Distance 5: Mark-Recapture Distance Sampling

Photo: Ullas Karanth

Photo: Steve Dawson

Photo: Thomas Norris

Photo: Rich Guenzel

Jeff Laake[1], David Borchers[2], Len Thomas[2]

[1]NMFS, Seattle. [2]CREEM, Univ. St. Andrews.

EURING Software Session
Jan 14th 2007

Outline

- Introduction – Conventional distance sampling
- Mark-recapture distance sampling (MRDS)
- Software
  - Distance
  - MRDS in Distance
  - R interface
- Example – Crabeater seals
- Other things Distance can (or will) do

References

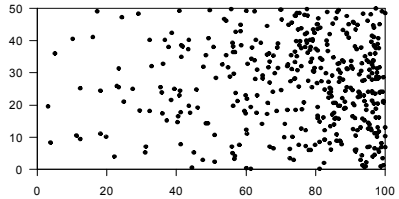
- Laake, J.L. and Borchers, D.L. 2004. Methods for incomplete detection at distance zero. In Advanced Distance Sampling, S. Buckland, D. Anderson, K. Burnham, J. Laake, D. Borchers, and L. Thomas (eds), 108–189. Oxford: Oxford University Press.
- Borchers, D. L., J. L. Laake, C. Southwell, and C. G. M. Paxton. 2006. Accommodating Unmodeled Heterogeneity in Double-Observer Distance Sampling Surveys. Biometrics 62: 372-378.
- Thomas, L., Laake, J.L., Strindberg, S., Marques, F.F.C., Buckland, S.T., Borchers, D.L., Anderson, D.R., Burnham, K.P., Hedley, S.L., Pollard, J.H., Bishop, J.R.B. and Marques, T.A. 2006. Distance 5.0. Release 2. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/

Introduction

Conventional Distance Sampling

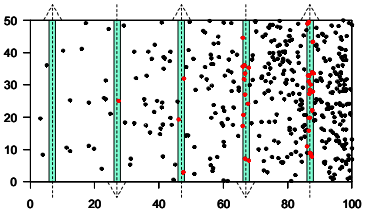
Goal: Estimate size of a closed population

- E.g.:



- Complete count: Population size, N = 412

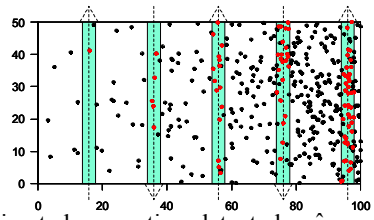
Plot sampling – e.g. strip transect



$$\text{Estimated population size} = \frac{\text{number in strips}}{\text{proportion of area surveyed}}$$

$$\hat{N} = \frac{n}{p_c}$$

Distance sampling – line transect



Estimated proportion detected $= \hat{p}_d$

Estimated number in strips, $\hat{N}_c = \dfrac{n}{\hat{p}}$

Estimated population size, $\hat{N} = \dfrac{\hat{N}_c}{p_c} = \dfrac{n}{p_c \hat{p}_d}$

Aside: General Horvitz-Thompson-like estimator

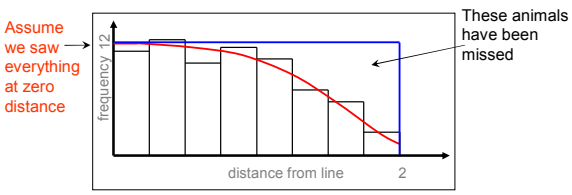
- If coverage probability and/or detection probability vary by individual, can use

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{p_{c,i}\hat{p}_{d,i}}$$

- In conventional distance sampling we (mostly) don't need this, because we have
  - uniform coverage probability by design
  - pooling robustness

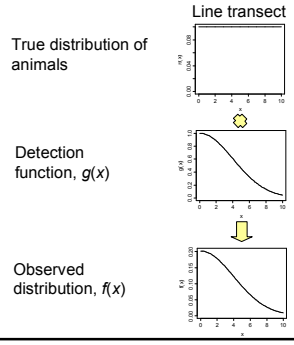
How do we estimate p_d?

