# Capture-recapture models in ecology: multi-state developments

Rachel McCrea

German Statistical Week

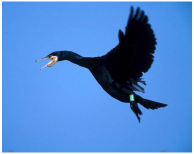
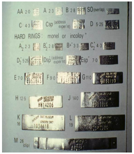
Hamburg, September 2015

# COLLABORATORS

- Hannah Worthington
- Ming Zhou
- Eleni Matechou
- Diana Cole
- Ruth King
- Richard Griffiths

# Outline

# INDIVIDUAL MARKING

## Capture-recapture data

- 1 0 0 1 0
- 1 1 0 1 1
- 0 0 1 0 1
- ...

# Closed population model, $M_t$

- $p_t$: probability an individual is captured at occasion $t$.
- Capture-recapture data and probabilities
  - `1 0 0 1 0` $p_1(1-p_2)(1-p_3)p_4(1-p_5)$
  - `1 1 0 1 1` $p_1p_2(1-p_3)p_4p_5$
  - `0 0 1 0 1` $(1-p_1)(1-p_2)p_3(1-p_4)p_5$
  - ...

# CLOSED POPULATION MODEL

- Some individuals will not be captured at all during the study;
- The encounter history for these individuals is given by
  - `0 0 0 0 0` $\qquad (1-p_1)(1-p_2)(1-p_3)(1-p_4)(1-p_5)$
- It is the number of individuals who are never captured that we need to estimate.

The likelihood has the form:

$$\boxed{L \propto \frac{N!}{(N-D!)} \prod_{i=1}^{D} \Pr(h_i) \times \Pr(h_0)^{N-D}} \tag{1}$$

- $h_i$: observed encounter history for individual $i$;
- $h_i$: observed encounter history of never encountered;
- $N$: population size;
- $D$: number of observed individuals.

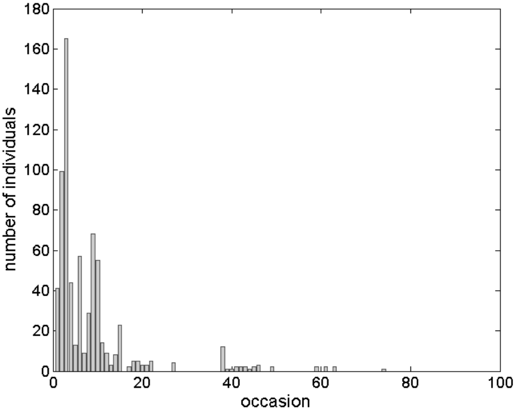
# CLOSED POPULATION MODEL, $M_b$

- $p$: probability of initial capture;
- $c$: probability of subsequent capture.

- Capture-recapture data and probabilities
  - `1 0 0 1 0` $p(1-c)(1-c)c(1-c)$
  - `1 1 0 1 1` $pc(1-c)cc$
  - `0 0 1 0 1` $(1-p)(1-p)p(1-c)c$
  - ...

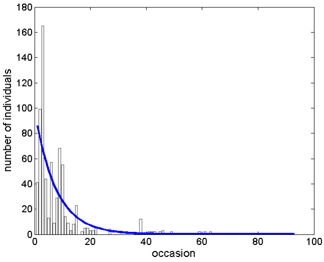
# REMOVAL DATA

$n_t$: size of sample removed at sample $t$.



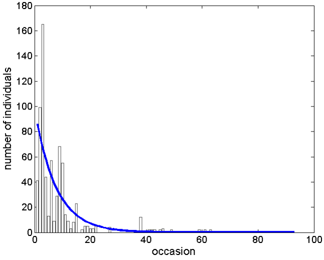
# LINK TO MODEL $M_b$

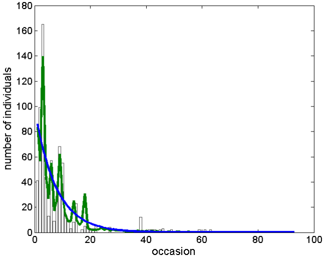
- Basic geometric model
  Pr(individual is removed at occasion t) $= (1-p)^{t-1}p$
- Same model as used for time to conception for human couples;
- Equivalent to estimating $p$ in $M_b$, and assuming $c = 0$.



