Distance sampling primer

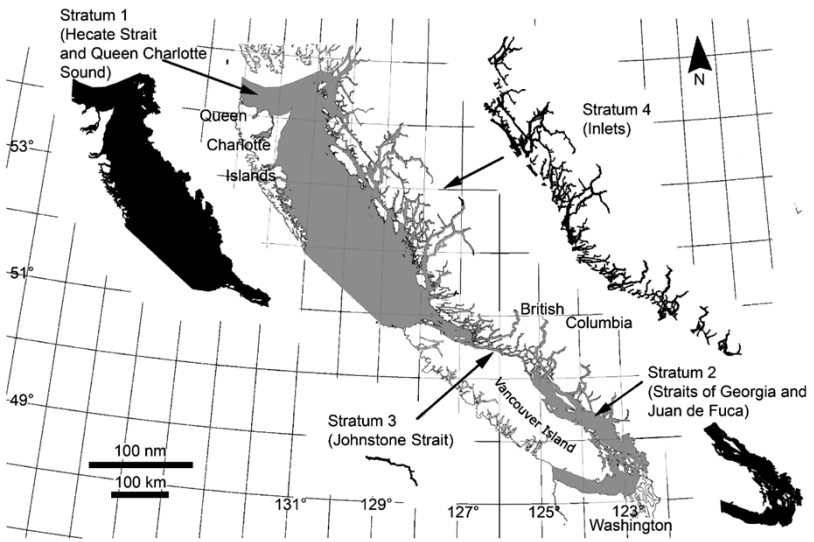
Len Thomas

How many?

- Goals:
  - Estimate number of marine mammals (“abundance”) of a given species in a defined study region in some defined time period
  - Estimate density (abundance per unit area)
  - Quantify uncertainty in our estimates

- Additional possible goals:
  - Multi-species survey
  - Density surface model (“map”)
  - Temporal trends

