Multiple covariate distance sampling (MCDS)

- Aim: Model the effect of additional covariates on detection probability, in addition to distance, while assuming probability of detection at zero distance is 1

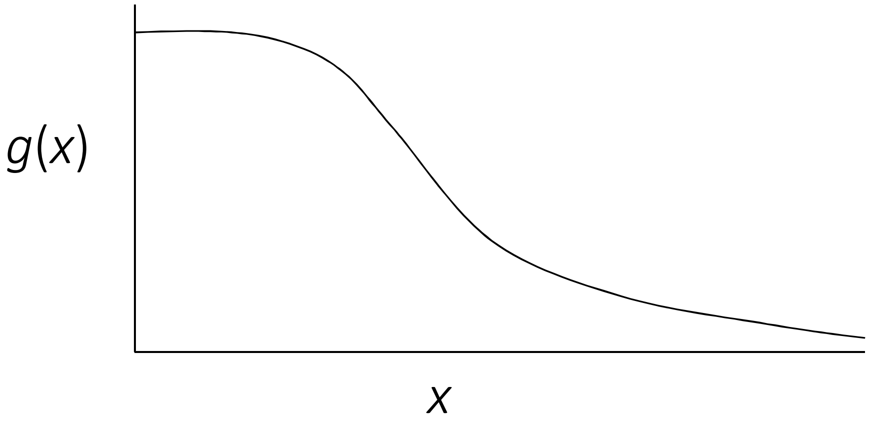
- References:
  - Marques (F) and Buckland (2004) Covariate models for the detection function. Chapter 3 in Buckland *et al*. (eds). Advanced Distance Sampling.
  - Marques (T) *et al*. (2007) Improving estimates of bird density using multiple covariate distance sampling. The Auk 127: 1229-1243.
  - Section 5.3 of Buckland *et al*. (2015) Distance Sampling: Methods and Applications

Contents

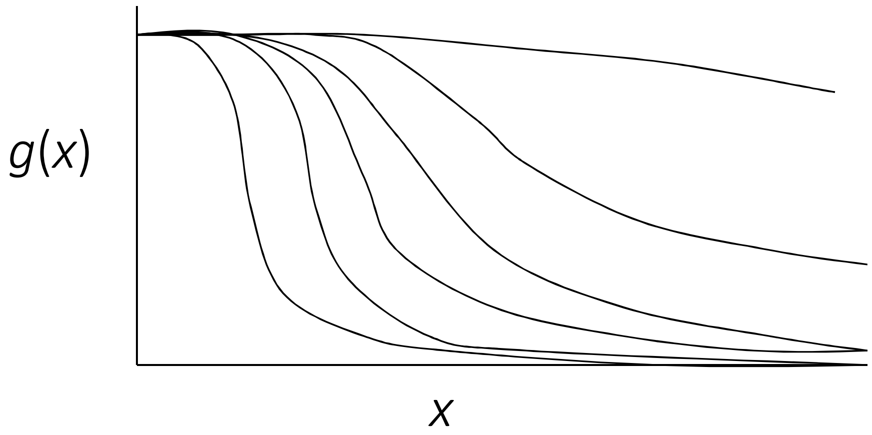
- Why additional covariates?
- Multiple covariate models
- Estimating abundance
- MCDS in Distance
- Complications
  - *Clustered populations*
  - *Adjustment terms*
  - *Stratification*
- MCDS analysis guidelines

Why additional covariates?