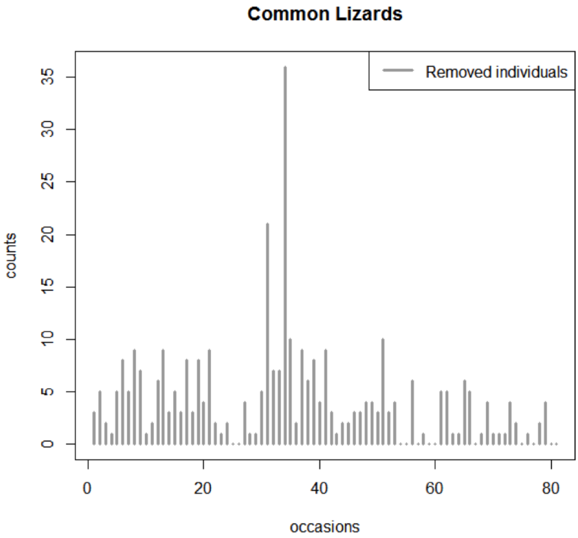
# LINK TO MODEL $M_b$

- Basic geometric model
  Pr(individual is removed at occasion t) $= (1-p)^{t-1}p$
- Same model as used for time to conception for human couples;
- Equivalent to estimating $p$ in $M_b$, and assuming $c = 0$.

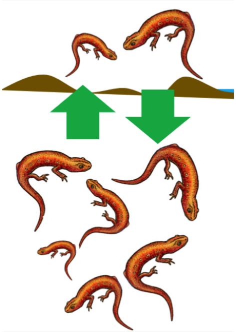


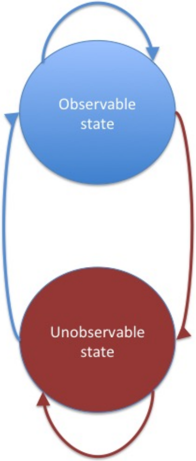
# WHY DO DATA EXHIBIT UNEXPECTED PEAKS?



# AN UNDERGROUND CITY?

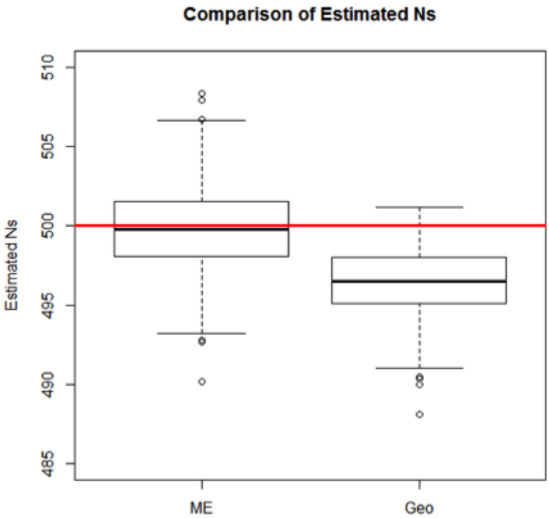


# MULTISTATE REMOVAL MODEL

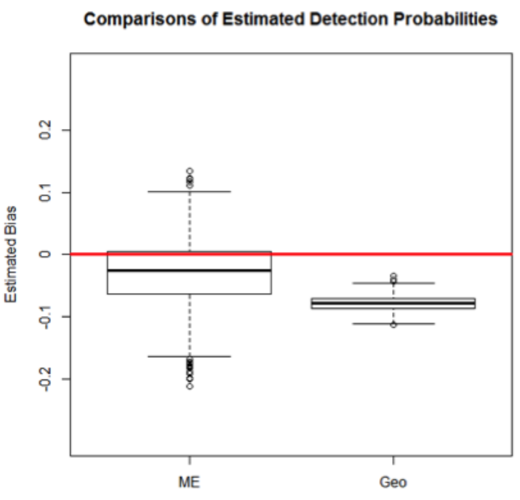



- Develop a two-state model, with one unobservable state with capture probability of 0;
- Naturally fits into a multievent framework, which is an HMM.