davidsuzuki.org

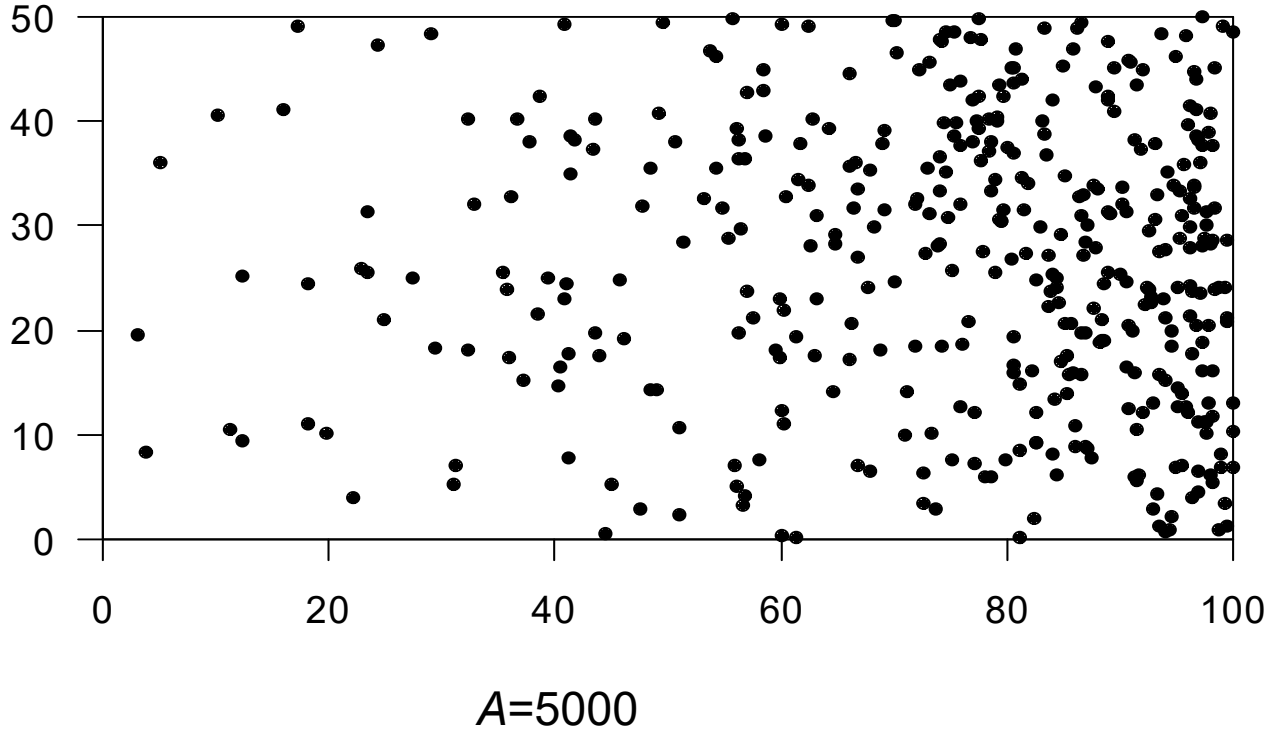


Complete census

- Method: count everything!

  N = 412

  D = 412/5000 = 0.0824



Plot sampling (strip transect)

- Let

  k = number of strips = 5

  L = total line length = 50x5 = 250

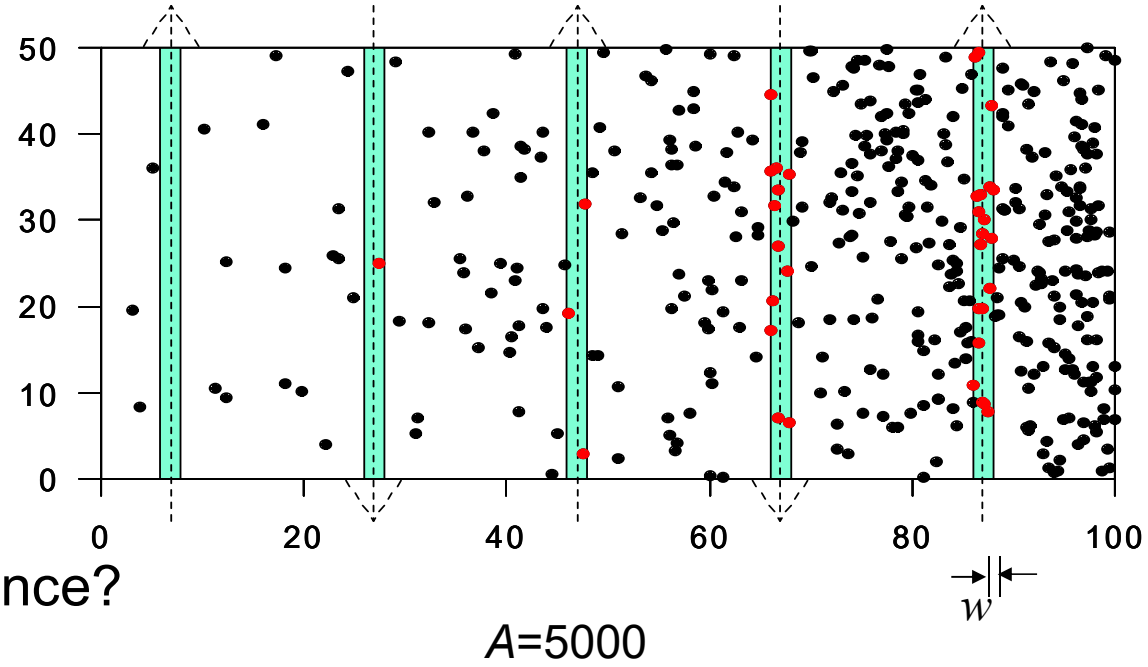
  w = the strip half-width = 1

  a = area of region covered

  = $2wL$ = 2x1x250 = 500

  n = number of animals counted = 36

