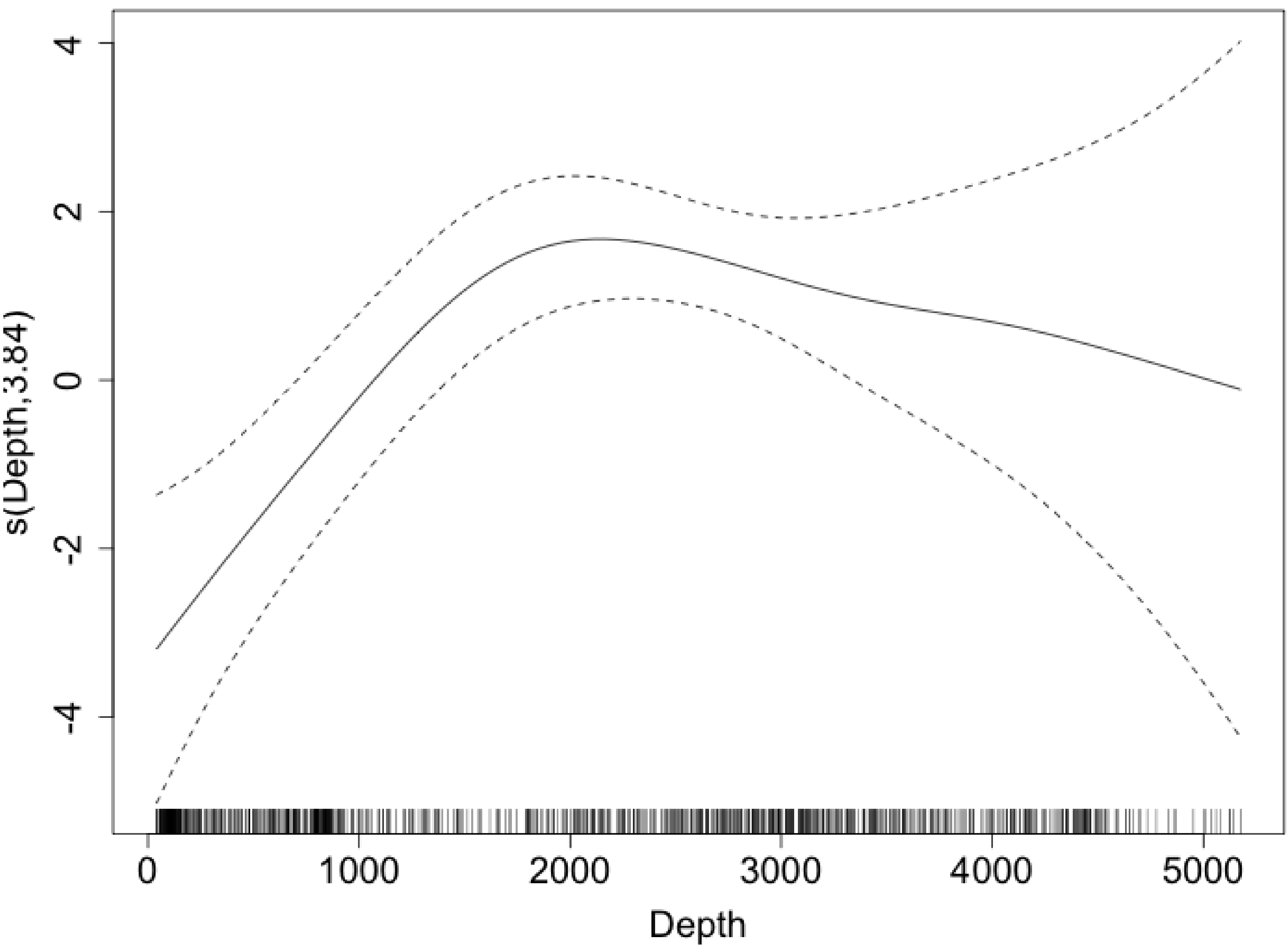
Chopping up transects



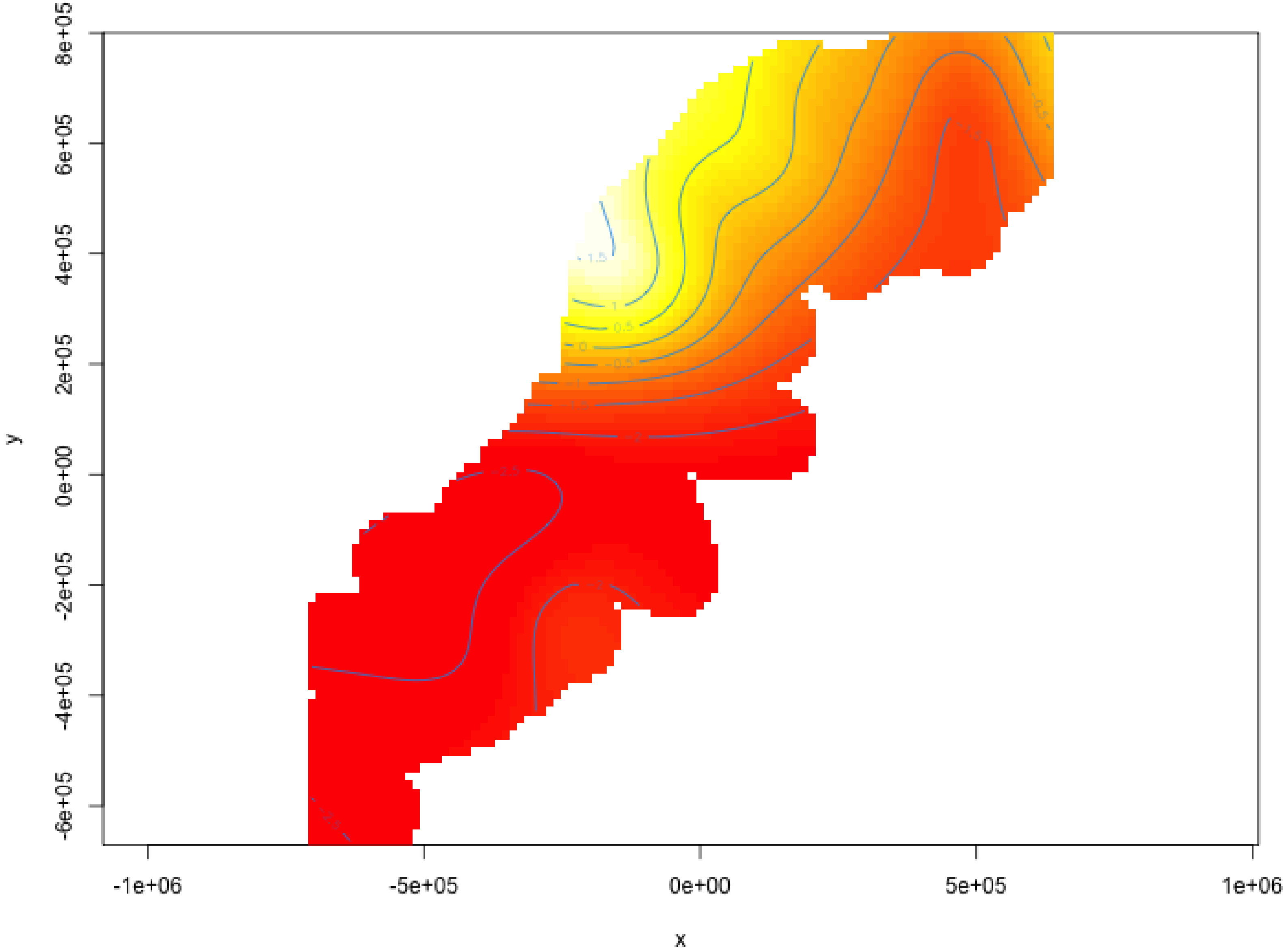
Physeter catodon by Noah Schlottman

Flexible, interpretable effects

Smooth response

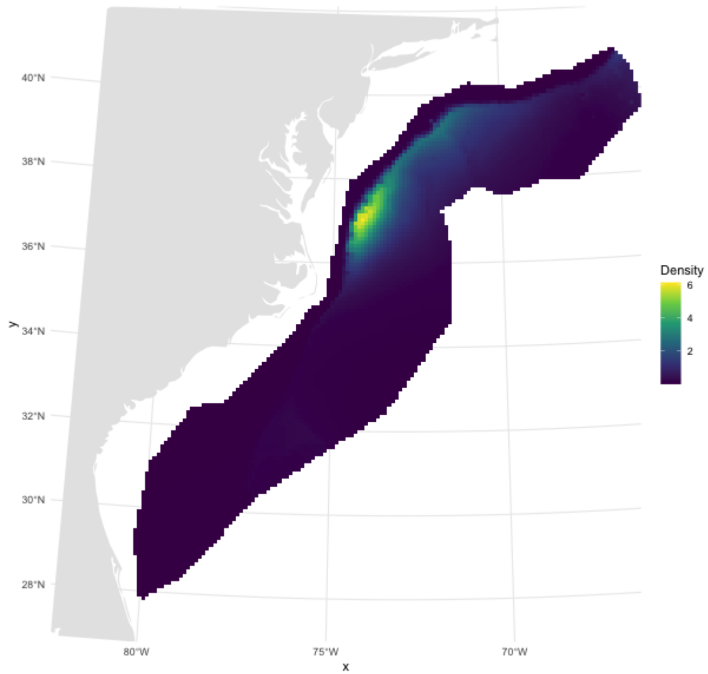


Explicit spatial effects



Predictions

