New models for reptile and amphibian removal data

Eleni Matechou, Rachel McCrea, Byron Morgan, Richard Griffiths, David Sewell and Brett Lewis

26 June 2014

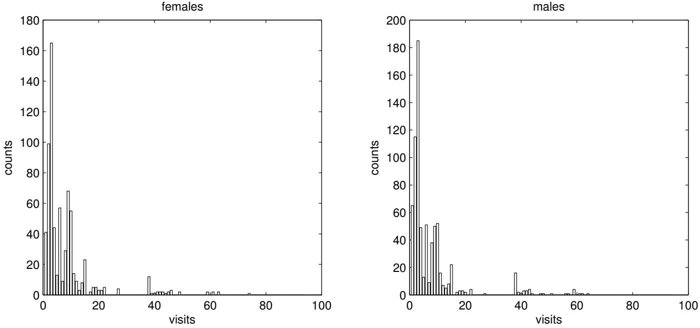
University of Kent

ncse National Centre for Statistical Ecology

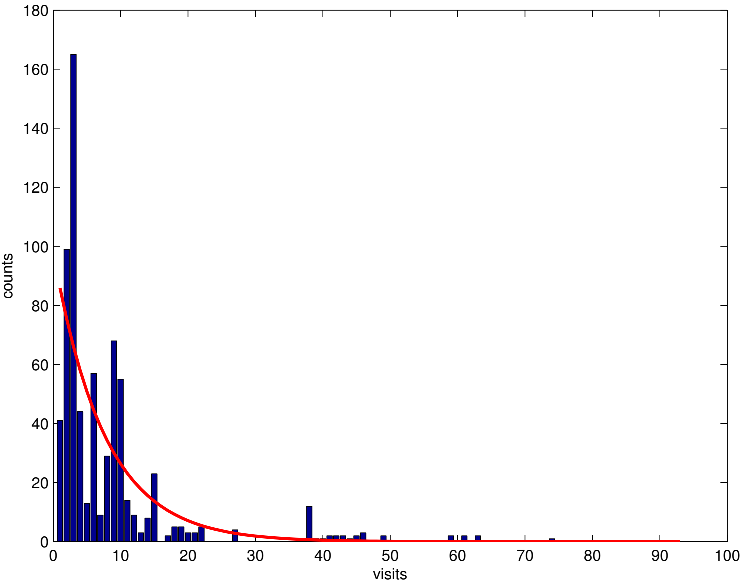
Outline

- Data
- Simple Removal model
- Covariates.
- Timed clearance
- Stopover models
- Discussion and recommendations.