- From this, how do we estimate abundance?



Intuitive estimator of abundance

- I saw 36 animals
- I covered 500/5000 = 1/10th of the study region
- So, I estimate there are 36/(1/10) = 36x10 = 360 animals

$$\hat{N} = \frac{n}{a/A} = \frac{nA}{a} = \frac{36 \times 5000}{500} = 360$$

Concept – Plot sampling

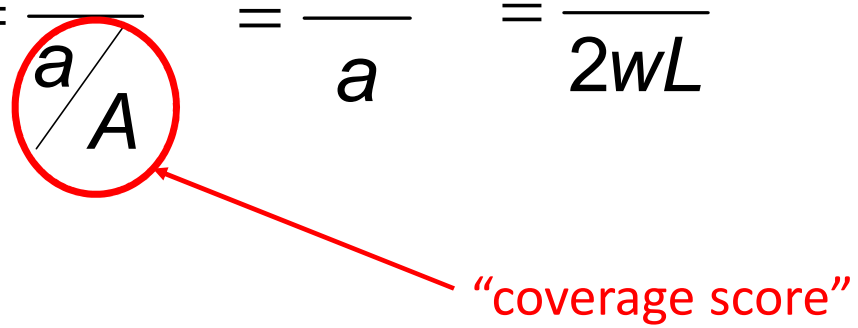
- Step 1: How many in <u>covered</u> region, N_a?