Predictions over an arbitrary area



- Don't want to be restricted to predict on segments
- Predict within survey area
- Extrapolate outside (with caution)
- Working on a grid of cells

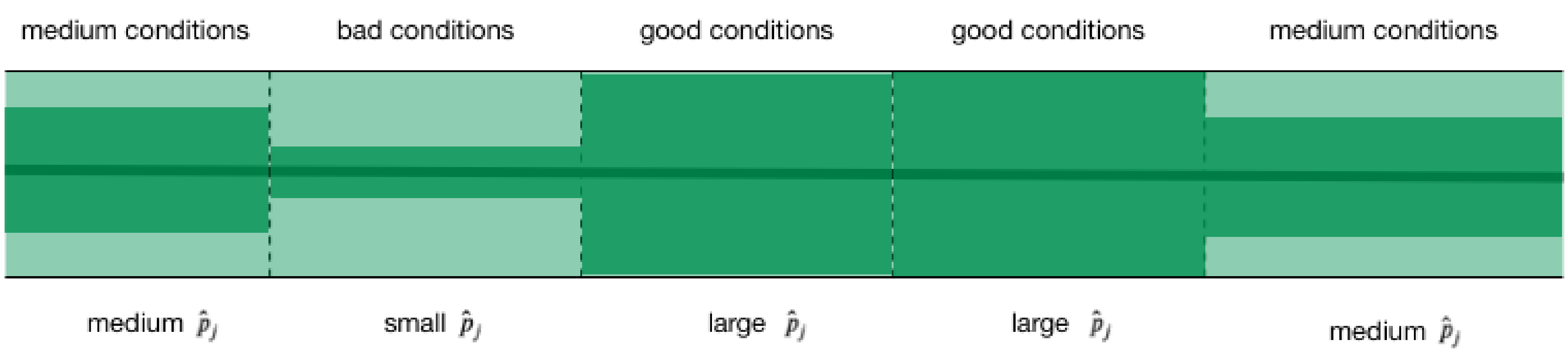
Detection information

Including detection information

- Two options:
  - adjust areas to account for **effective effort**
  - use **Horvitz-Thompson estimates** as response

Count model `count~...`

- Area of each segment, A_j
  - use $A_j\hat{p}_j$
- think effective strip width ($\hat{\mu} = w\hat{p}$)
- Response is counts per segment
- "Adjusting for effort"
- "Count model"