- Record distance from line to each detected object



$$\hat{p}_d = \frac{\text{area under curve}}{\text{area of rectangle}}$$

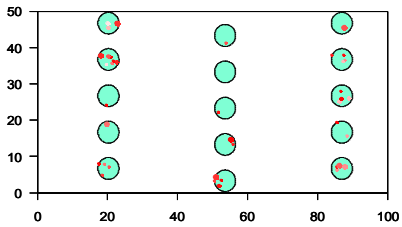
$g(x)$ and $f(x)$



- Specify parametric form for g(x)
- Fit f(x) to observed data
- Estimate p_d

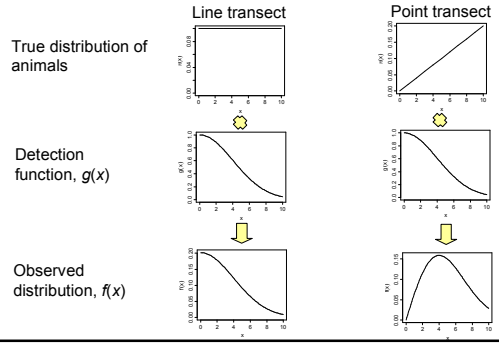
$$\hat{p}_d = \frac{1}{w\hat{f}(0)}$$

Point transects



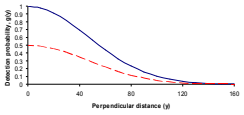
Measure radial distance from point to object

Estimate p_d in a similar way



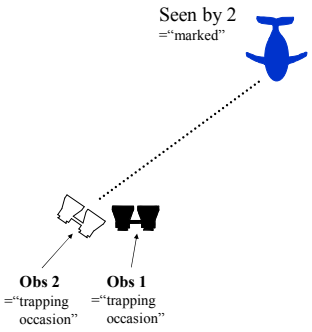
Key assumption: g(0)=1

- Key assumption is that we detect all animals at 0 distance (on the line or at the point)
- Violation causes negative bias – e.g., if g(0)=0.8, estimates of N are 80% of true N
- Nothing in the distance data to enable us to tell if g(0)=1
- Need additional data…

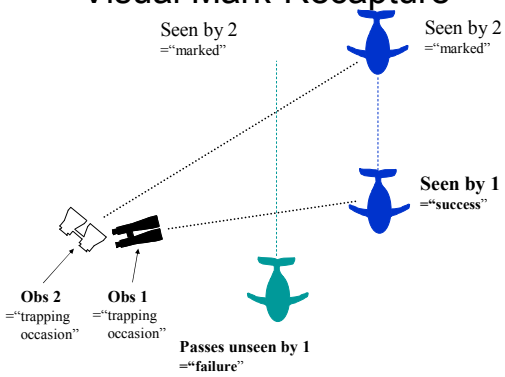


Mark-recapture distance sampling

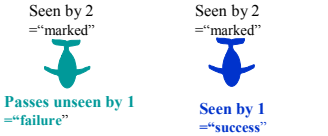
Visual Mark-Recapture



Visual Mark-Recapture



Visual Mark-Recapture



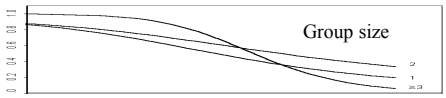
- We know 2 animals passed (because Obs 2 saw them)
- Of these, Obs 1 saw 1
- So **estimate:**
  Pr(Obs 1 sees) = $\hat{p}_1 = \frac{1}{2} = \frac{n_{12}}{n_2}$ = $\frac{\text{number "duplicates"}}{\text{number seen by 2}}$

Heterogeneity

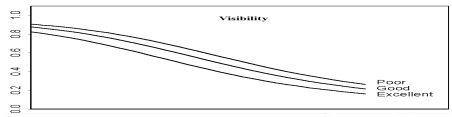
- **BUT** we know that heterogeneity between animals in *p* causes positive bias in estimated *p*, and hence negative bias in estimated *N*
- So, include covariates where we have them…

Potential kinds of covariate

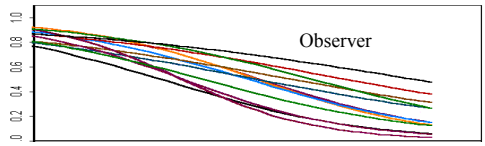
- The **animals** themselves (distance, size ...)



- The **environment** (sea state, ground cover, ...)



- The kind of **survey effort** (the observers, their platforms, ...)



Conditional Detection Function estimated by binary regression

- can include any observable explanatory variables in the detection function;
- can easily be done with standard statistical packages;
- provides model selection tools

Configuration: Trial-Observer

Observer 2
↓
sets up trials for
↓
Observer 1
↓
to estimate p_1

The Observer at the end of an arrow must be independent of the Observer at the start of the arrow

Configuration: Independent Observer

to estimate p_2
↑
Observer 2
↓
sets up trials for
↑
↓
Observer 1
↑
↓
to estimate p_1

$$p_{\cdot} = p_1 + p_2 - (p_1 p_2)$$

The Observer at the end of an arrow must be independent of the Observer at the start of the arrow

Abundance Estimation

- Trial-Observer

$$\hat{N}_c = \sum_{\text{all } i \text{ seen by } 1} \frac{1}{\hat{p}_1(x_i, \cdots)}$$