Plot sampling: $N_a = n$

- Step 2: Given N_a, how many in <u>study</u> region, N

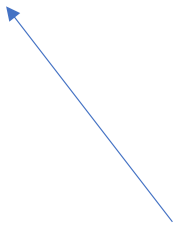
If transects placed at random: $\hat{N} = \frac{N_a}{a/A}$

- Overall: $\hat{N} = \frac{n}{a/A} = \frac{nA}{a} = \frac{nA}{2wL}$



Assumptions – plot sampling

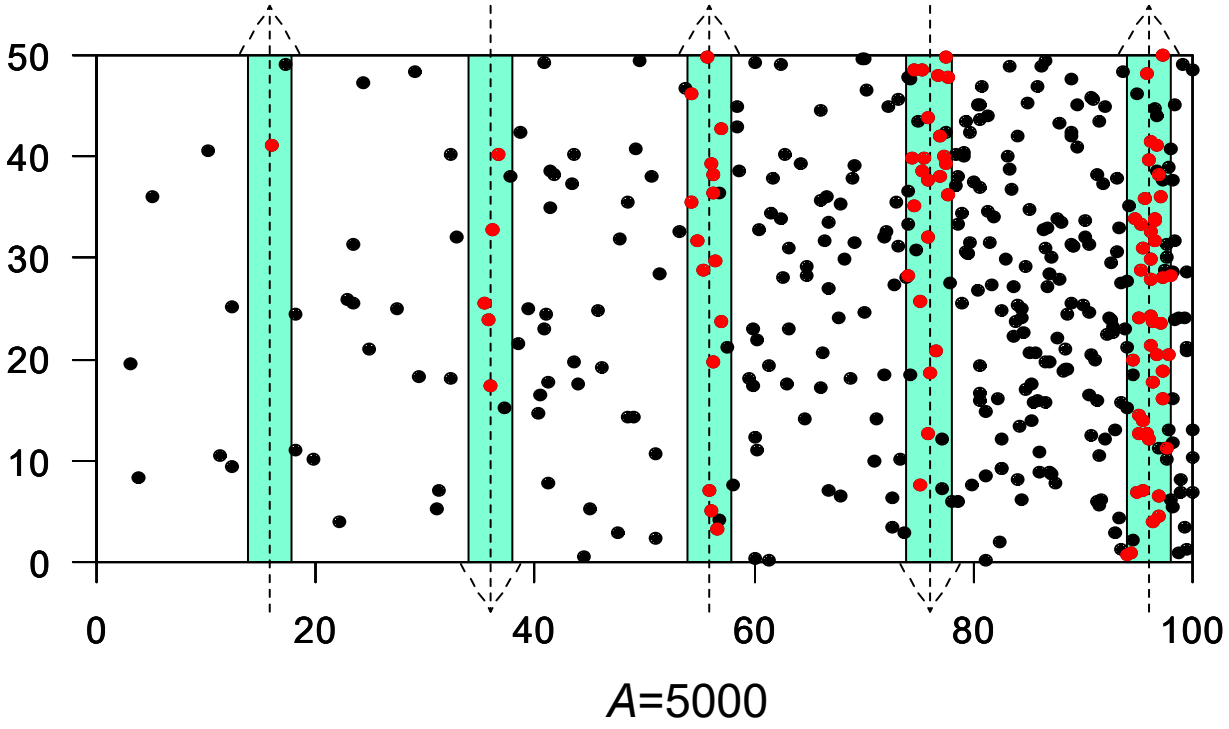
- Step 1: How many in <u>covered</u> region, N_a?
  - Detect everything in the transects
- Step 2: Given N_a, how many in <u>study</u> region, N
  - Transects are located randomly



This assumption is about the survey design.
So, we call this type of method "design-based".