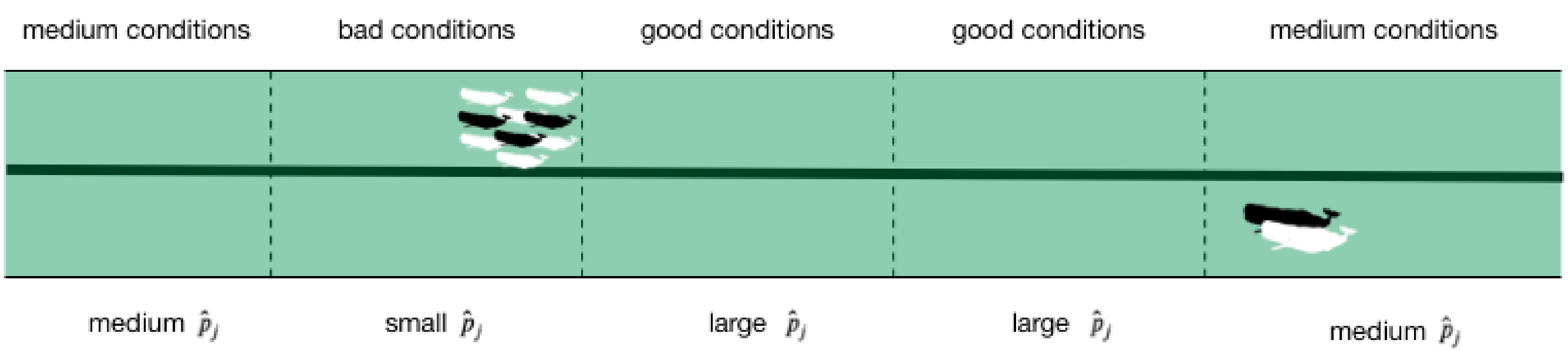
Estimated abundance

`abundance.est~...`

- Effort is area of each segment
- Estimate H-T abundance per segment

$$\hat{n}_j = \sum_i \frac{s_i}{\hat{p}_i}$$

(where the i observations are in segment j)



Detectability and covariates

- 2 covariate "levels" in detection function
  - "Observer"/"observation" -- change **within** segment
  - "Segment" -- change **between** segments
- "Count model" only lets us use segment-level covariates
- "Estimated abundance" lets us use either

When to use each approach?

- Generally "nicer" to adjust effort
- Keep response (counts) close to what was observed
- **Unless** you want observation-level covariates

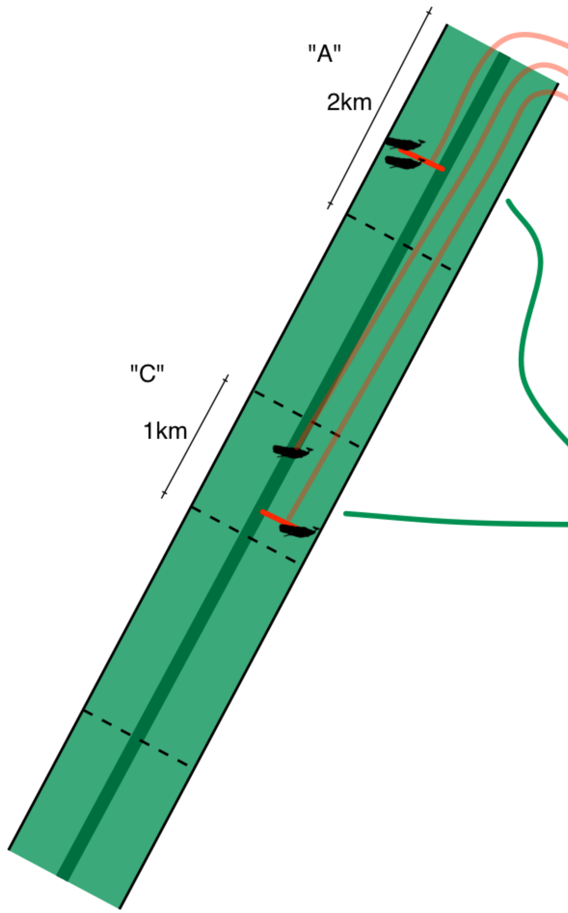
Data requirements

What do we need?

- Need to "link" data
    - ✅ Distance data/detection function
    - ✅ Segment data
    - ✅ Observation data (segments 🔗 detections)

More info on course website.

Density surface model data setup for package `dsm`



Detection data

1 line per detection

object	distance	size	observer	...
1	0.5	2	1	...
...	...	...	...	...
101	0	1	1	...
102	0.45	1	1	...
...	...	...	...	...