In conventional distance sampling (CDS) analysis all factors affecting detectability, except distance, are ignored



In reality, many factors may affect detectability



Sources of heterogeneity:

Object : species, sex, cluster size

Effort: observer, habitat, weather

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Examples of heterogeneity 1

Effect of time of day on Rufous Fantail birds in Micronesia (point transects). Ramsey et. al. 1987. Biometrics 43:1-11

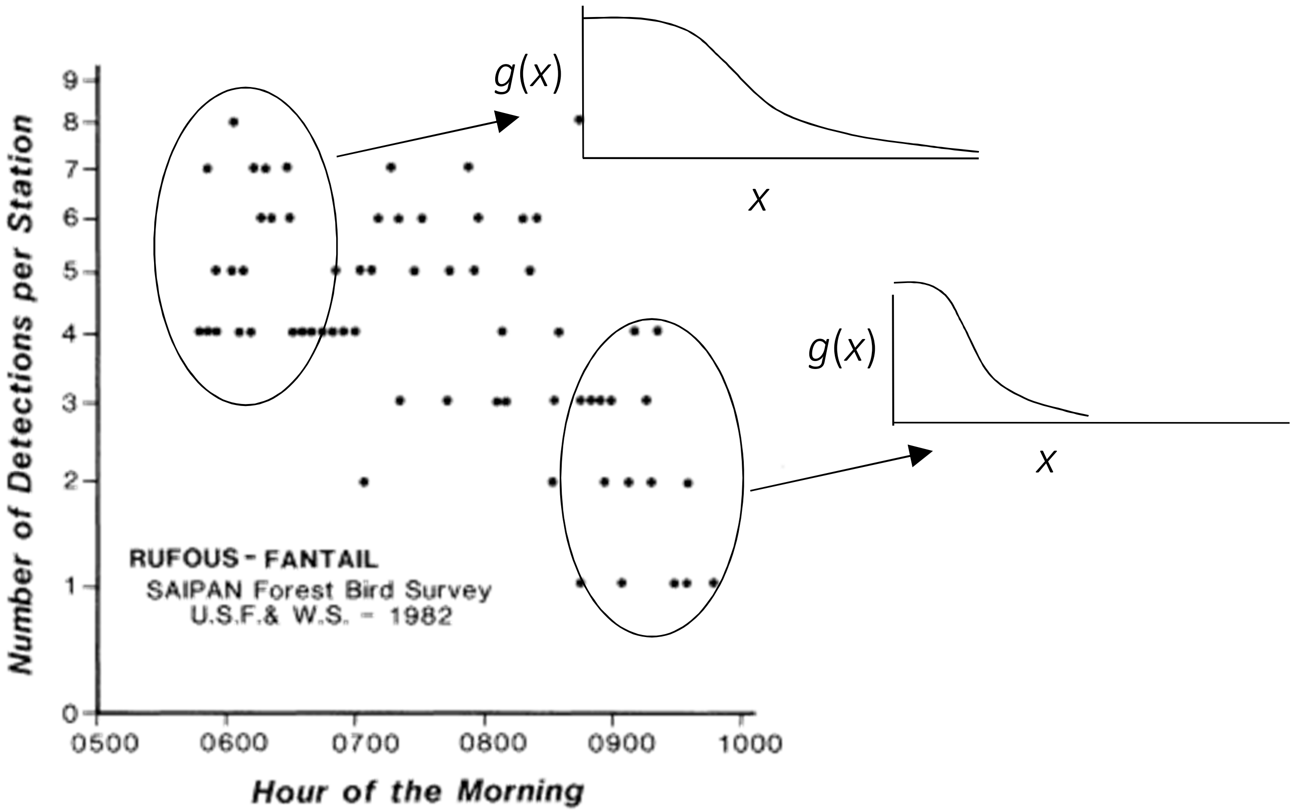


Figure 1. Station counts of Rufous Fantails on Saipan appear higher in the early morning hours than in the late morning ($n = 64$, $r = -.60$).