Distance (line transect) sampling

- Extend plot sampling so don't have to assume all animals in the covered region are detected
- Here
  - w = 2 (strip can be wider, as don't have to see everything)
  - a = 1000
  - n = 68 (more animals seen)
- Let
  - P_a = proportion of animals detected within covered region
- Imagine we know (or can estimate) $\hat{P}_a = 0.7$



Intuitive estimator of abundance

- I saw 68 animals
- The estimated proportion seen was 0.7
- So, I estimate the true number of animals in the strips was 68/0.7 = 97.1
- I covered 1000/5000 = 1/5th of the study region
- So, I estimate there are 97.1/(1/5) = 485.7 animals

$$\widehat{N} = \frac{n/\widehat{P}_a}{a/A} = \frac{n}{a/A\,\widehat{P}_a} = \frac{68}{(1000/5000) \times 0.7} = 485.7$$

Concept – Distance sampling

- Step 1: How many in <u>covered</u> region, N_a?

  Distance sampling: $\hat{N}_a = n / \hat{P}_a$

- Step 2: Given N_a, how many in <u>study</u> region, N

  If transects placed at random: $\hat{N} = \frac{\hat{N}_a}{a/A}$

- Overall: $\widehat{N} = \frac{n}{a/A\,\hat{P}_a} = \frac{nA}{a\hat{P}_a} = \frac{nA}{2wL\hat{P}_a}$

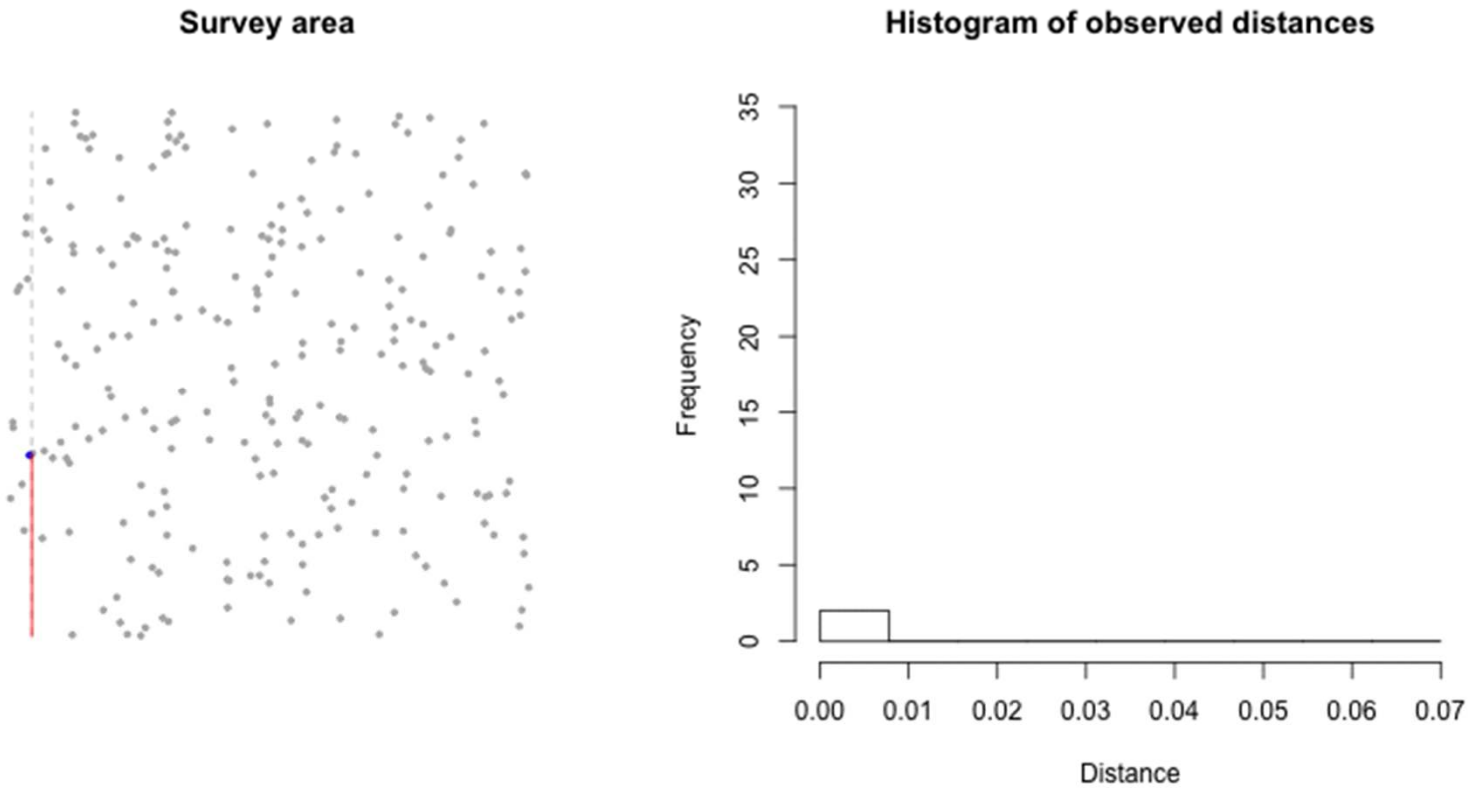
  "detection probability"