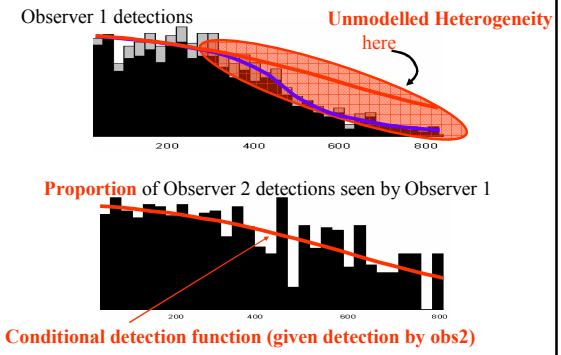
(note)

- Independent Observer

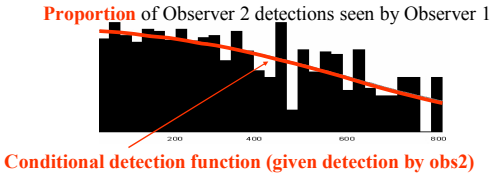
$$\hat{N}_c = \sum_{\text{all } i \text{ seen}} \frac{1}{\hat{p}_{\cdot}(x_i, \cdots)}$$

(note)

Is there still a Problem?

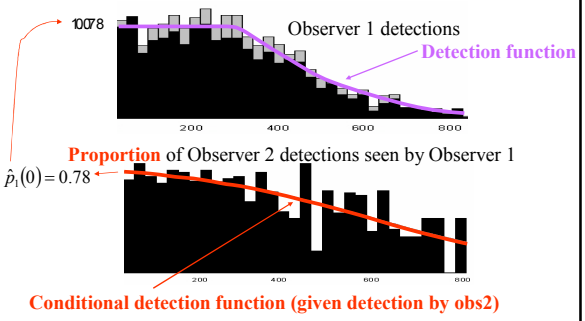


"Full Independence" (FI) Model



"Point Independence" (PI) Model

Independence at 0 only

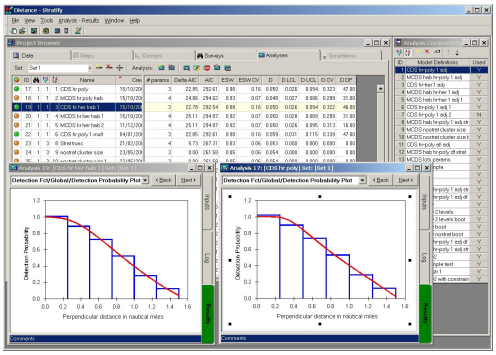


Critical Assumptions of Mark-Recapture Distance Sampling

- Have the required independence between observers
- No unmodelled heterogeneity
  - anywhere (full independence)
  - at some specified distance (point independence)
  - field methods can help – e.g., visual acoustic
- Duplicates (resightings) known
  - Methods exist to allow for duplicate uncertainty
- Animals distribution wrt distance is known (point independence)

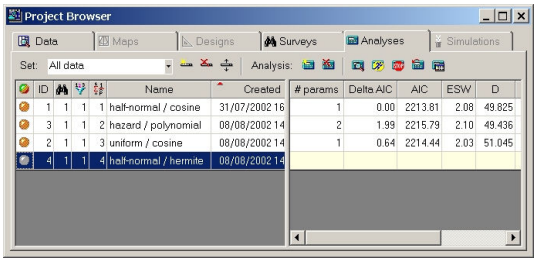
Software

Distance

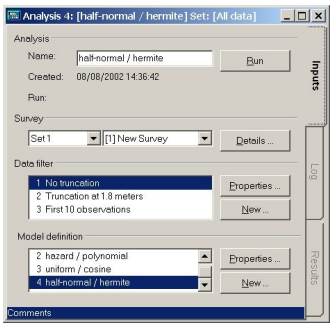


Analysis Browser

(Analysis tab of project browser window)