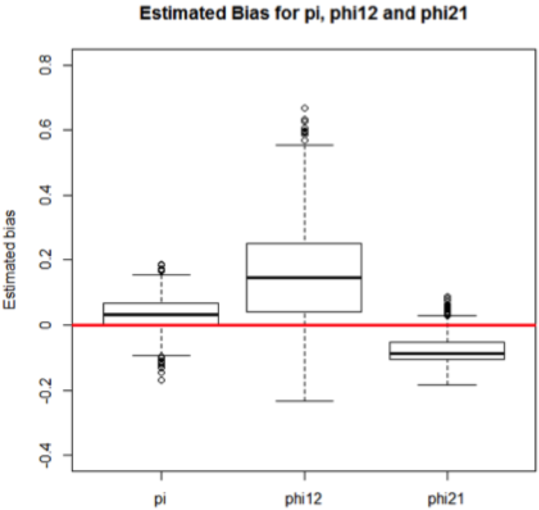
# SIMULATION RESULTS



# SIMULATION RESULTS



# Simulation results



# PARAMETER REDUNDANCY

- A model is parameter redundant if you cannot estimate all of the parameters;
- Parameter redundancy is diagnosed by forming a derivative matrix $D = \partial\kappa/\partial\theta$ where $\kappa$ denotes an exhaustive summary for a model that provides a unique representation of the model and $\theta$ denotes the parameters;
- If $\text{rank}(D) = \dim(\theta)$, all parameters are estimable;
- If $\text{rank}(D) < \dim(\theta)$ the model is parameter redundant.

# PARAMETER REDUNDANCY

- Model $\pi, p, \psi_{12}, \psi_{21}$ is parameter redundant;
- The estimable parameters are: $\pi p$, $p\psi_{21}$ and $p(\psi_{12} - 1) - \psi_{12} - \psi_{21}$.
- If $p$ is modelled using a temporal covariate, the model is full rank.

# Jolly-Seber model

- The studied population might not be closed, but still want to be able to estimate population size;
- Parameters for the Jolly-Seber model:
  - $N$: population size;
  - $\beta_t$ : proportion of individuals first available for capture at occasion t+1;
  - $p_t$: probability an individual is captured at occasion t;
  - $\phi_t$: probability an individual present in the study area at occasion t remains in the study area until occasion t+1.

# JOLLY-SEBER MODEL

- When forming the probability of an observed encounter history we need to sum over possible entry and departure times.
  - Suppose individual $i$ is first captured at occasion $f_i$ and last captured at occasion $l_i$;
  - $x_{ij} = 1$ if individual $i$ is captured at occasion $j$, $x_{ij} = 0$ otherwise.

$$\Pr(h_i) = \sum_{b=1}^{f_i} \sum_{d=l_i}^{T} \beta_{b-1} \left( \prod_{j=b}^{d-1} \phi_j \right) (1 - \phi_d) \left\{ \prod_{j=b}^{d} p_j^{x_{ij}} (1 - p_j)^{1-x_{ij}} \right\}$$

# Jolly-Seber model

Corresponding probability of an individual not captured during the study:

$$\Pr(h_0) = \sum_{b=1}^{T} \sum_{d=1}^{T} \beta_{b-1} \left( \prod_{j=b}^{d-1} \phi_j \right) (1 - \phi_d) \left\{ \prod_{j=b}^{d} (1 - p_j) \right\}$$

The likelihood, once again, has the same form:

$$\boxed{L \propto \frac{N!}{(N - D!)} \prod_{i=1}^{D} \Pr(h_i) \times \Pr(h_0)^{N-D}} \tag{1}$$

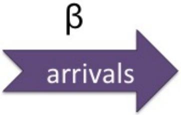
# STOPOVER MODEL

- Generalised version of the Jolly-Seber model (Pledger et al, 2009)
- Parameters are defined to be age-dependent, where **age** corresponds to the time spent in study area:
  - $N$: population size;
  - $\beta_t$ : proportion of individuals first available for capture at occasion t+1;
  - $p_t(a)$: probability an individual which entered the study $a$ occasions previously is captured at occasion t;
  - $\phi_t(a)$: probability an individual present in the study area at occasion t, which entered the study $a$ occasions previously, remains in the study area until occasion t+1.
- Can naturally be expressed in an HMM framework.

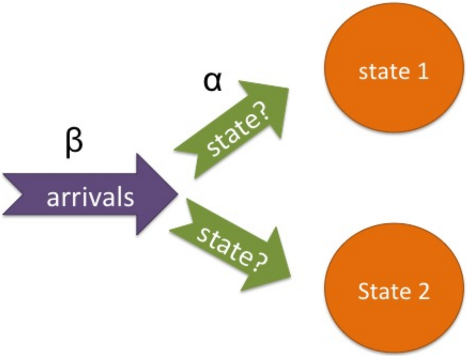
# MULTISTATE STOPOVER MODEL

- Individuals may be captured in different states;
- Multistate extensions exist for many capture-recapture models;
- Demonstrate that its possible to build transitions and state-dependence into the basic stopover model;
- HMM provides a useful, efficient framework for this.