Included in the detection function, additional columns for detection covariates

Segment table

1 line per segment

Sample.Label	Effort	...
A	2	...
...	...	...
C	1	...
...	...	...

Effort data, extra columns for environmental covariates (space, depth, altitude etc)

Observation table

1 line per detection

object	Sample.Label	size	distance	...
1	A	2	0.5	...
...	...	...	...	...
101	C	1	0	...
102	C	1	0.45	...
...	...	...	...	...

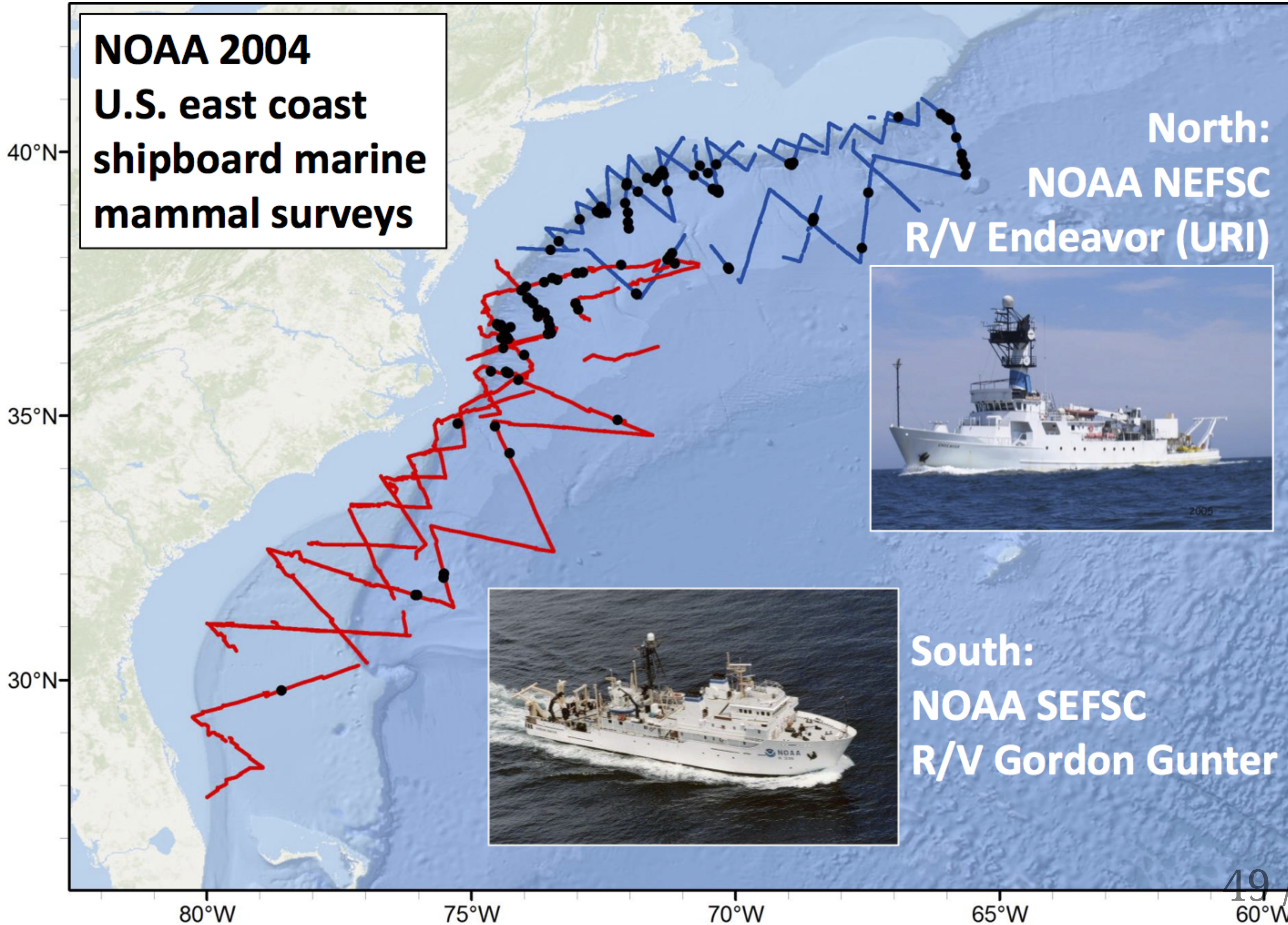
Links detections to effort data

See `?"dsm-data"` for more information on these tables.