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Examples of heterogeneity 2

Effect of sea state (and other covariates) on sea turtles in the Eastern Tropical Pacific (shipboard line transects). Beavers and Ramsey, 1998, J. Wildl. Manage. 63: 948-957

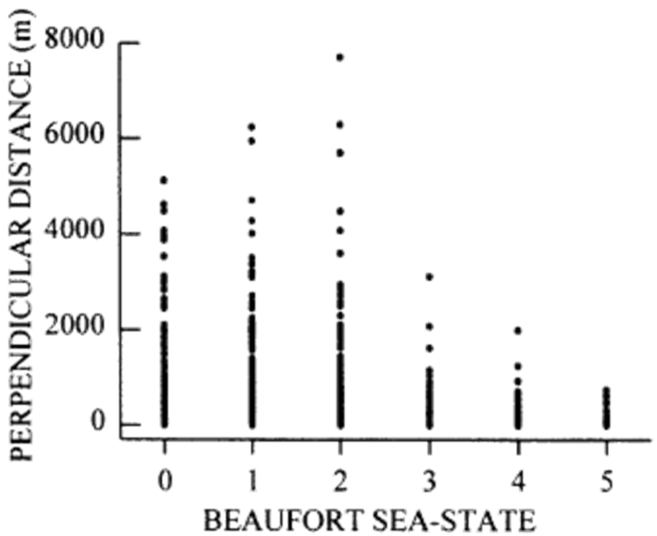


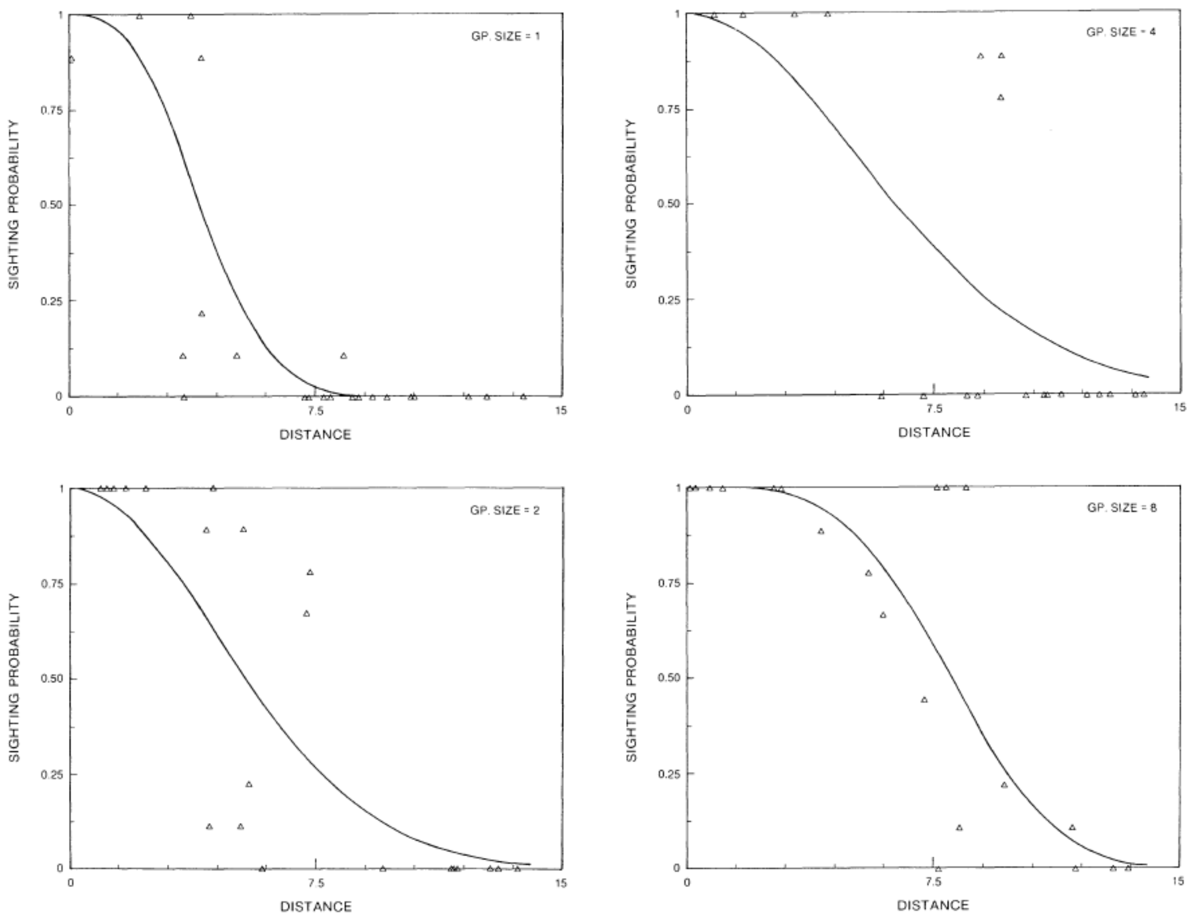
Fig. 2. Covariates of air temperature, sea surface temperature, and Beaufort sea-state plotted against unadjusted, ungrouped perpendicular sighting distances (m) of sea turtles in the eastern tropical Pacific, 1989–90.