Analysis Details



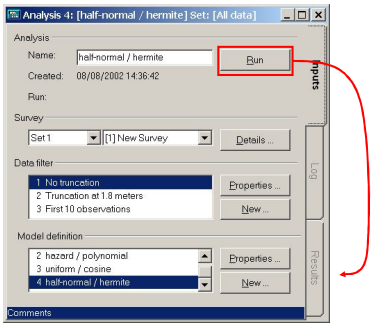
Components of an Analysis

1. Survey
2. Data Filter
3. Model Definition

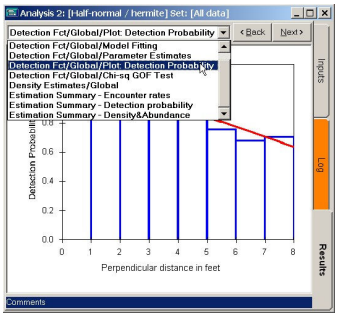
Data

Results

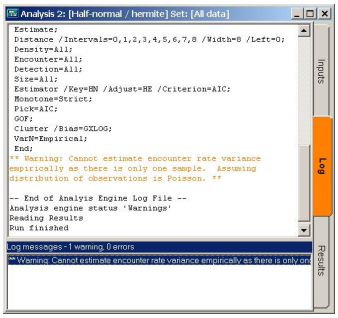
Running an analysis



Viewing results



Looking in the Log for errors and warnings

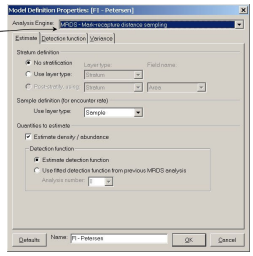


MRDS Analysis Engine

- Main capabilities:
  - Line transects (point transects to follow…)
  - Independent observer and trial configurations (removal is another possibility, and is to follow…)
  - Full and point independence
  - Distance data as “exact” measurements or grouped into intervals
- Implemented as an R library
  - Called behind-the-scenes from the Distance GUI

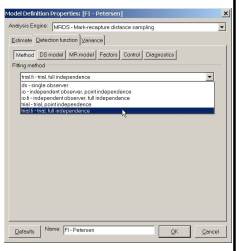
MRDS analyses

- Select MRDS engine in Model Definition
- Options presented in a tabbed dialog box



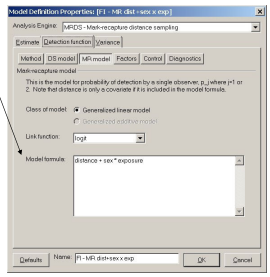
Detection function tab

- 5 methods at present
  - ds – CDS and MCDS (but no adjustment terms)
  - IO (independent observer) – both point and full independence
  - Trial – both point and full independence
- Choice of method determines which model you need
  - DS model = distance sampling model.
    - half-normal or hazard rate, optionally with covars in the scale parameter
  - MR model = mark recapture model
    - GLM with logit link



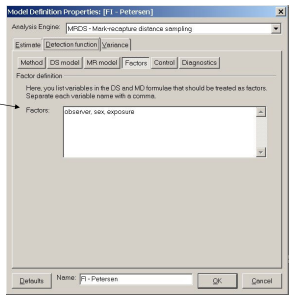
Model formulae

- Type in variable names joined by “+” (main effect), “:” (interaction), “*” (main effect + interaction)



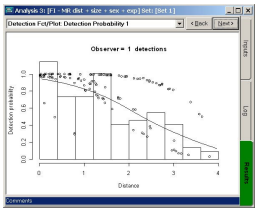
Factors

- Need to specify which variables in the formulae are factors



Results

- Produces
  - diagnostics (qq plots, detection function plots, goodness-of-fit tests)
  - parameter estimates, and estimated density and abundance
- Can customize plots (in Preferences)
- Plots stored as graphics files in a folder “R” within project data folder
- Results optionally stored in an .Rdata file, so if you know R software you can access them



R interface

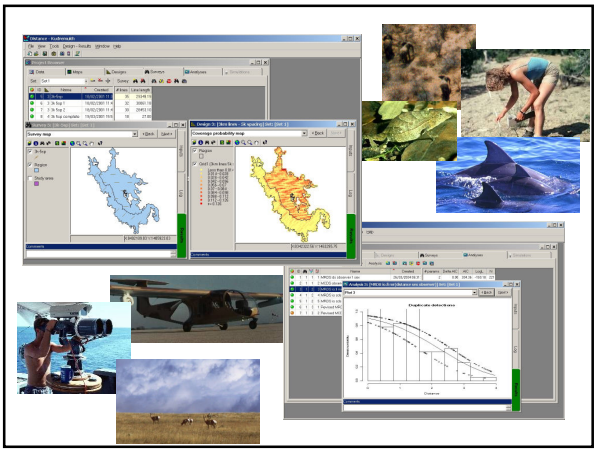
- MRDS implemented as a library (mrds) in R
- Primary functions are
  - ddf() – fits detection functions
  - dht() – estimates density and abundance given fitted detection function
- Both are documented in R help, but best way to learn to use them is by running analyses in debug mode in Distance

Crabeater seal example

Crabeater seal example

- To demonstrate:
  - Data structure
  - Analyses: FI and PI
  - Results
  - Running in debug mode
- Mention golf tees

Other capabilities of Distance

Capabilities of Distance

- Analysis:
  - Conventional distance sampling (CDS)
  - Multiple covariate distance sampling (MCDS)
  - Mark-recapture distance sampling (MRDS) (more to come!)
  - Density surface modelling (DSM) – to come!
  - Unequal coverage designs – to come!
- Survey design
  - Automated survey design – more to come!
- Simulation
  - to come I hope!