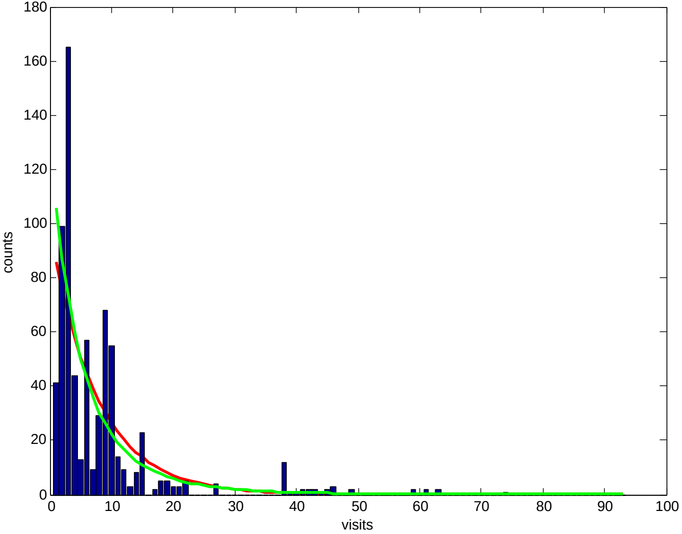
Basic Data: Great Crested Newts



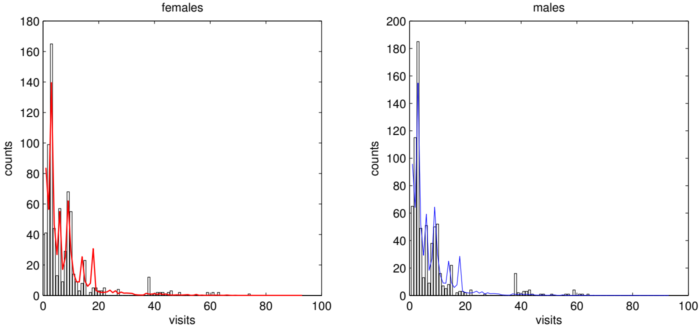
Result of fitting geometric model



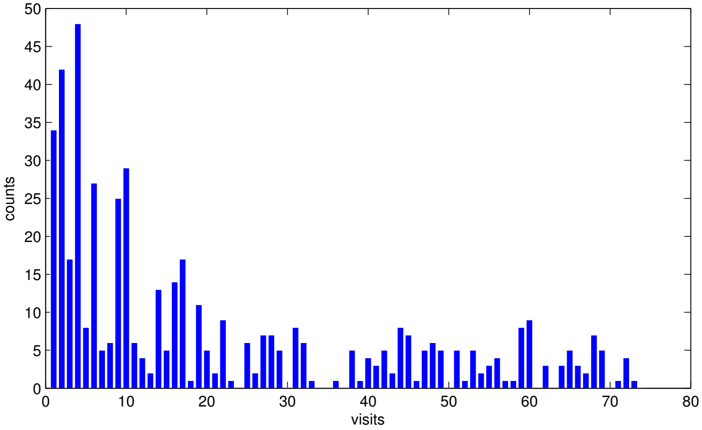
and then adding over dispersion

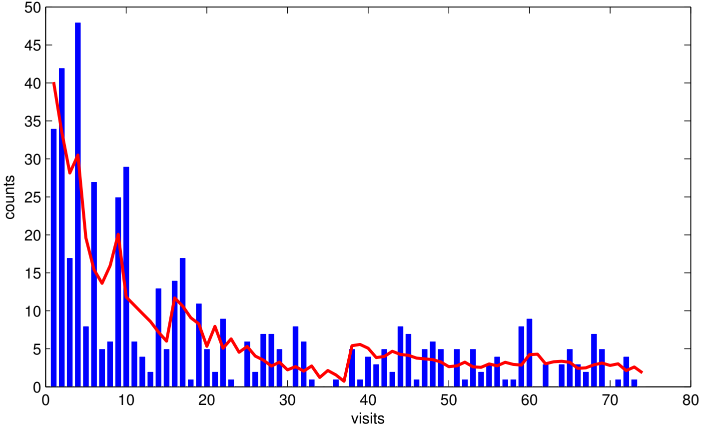


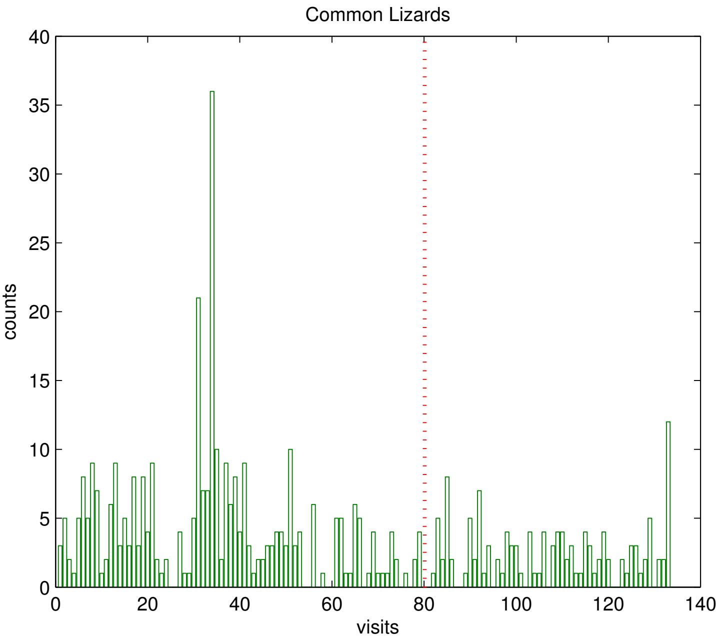
Geometric removal model with covariate

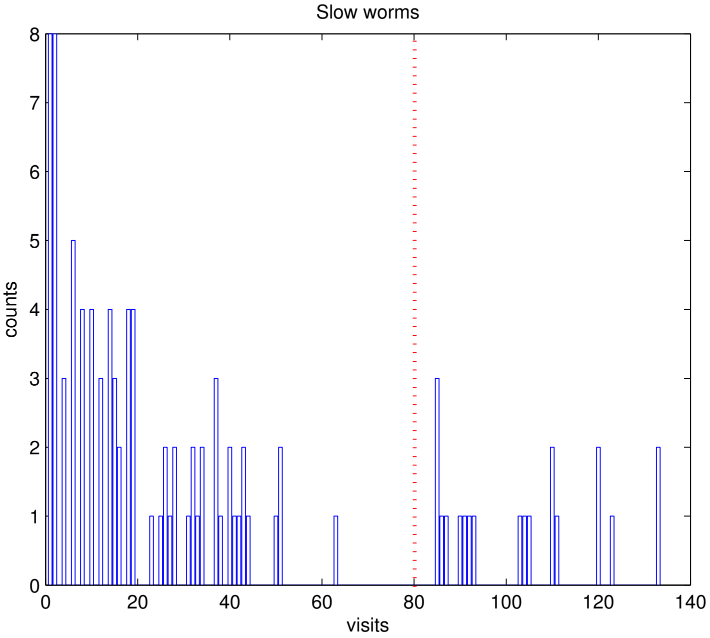


Timed clearance




counts
visits
0
5
10
15
20
25
30
35
40
45
50
0
10
20
30
40
50
60
70
80


Common Lizards
counts
visits
0
5
10
15
20
25
30
35
40
0
20
40
60
80
100
120
140


Slow worms
counts
visits
0
1
2
3
4
5
6
7
8
0
20
40
60
80
100
120
140

Modelling

- The simple removal model dates from papers by Moran (1951) and Zippin (1958)
- It assumes a constant capture probability, p.
- The other parameter is the desired population size, N.
- The same model applies to fecundability data, which record months to conception for human couples.
- 100 Smokers: 29 16 17 4 3 9 4 5 1 1 1 3 7; $\hat{N} = 95$.
- 486 Non-smokers: 198 107 55 38 18 22 7 9 5 3 6 6 12; $\hat{N} = 476$.

Three simple probability distributions

- Geometric: $\Pr(X = i) = (1-p)^{i-1}p$
- Multinomial:
  $\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_T = x_t) = \frac{N!}{x_1!x_2!,\ldots,x_t!}\prod_{i=1}^{T} p_i^{x_i}$
- Beta-geometric:
  Here we give p a Beta distribution.

Notation and likelihood

- N: initial population size