Prepared by David L Miller, University of St Andrews

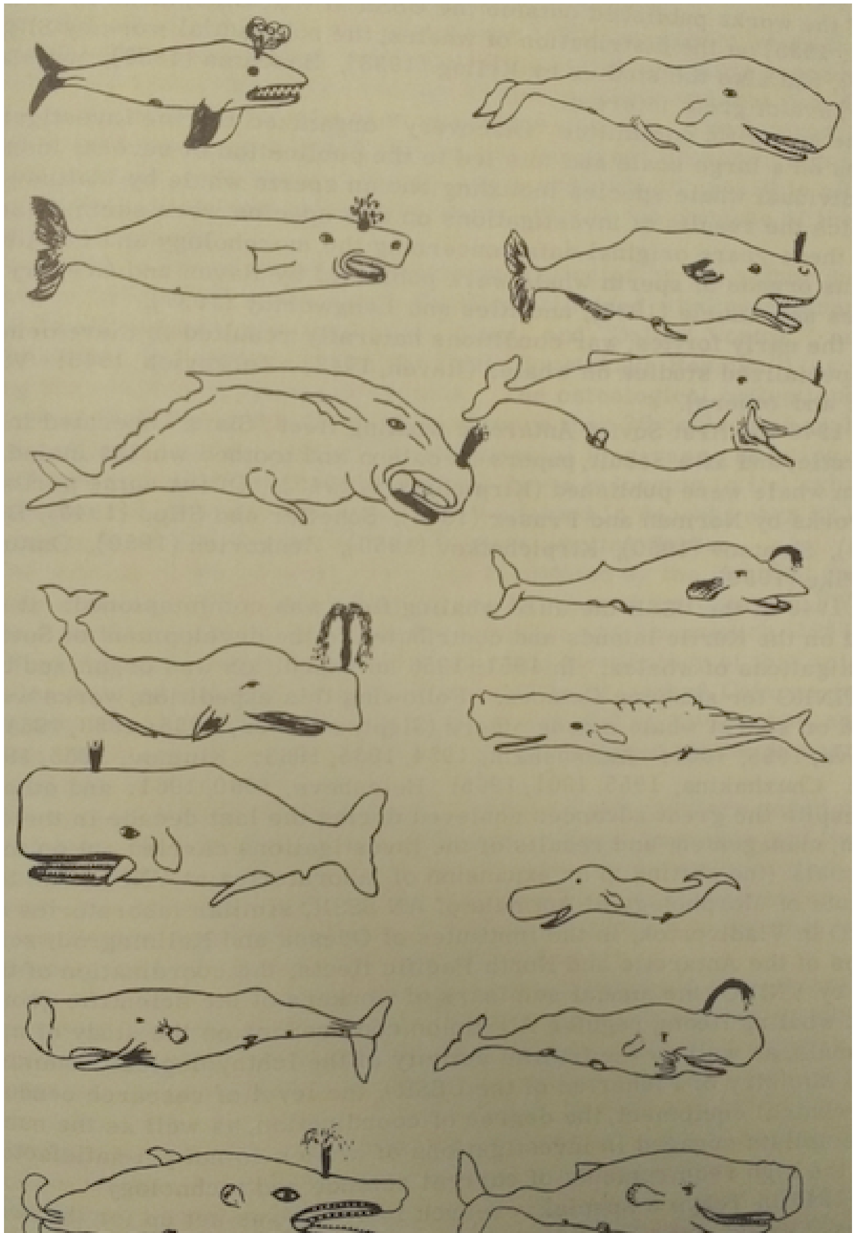
Example data

Example data



Example data

Sperm whales



- Hang out near canyons, eat squid
- Surveys in 2004, US east coast
- Thanks to Debi Palka (NOAA NEFSC), Lance Garrison (NOAA SEFSC) for data. Jason Roberts (Duke University) for data prep.

Recap

- Model counts or estimated abundance
- The effort is accounted for differently
- Flexible models are good
- Incorporate detectability
- 2 tables + detection function needed