  "coverage score"

- So how do we estimate P_a?

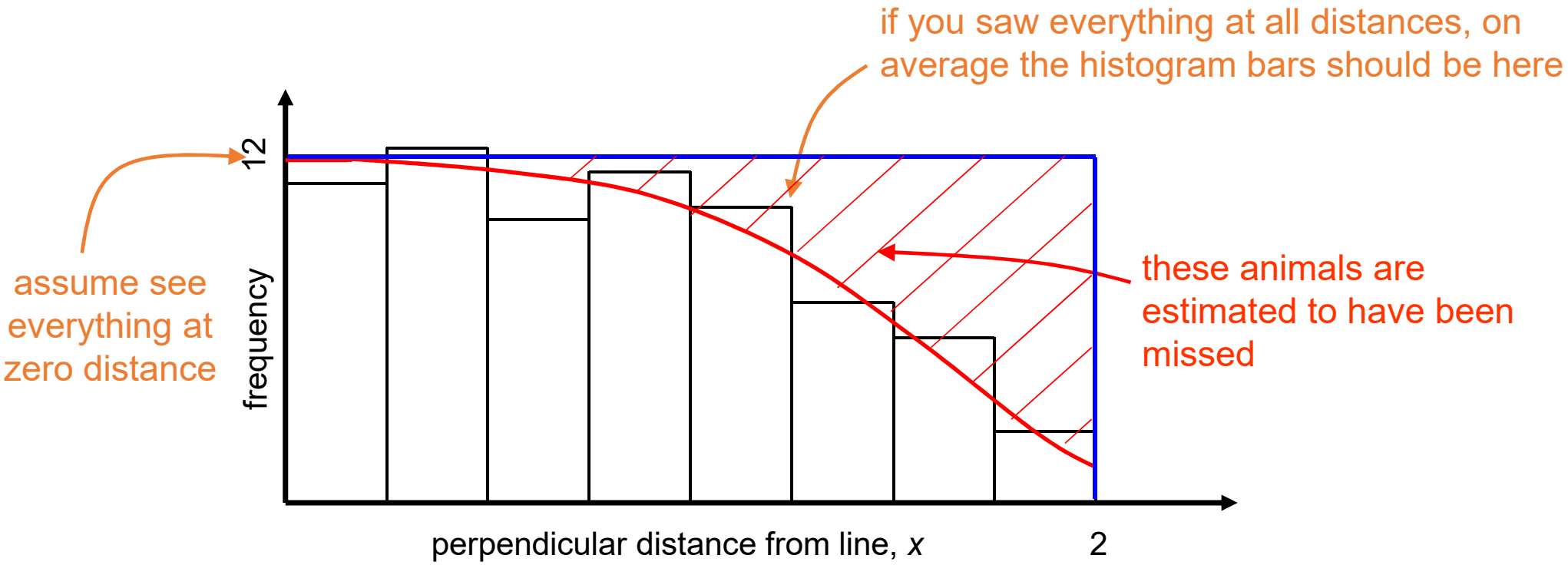
Estimating P_a

Obtain perpendicular distance, *x*, from transect line to each observed object

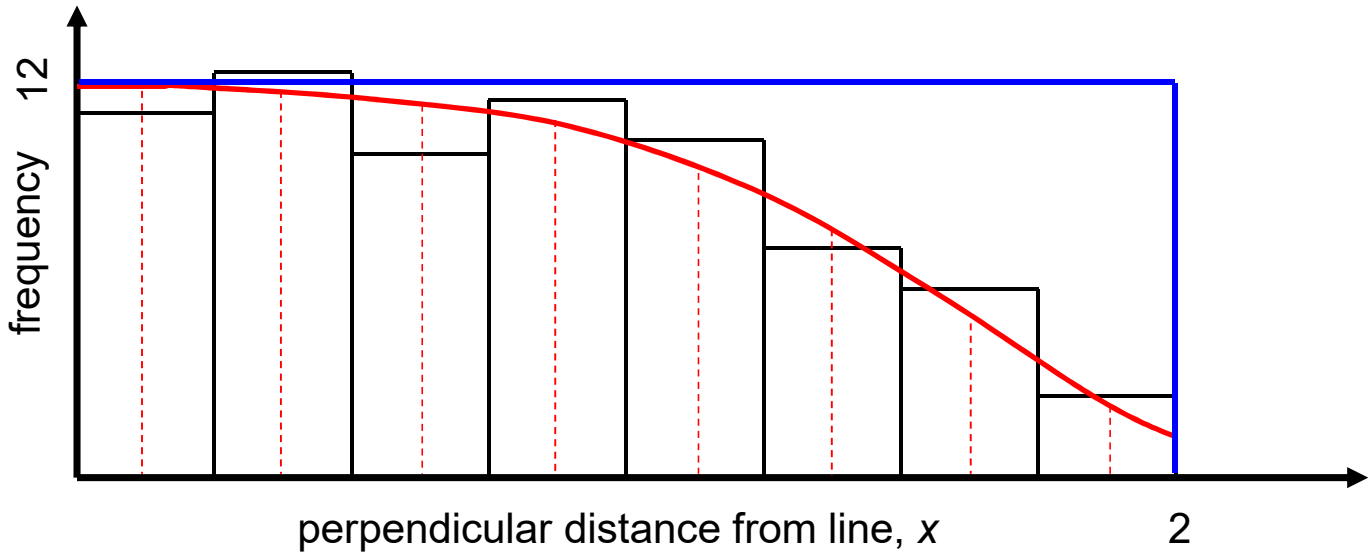


Can record perpendicular distance, or
radial distance and angle from trackline to detection

Estimating P_a



Estimating P_a



$$\hat{P}_a = \frac{\text{area under curve}}{\text{area under rectangle}}$$

- Area of rectangle = 12x2 = 24
- Area under curve = 0.25x(12+11.5+11+10.5+9+7+4+3) = 17
- So

$$\hat{P}_a = \frac{17}{24} = 0.7$$

Assumptions – distance sampling (line transect)

- Step 1: How many in covered region, N_a?
  - Detect everything at zero distance
  - Animals uniformly distributed within survey strips
    - Assured if you place lines randomly
  - Survey process is a snapshot (or animals don't move)
  - Distances are measured accurately
  - (Detections are independent)
- Step 2: Given N_a, how many in study region, N
  - Transects are located randomly

Model of animal distribution

Also have a model of the detection process

So this method is "model-based"