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Examples of heterogeneity 3

Effect of cluster size on beer can detectability. Otto and Pollock, 1990, Biometrics 46: 239-245



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Why worry about heterogeneity?

In CDS, we use models that are pooling robust, so why worry about heterogeneity?

- Pooling robustness works for all but extreme levels of heterogeneity
- Potential bias if density is estimated at a 'lower level' than detection function (e.g. density by geographic region, detection function global)
- Could potentially increase precision of detection function estimate
- Interest in sources of heterogeneity in their own right (e.g. group size)

Dealing with heterogeneity

Stratification

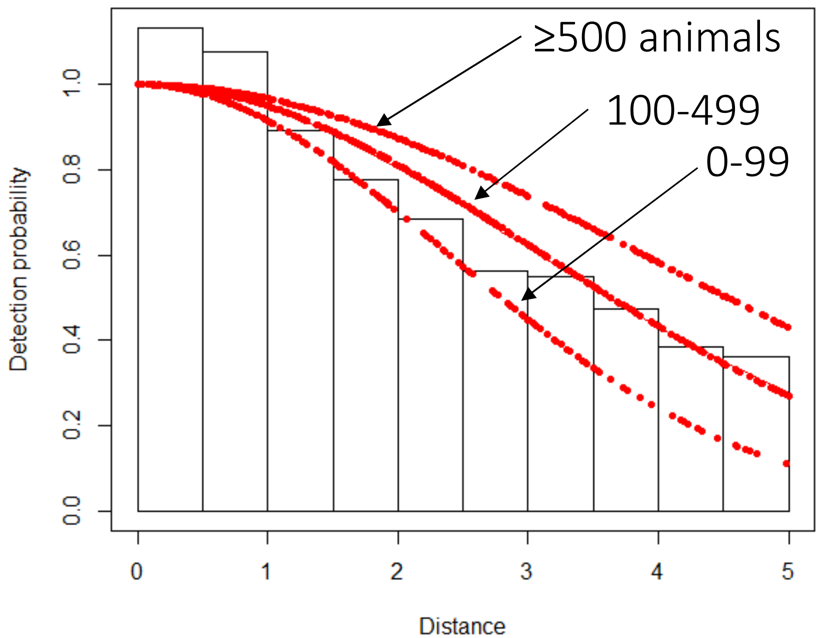
Requires estimating separate detection function parameters for each stratum,

- often not possible due to lack of data

Model as covariates in detection function

Allows a more parsimonious approach:

- can model effect of numerical covariates
- can 'share information' about detection function *shape* between covariate levels



CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Multiple covariate models Recap of CDS models

$$g(x) = \Pr[\text{animal at distance } x \text{ is detected}]$$

$$= k(x)\left[1 + \sum_{j=1}^{m} a_j p_j(x_s)\right] / c$$

Key function

j^{th} series adjustment term

Scaling constant to ensure $g(0) = 1$

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

CDS models continued

Key functions

Shape parameter

Hazard rate $k(x) = 1 - \exp\left[-\left(\frac{x}{\sigma} \right)^{-b} \right]$

Half-normal $k(x) = \exp\left(\frac{-x^2}{2\sigma^2} \right)$

Uniform $k(x) = 1$

Scale parameter

Series adjustments

Cosine $\cos(j\pi x_s)$

Polynomial x_s^j

Hermite poly. $H_j(x_s)$

x_s are scaled distances

CREEM Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Modelling with covariates

$g(x,\mathbf{z})$ = Pr[animal at distance x and covariates **z** is detected]

Assume the covariates affect the ***scale*** of the key function, not its ***shape***. So choose key functions with a scale parameter

Let $$\sigma(z) = \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j z_j\right)$$

e.g. Hazard rate $$k(x,z) = 1 - \exp\left[-\left(\frac{x}{\sigma(z)}\right)^{-b}\right]$$

Half normal $$k(x,z) = \exp\left(\frac{-x^2}{2\sigma(z)^2}\right)$$

k is used here to denote the "key" function

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Modelling with covariates

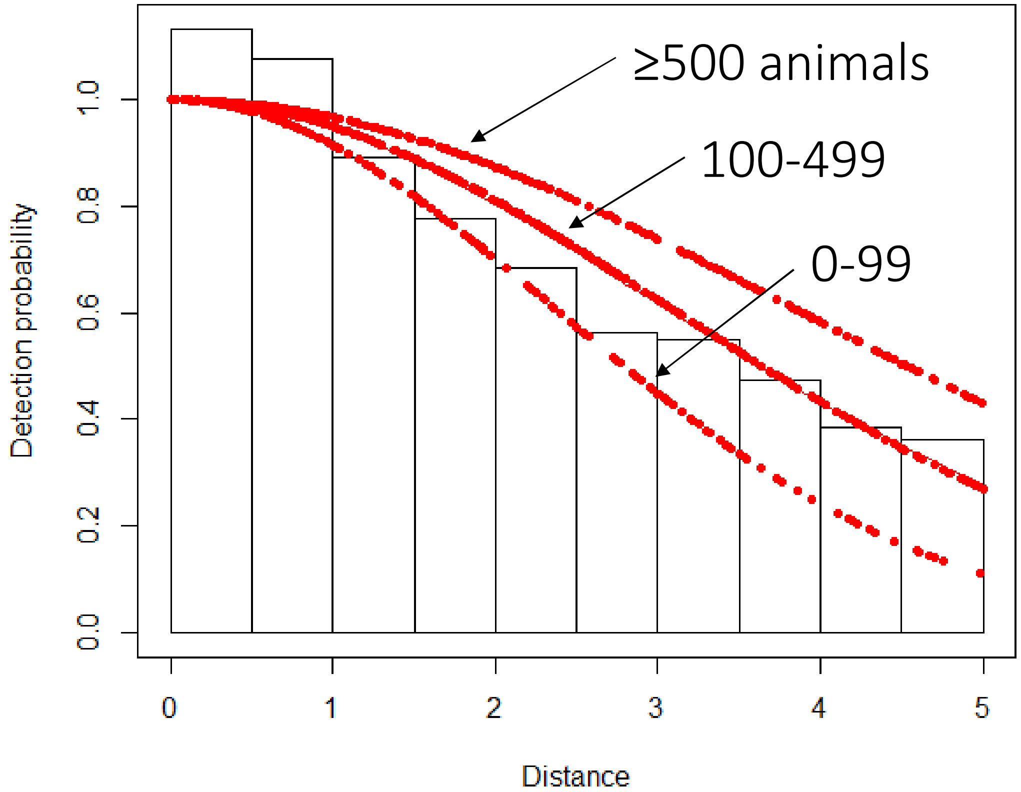
Example: Dolphin tuna vessel data

Model: half-normal, with no adjustments

Covariate: cluster size as factor (3 levels) with dummy variables, s_{d1} and s_{d2}

$$g(x, s) = exp\left(\frac{-x^2}{2\sigma(s)^2}\right)$$

$$\sigma(s) = exp(\beta_0 + \beta_1 s_{d1} + \beta_2 s_{d2})$$



CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Estimating abundance without covariates using Horvitz-Thompson estimator