- n_k: size of the k^{th} sample removed from the population, $k = 1, 2, \ldots, T$
- $x_k = \sum_{j=1}^{k-1} n_j$, $k = 2, 3, \ldots, T+1; x_1 = 0$.

For example: $\mathbf{n} = (65, 115, 185, \ldots)$

$$\mathbf{x} = (0, 65, 180, 365, \ldots).$$

We then form the likelihood:

$$\mathrm{L}(N, p; \mathbf{n}) = \frac{N!}{(\prod_{k=1}^{T} n_k!)(N - x_{T+1})!} \left[\prod_{k=1}^{T} \{p(1-p)^{k-1}\}^{n_k} \right] (1-p)^{T(N - x_{T+1})}$$

Maximum-likelihood estimates

The likelihood simplifies to give

$$\mathrm{L}(N, p; \mathbf{n}) = \frac{N!}{(\prod_{k=1}^{T} n_k!)(N - x_{T+1})!} p^{x_{T+1}} (1-p)^{TN - \sum_{k=1}^{T+1} x_k}$$

Maximum-likelihood estimates of the two parameters are given by the solutions to the equations:

$$\hat{N} = \frac{x_{T+1}}{1-(1-\hat{p})^T}$$

$$\frac{\hat{p}}{1-\hat{p}} - \frac{T(1-\hat{p})^T}{1-(1-\hat{p})^T} = \frac{\sum_{k=1}^{T}(k-1)n_k}{x_{T+1}}$$

Covariates

- In practice we can expect the capture probability to vary
- for fecundability, conception might vary with age and parity
- for animals capture might vary with temperature. For GCN, minimum air temperature was used
- We may have a logistic transformation, for a covariate w:

$$p = \frac{1}{1 + \exp(\alpha + \beta w)}.$$

- It is often necessary to choose the best covariate(s) from a relevant set.
- Overdispersion may also be included, eg., using a beta-geometric distribution.