Design-based, as with plot sampling

So, overall, a mix of design- and model-based

Quantifying uncertainty

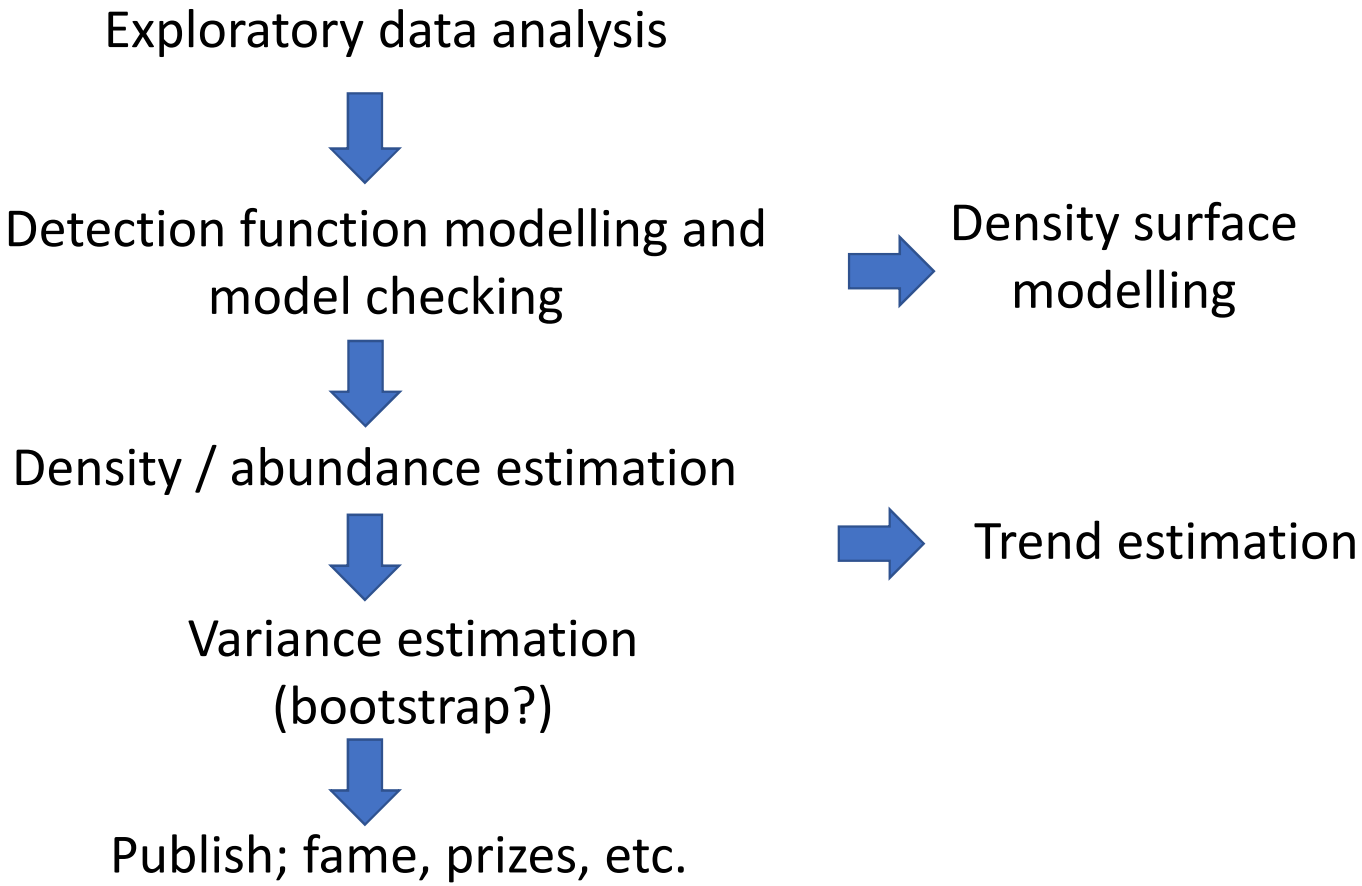
- Important!
- Where does the uncertainty come from?

Gui Bortolotto

- Detection function – model-based
- Encounter rate – design-based
  - Variation in counts between lines

- Independent, so can easily combine to produce an overall variance estimate (“delta method”)
- Alternative: nonparametric bootstrap (resample lines)

Conventional distance sampling (CDS) analysis workflow

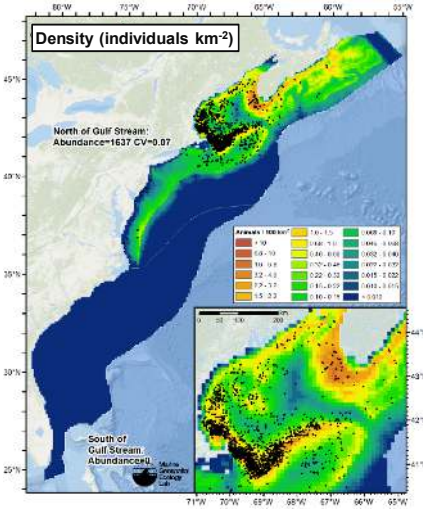


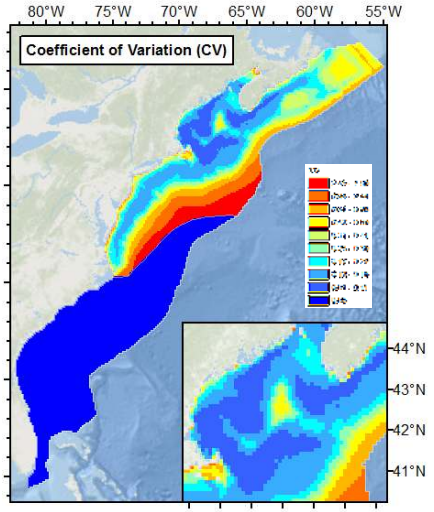
Introducing subsequent talks

- Good design is key! – Cornelia
- Key assumption – perfect detection on the trackline.
  Commonly violated for marine mammals. What to do? – David
- Neglected assumption – no movement.
  Commonly violated for marine mammals. What do do? – Richard
- Heard but not seen – passive acoustics. – Tiago
- Replacing design assumptions with model – density surface model – Jason.