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{Pr[animal\ included]} = \sum_{i=1}^{n} \frac{1}{\left[\frac{2wL\hat{P}_a}{A}\right]} = \frac{nA}{2wL\hat{P}_a}$$

Remember:
x's are the distances
and $P_a = {}^{\mu}/_{w}$

Recall that *f(x)* = pdf of observed *x*'s $= \frac{g(x)}{\int g(x)dx} = \frac{g(x)}{\mu} = \frac{g(x)}{wP_a}$

Because $g(0)$=1 by assumption, then *f*(0) = g(0)/μ = 1/μ = 1/ wP_a

So $\hat{N} = \frac{nA}{2wL\hat{P}_a} = \frac{nA}{2L} . \hat{f}(0)$

CREEM
Centre for Research into Ecological and Environmental Modelling

University of St Andrews

Estimating abundance with covariates

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{Pr[animal\ included]} = \sum_{i=1}^{n} \frac{1}{\left[\frac{2wL\hat{P}_a(z_i)}{A}\right]} = \frac{A}{2wL}\sum_{i=1}^{n} \frac{1}{\hat{P}_a(z_i)}$$

Now $f(x|\mathbf{z}) = \frac{g(x,\mathbf{z})}{\int g(x,\mathbf{z})dx} = \frac{g(x,\mathbf{z})}{\mu(\mathbf{z})} = \frac{g(x,\mathbf{z})}{wP_a(\mathbf{z})}$

Because $g(0,\mathbf{z})=1$ by assumption, then $f(0|\mathbf{z}) = {}^{g(0,\mathbf{z})}/_{\mu(\mathbf{z})} = {}^{1}/_{\mu(\mathbf{z})} = {}^{1}/_{wP_a(\mathbf{z})}$

So

$$\hat{N} = \frac{A}{2wL}\sum_{i=1}^{n} \frac{1}{\hat{P}_a(0|\mathbf{z}_i)} = \frac{A}{2L}\sum_{i=1}^{n} \hat{f}(0|\mathbf{z}_i)$$

Note similarity to CDS estimator

MCDS in Distance

In `ds` command, specify covariates in `formula` argument

```
ds(data, key, formula)
```

E.g. `ds(data=Dolphin, key="hn", formula=~size.class)`

Covariate type:

- Factor covariates classify the data into distinct classes or levels. Can be numerical or text. One parameter per factor level.
- Non-factor (i.e., continuous) covariates must be numerical (integer or decimal). One parameter per covariate + 1 for the intercept.

Complications 1. Clustered populations

When cluster size is a covariate:

- Distance recognizes cluster size because column is called `size` (i.e. reserved word)

E.g. `ds(data=Dolphin, key="hn", formula=~size)`

$$\hat{N}_{group} = \sum_{i=1}^{n} \frac{1}{Pr[group\ i\ included]} \qquad \hat{N} = \sum_{i=1}^{n} \frac{size\ of\ group\ i}{Pr[group\ i\ included]}$$

Estimate of group size is given by $\hat{E}[s] = \frac{\hat{N}}{\hat{N}_{group}}$

MCDS analysis guidelines

Choose covariates that are:
- independent of distance
- not strongly correlated with each other

Specifying the model:
- factor covariates generally harder to fit
- check convergence and monotonicity
- add only one covariate at a time
- where necessary, use starting values and bounds for parameters
- consider reducing the truncation distance, w, if more than 5% of the $P_a(z_i)$ are <0.2, or if any are less than 0.1