Model for timed clearences

- For the timed clearance data, at fixed times parts of the study area were cleared.
- This was modelled by assuming a global population of animals of size N, to be estimated.
- Fractions of this number were assumed to be available for capture during each time interval. Cf Stop-over modelling.
- Using maximum likelihood, the fractions were estimated as: 0.45, 0.10, 0.17, 0.28, 0.00.
- The estimated number of animals not observed was $\hat{N}_0 = 51$.

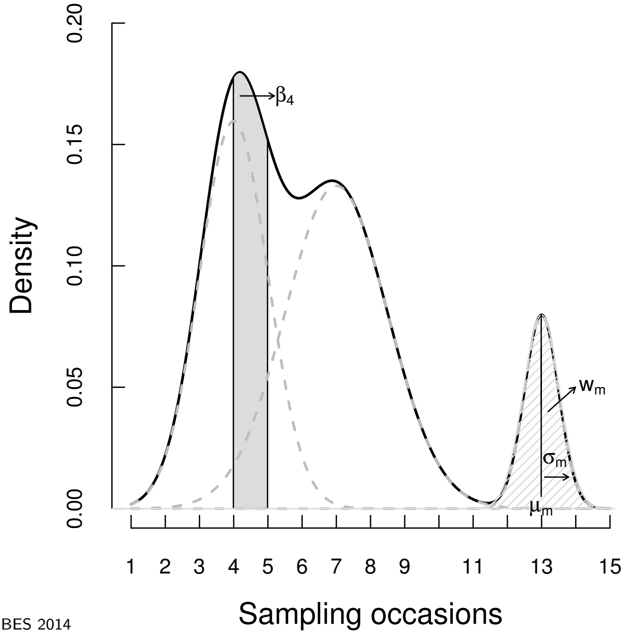
Stop-over model parameters

- N: population size; M: number of arrival groups.
- w_m, μ_m and σ_m, $m = 1, \ldots, M$: The population fractions, mean arrival times and standard deviations of arrival times of the M arrival groups, $\sum_{m=1}^{M} w_m = 1$. The population fraction that arrived between occasions $b-1$ and b is denoted by β_{b-1}. In terms of the mixture components,

$$\beta_{b-1} = \sum_{m=1}^{M} w_m \left\{F_m(b) - F_m(b-1)\right\}, \; b = 2, \ldots, T-1$$

where $F_m(b) = P(X \leq b)$ when $X \sim N(\mu_m, \sigma_m^2)$. The first and last intervals are open-ended with
$\beta_0 = \sum_{m=1}^{M} w_m F_m(1)$ and
$\beta_{T-1} = 1 - \sum_{m=1}^{M} w_m F_m(T-1)$, $\forall m$, ensuring that the entry parameters sum to 1 i.e. $\sum_{b=1}^{T} \beta_{b-1} = 1$.

Stop-over model: $M = 3$



Forming the likelihood

If an individual belongs to category t, $t = 1, \ldots, T$, then it was removed on sampling occasion t. The unknown number of individuals that were never detected and therefore never removed belong to category $T + 1$.

The probability of belonging to category t, γ_t, is:

$$\gamma_t = \begin{cases} \sum_{b=1}^{t} \left[\beta_{b-1} \left\{\prod_{k=b}^{t-1}(1-p_k)\right\}\right] p_t, & t = 1, \ldots, T \\ 1 - \sum_{t=1}^{T} \gamma_t = \sum_{b=1}^{T} \left[\beta_{b-1} \prod_{k=b}^{T}(1-p_k)\right], & t = T+1 \end{cases},$$

The likelihood is multinomial with $T + 1$ cells, γ_t, $t = 1, \ldots, T + 1$ probabilities and n_t, $t = 1, \ldots, T + 1$ frequencies.