Density surface models – pros and cos

- Pros
  - Don’t need a design! (but can you still do detection function estimation?)
  - Can explain variation between lines – so potentially smaller variance
  - Can make better inferences into small areas “borrow strength across space and time”
  - Can aid understanding of factors driving variation in marine mammal density
    - Warning – correlation not causation!
  - You get a map!
- Cons
  - Less robust than design-based methods – all models are wrong!
  - Results can be sensitive to model used
  - Some methods not accessible to most users

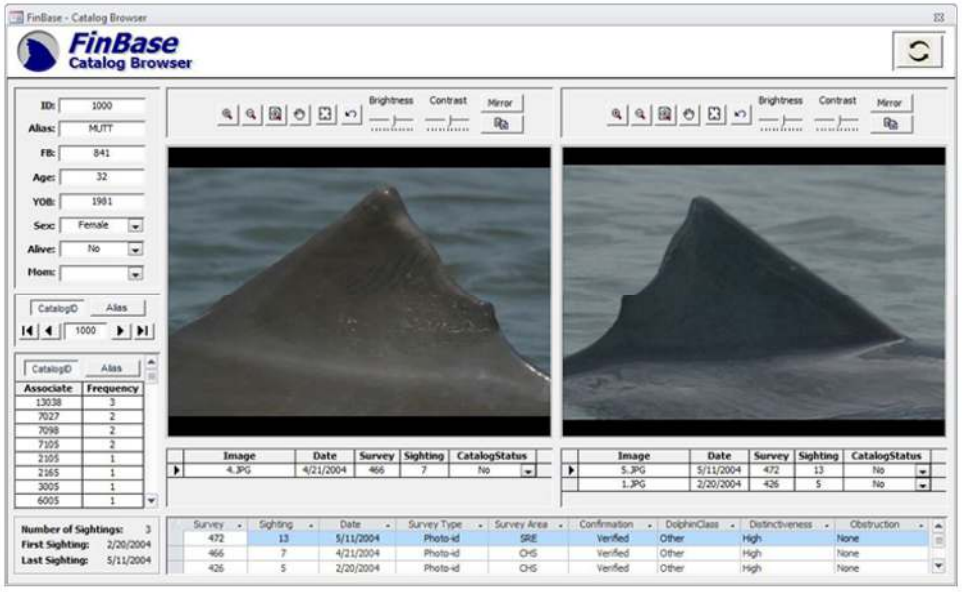




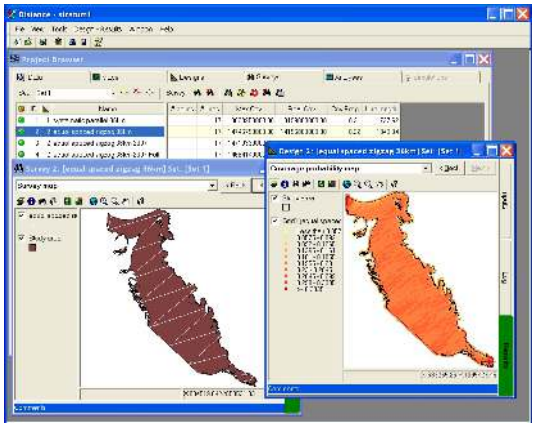
Alternative to distance sampling

- Mark recapture – where animals can be uniquely identified
  - Photo-ID
  - Spatial capture-recapture
  - Close-kin methods

Photos: Keith Mullin / NOAA



Where can I find out more?



- Clearinghouse for info: http://distancesampling.org
- Free online workshop: https://workshops.distancesampling.org/
- Interactive online and face-to-face workshops
- Text books; bibliography of papers at distancesampling.org site
- Google group: distance-sampling
- Free software:
  - Distance for Windows – distancesampling.org
  - Distance sampling packages in R – distancesampling.org
    - Development versions at: https://github.com/DistanceDevelopment/
    - To install them, first install the devtools package, then `devtools::install_github("DistanceDevelopment/[package]", build_vignettes = TRUE)`
    - E.g., `devtools::install_github("DistanceDevelopment/Distance", build_vignettes = TRUE)`