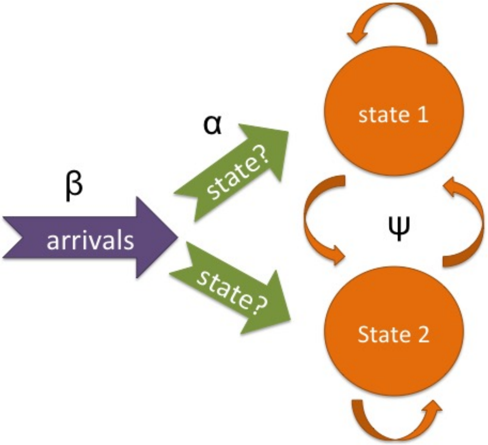
# MULTISTATE STOPOVER MODEL



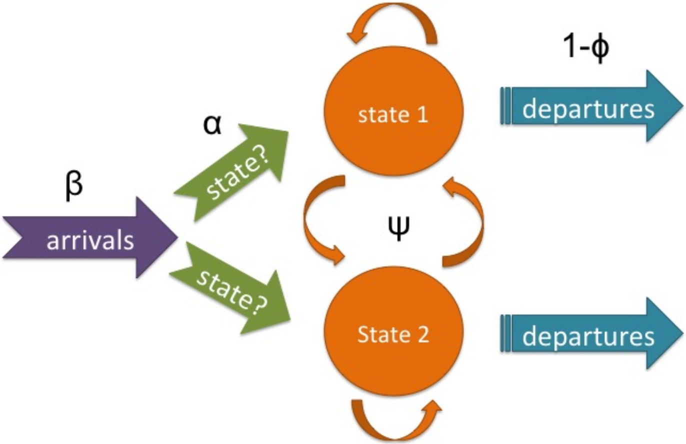
# MULTISTATE STOPOVER MODEL



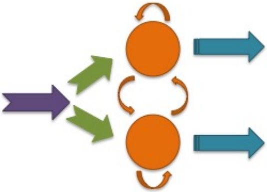
# MULTISTATE STOPOVER MODEL

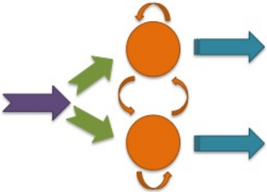


# Multistate stopover model

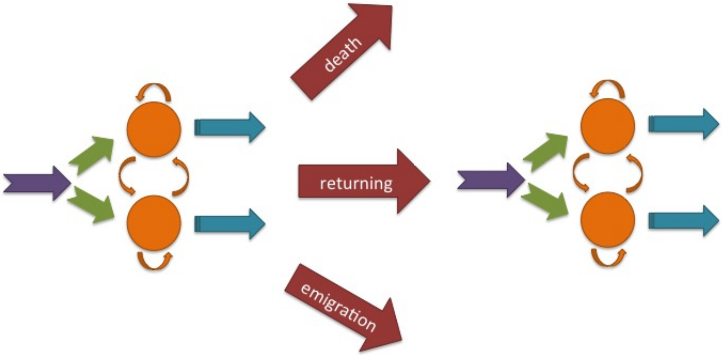


# INTEGRATING OVER MULTIPLE YEARS

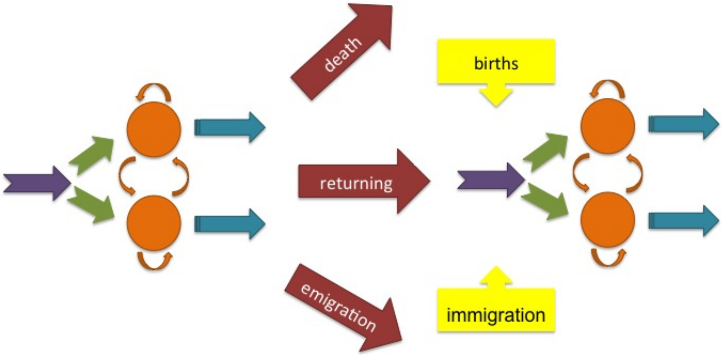




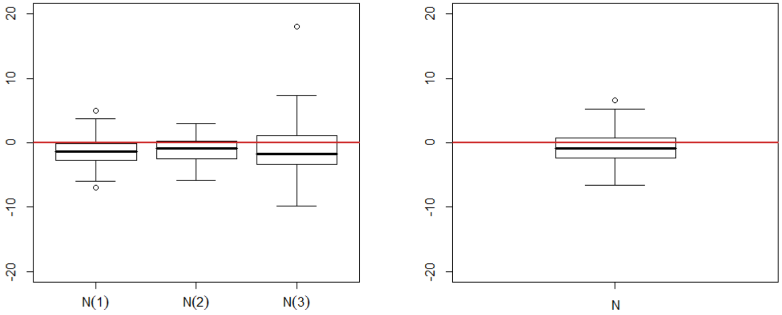
# Integrating over multiple years



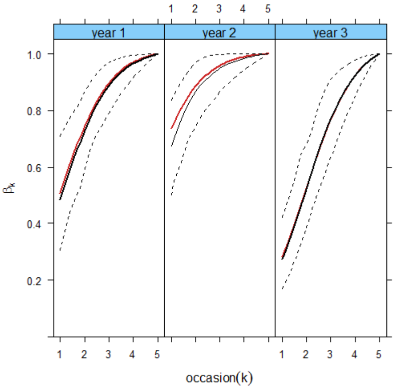
# INTEGRATING OVER MULTIPLE YEARS

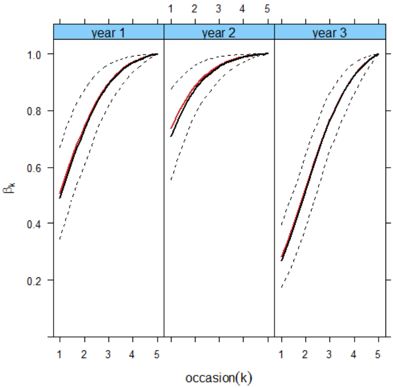


# Simulation results

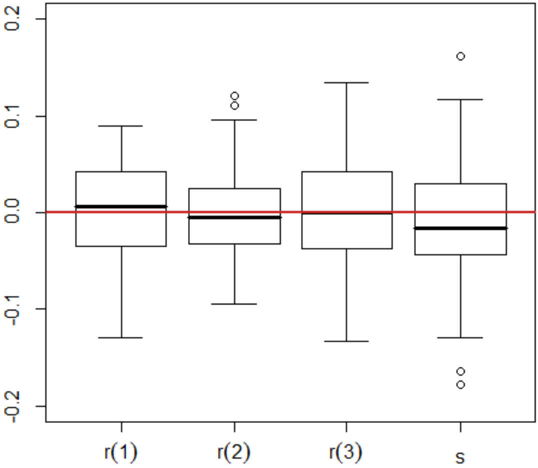


# Simulation results



# SIMULATION RESULTS



# ADVANTAGES

- General framework, with other models forming a special case;
  - Robust design (closed and open);
  - Closed population models - including a multistate closed population model (Worthington et al, 2015);
  - Stopover and Jolly-Seber models;
- Using all available data in a coherent model - compare Besbeas et al (2002);
- Natural generalisation of model selection methods for multistate models
  - Transdimensional simulated annealing (Brooks et al, 2003);
  - Step-wise procedures using score tests (McCrea and Morgan, 2011);

# Discussion

- Removal modelling:
  - Developed a new model for individuals moving into unobservable states;
  - Matechou et al (2015) has relaxed the assumption of closure within removal models and these methods could be included in the multievent removal framework;
  - Further investigation of the poor performance of near-redundant models.
- Stopover modelling:
  - HMM framework has provided an efficient approach for dealing with complex capture-recapture data;
  - Integrating the analysis of multiple years of data has improved precision and accuracy of parameter estimates;
  - Assessment of goodness-of-fit is an active area of research.

# REFERENCES


Besbeas, P., Freeman, S. N., Morgan, B. J. T. and Catchpole, E. A. (2002) Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58**, 540–547.

Brooks, S. P., Friel, N. and King, R. (2003) Classical model selection via simulated annealing. *Journal of the Royal Statistical Society Series B*. **65**, 503–520.

Matechou, E., McCrea, R.S., Morgan, B. J. T., Nash, D. and Griffiths, R. (2015) Renewal models for removal data. *In Revision.*

McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of capture-recapture data*. Chapman and Hall/CRC Press. Boca Raton.

McCrea, R. S. and Morgan, B. J. T. (2011) Multi-site mark-recapture model selection using score tests. *Biometrics*, **67**, 234–241.

Pledger, S., Efford, M., Pollock, K. H., Collazo, J. A. and Lyons, J. E. (2009) Stopover duration analysis with departure probability dependent on unknown time since arrival. *Environmental and Ecological Statistics*. **3**, 349–363.

Schwarz, C. J. and Arnason, A. N. (2006) A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*. **52**, 860–873.

Worthington, H. (2015) *The statistical development of multistate stopover models.* PhD Thesis, University of St Andrews.

Worthington, H., McCrea, R. S., King, R. and Griffiths, R. (2015) Estimation of population size when capture probabilities depend on individual discrete state information. In prep.