Stop-over likelihood

The parameters are given by

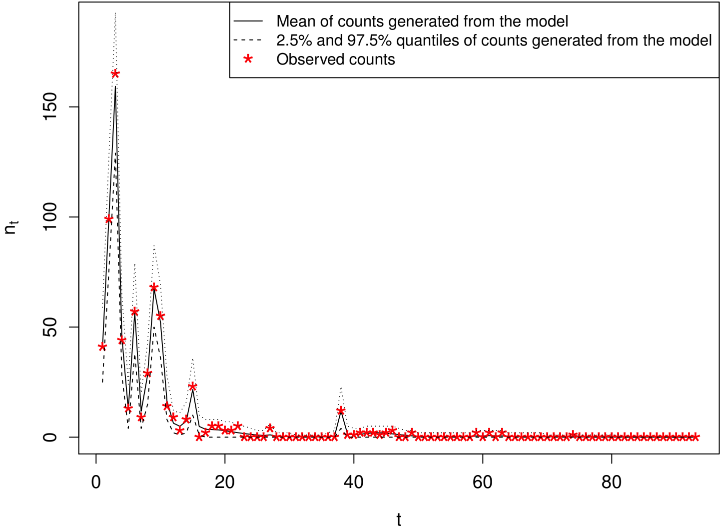
$$\boldsymbol{\theta} = (M, N, (w_m, \mu_m, \sigma_m)_{m=1,\dots,M}, (p_t)_{t=1\dots,T})$$

and the likelihood is:

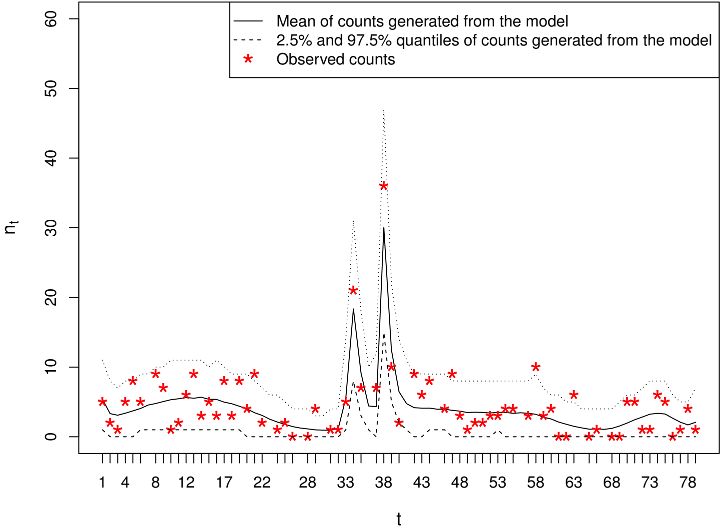
$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{N!}{\left(\prod_{t=1}^{T} n_t!\right)(N-D)!}\left\{\prod_{t=1}^{T} \gamma_t^{n_t}\right\}\gamma_{T+1}^{N-D},$$

where $D = \sum_{t=1}^{T} n_t$. We assume constant capture in the applications.

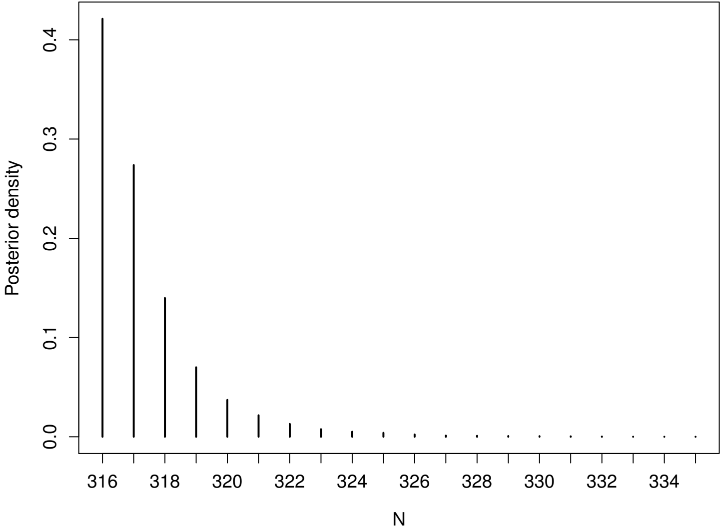
Stop-over, RJMCMC, GCN



Stop-over, RJMCMC, lizard



Posterior distribution for N



References

1. Dorazio, Jelks and Jordan (2005)*Biometrics*, **61**, 1093–1101.
2. Matechou, E., Nicholls, G., Morgan, B.J.T., Collazo, J.A. and Lyons, J.E. (2014). Bayesian mixture models for stopover data. Submitted for publication.
3. Moran (1951) *Biometrika*, **38**, 307–311.
4. Ridout and Morgan (1991) *Biometrics*, **47**, 1423–1433.
5. Zippin (1958) *J. Wildlife Management*, **22**, 82–90.