Explicitly incorporating stochastic availability processes in surveys of marine mammal abundance

Roland Langrock
David Borchers
Hans Skaug
Walter Zucchini

Universities of St Andrews, Bergen and Göttingen

Recap: conventional distance sampling

- model detection probability as a function of distance
- Crucial assumption: $g(0) = 1$ (certain detection on the trackline)
- serves as an "anchor" for quantifying the number of animals missed by the survey
- $g(0) \neq 1 \Rightarrow$ abundance estimates (negatively) biased
- for marine mammals typically $g(0) \neq 1$

Existing methods:

1.) correction factors

2.) mark-recapture distance sampling

3.) explicitly model the availability of animals via Poisson process

Recap: conventional distance sampling

- model detection probability as a function of distance
- Crucial assumption: $g(0) = 1$ (certain detection on the trackline)
- serves as an "anchor" for quantifying the number of animals missed by the survey
- $g(0) \neq 1 \Rightarrow$ abundance estimates (negatively) biased
- for marine mammals typically $g(0) \neq 1$

Existing methods:

1.) correction factors

2.) mark-recapture distance sampling

3.) explicitly model the availability of animals via Poisson process

Incorporate availability process in DS

Two main reasons for missing whales in DS surveys:

1.) animal is underwater

2.) animal is at surface but not detected

Aim: quantify 1.) and 2.) formulate a corresponding model for sightings.

Strategy:

- quantify availability via stochastic process (either in discrete or in continuous time)
- model conditional detection probability, given the animal is available, as a function of distance from animal to observer

Incorporate availability process in DS

Two main reasons for missing whales in DS surveys:

1.) animal is underwater

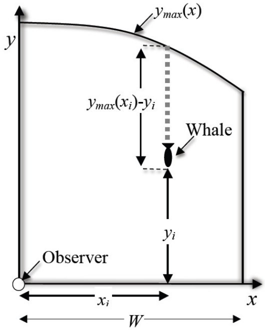
2.) animal is at surface but not detected

Aim: quantify 1.) and 2.) formulate a corresponding model for sightings.

Strategy:

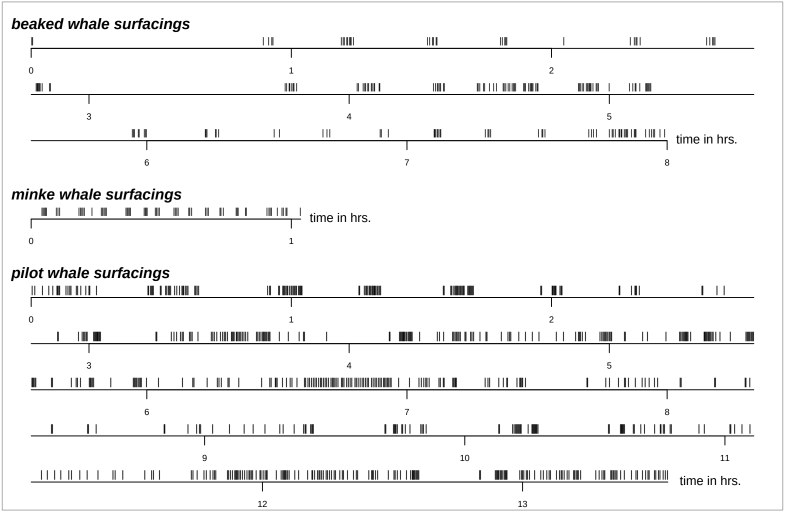
- quantify availability via stochastic process (either in discrete or in continuous time)
- model conditional detection probability, given the animal is available, as a function of distance from animal to observer

Model illustration (for a single whale)

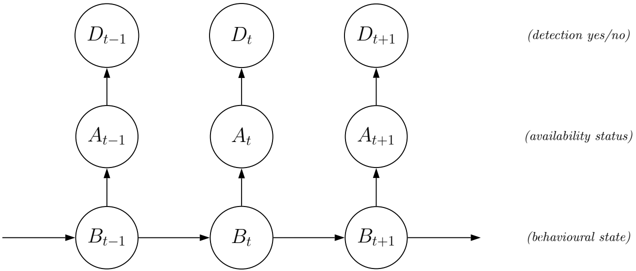


- observer at $(x = 0, y = 0)$
- assume animal movement is negligible
  $\rightarrow$ perpendicular distance fixed
- availability events occur according to some stochastic process
  (operating on the forward distance scale)
- detection process = thinned surfacing process (thinned according to some detection function)

Some motivating real surfacings data



Strategy I: Hidden Markov model formulation



- t: time (e.g. in seconds)
- N states (corresponding to "diving", "resting", ...)
- B_t: Markov or semi-Markov chain (not observed!)
- A_t is either 0 or 1, the probabilities of which depend on current state of B_t
- D_t depends on A_t and distance of animal to observer at time t

Strategy II: Markov-modulated Poisson process model formulation

- same principle as for HMMs, but model in continuous time
- extension of the Poisson process approach by Schweder and Skaug (1999)
- N-state continuous-time Markov process for switching between N behavioural states
- rate of signals (availability events) in state i: λ_i
- detections are then events of such a process after thinning according to detection probability, e.g.

$$h(x, y) = \mu \exp\left(-\frac{x^\gamma + y^\gamma}{\sigma^\gamma}\right)$$

Remarks on statistical inference

- Data required:
  - survey data (perp. and forward distances from observer to sighted animals)
  - auxiliary data (e.g., tag data) on availability (otherwise identifiability issues)
- Strategy:
  1.) estimate parameters of availability process from auxiliary data
  2.) fixing those parameters, use survey data and the full model to estimate parameters determining the conditional detection probability (as a function of distance)
  3.) based on fitted model, compute effective strip width, abundance estimates, etc.

Remarks on statistical inference

- Data required:
  - survey data (perp. and forward distances from observer to sighted animals)
  - auxiliary data (e.g., tag data) on availability (otherwise identifiability issues)
- Strategy:
  1.) estimate parameters of availability process from auxiliary data
  2.) fixing those parameters, use survey data and the full model to estimate parameters determining the conditional detection probability (as a function of distance)
  3.) based on fitted model, compute effective strip width, abundance estimates, etc.

Current/future work & references

- provide R package
- double-observer data
- incorporating possible movement of the animal
- yet more flexible stochastic processes for surfacings

Borchers, D., Zucchini, W., Heide-Jørgensen, M., Cañadas, A., Langrock, R., 2013. Using hidden Markov models to deal with availability bias on line transect surveys. *Biometrics*, in press.

Langrock, R., Borchers, D., Skaug, H., 2013. Markov-modulated nonhomogeneous Poisson processes for modeling detections in surveys of marine mammal abundance. *Journal of the American Statistical Association*, in press.