

Open models for removal data

Eleni Matechou, Rachel S. McCrea, Byron J. T. Morgan,
Darryn Nash and Richard A. Griffiths

ISEC

Seattle, Tuesday, 28/06/16



OUTLINE

REMOVAL DATA
MODELS
RESULTS
SUMMARY
FUTURE WORK

To appear in Annals of Applied Statistics

OPEN MODELS FOR REMOVAL DATA

BY ELENI MATECHOU, RACHEL S. MCCREA, BYRON J.T. MORGAN,
DARRYN J. NASH AND RICHARD A. GRIFFITHS

University of Kent

Individuals of protected species, such as amphibians and reptiles, often need to be removed from sites before development commences. Usually, the population is considered to be closed. All individuals are assumed to i) be present and available for detection at the start of the study period and ii) remain at the site until the end of the study, unless they are detected. However, the assumption of population closure is not always valid. We present new removal models which allow for population renewal through birth and/or immigration, and population depletion through sampling as well as through death/emigration. When appropriate, productivity may be estimated and a Bayesian approach allows the estimation of the probability of total population depletion. We demonstrate the performance of the models using data on common lizards, *Zootoca vivipara*, and great crested newts, *Triturus cristatus*.

REMOVAL DATA

MODELS

RESULTS

SUMMARY

FUTURE WORK

POLICY

When protected species occur on a site scheduled for development, there may be a legal requirement for them to be relocated to an alternative site before the development can proceed.²

In the UK, such relocations, often required for amphibians and reptiles, are typically achieved by conducting regular surveys during which detected animals are relocated to a suitable alternative habitat.

²Germano, J. M., Field, K. J., Griffiths, R. A., Clulow, S., Foster, J., Harding, G. and Swaisgood, R. R. 2015. Mitigation-driven translocations: are we moving wildlife in the right direction? *Frontiers in Ecology and the Environment* **13** 100-105.

DATA

TABLE 1

Removal sampling: counts of common lizards removed in 2007. A dash indicates that no sampling took place on that day. Temp denotes the maximum daily temperature ($^{\circ}\text{C}$) recorded on each visit to the study site. Here, the number of removed individuals, $D = 213$.

Date	Count	Temp	Date	Count	Temp	Date	Count	Temp	Date	Count	Temp
24-May	0	20	02-Jul	2	15	10-Aug	7	19	18-Sep	6	16
25-May	-	-	03-Jul	7	17	11-Aug	-	-	19-Sep	6	16
26-May	-	-	04-Jul	3	16	12-Aug	-	-	20-Sep	6	14
27-May	-	-	05-Jul	1	15	13-Aug	12	18	21-Sep	4	14
28-May	-	-	06-Jul	1	15	14-Aug	8	20	22-Sep	-	-
29-May	-	-	07-Jul	-	-	15-Aug	1	18	23-Sep	4	14
30-May	-	-	08-Jul	-	-	16-Aug	10	14	24-Sep	1	14
31-May	-	-	09-Jul	1	16	17-Aug	9	16	25-Sep	5	11
01-Jun	0	22	10-Jul	-	-	18-Aug	-	-	26-Sep	5	12
02-Jun	-	-	11-Jul	-	-	19-Aug	-	-	27-Sep	0	12
03-Jun	-	-	12-Jul	2	17	20-Aug	6	17	28-Sep	2	12
04-Jun	-	-	13-Jul	-	-	21-Aug	-	-	29-Sep	-	-
05-Jun	1	21	14-Jul	-	-	22-Aug	-	-	30-Sep	1	14
06-Jun	0	21	15-Jul	-	-	23-Aug	-	-	01-Oct	3	16
07-Jun	2	21	16-Jul	0	18	24-Aug	-	-	02-Oct	0	16
08-Jun	0	23	17-Jul	0	18	25-Aug	-	-	03-Oct	1	16
09-Jun	-	-	18-Jul	3	18	26-Aug	-	-	04-Oct	0	16
10-Jun	-	-	19-Jul	-	-	27-Aug	-	-	05-Oct	0	16
11-Jun	1	28	20-Jul	0	13	28-Aug	-	-	06-Oct	0	16
12-Jun	1	20	21-Jul	-	-	29-Aug	-	-	07-Oct	-	-
13-Jun	-	-	22-Jul	-	-	30-Aug	-	-	08-Oct	0	16
14-Jun	-	-	23-Jul	1	18	31-Aug	-	-	09-Oct	1	14
15-Jun	0	13	24-Jul	-	-	01-Sep	-	-	10-Oct	1	15
16-Jun	-	-	25-Jul	0	16	02-Sep	9	18	11-Oct	1	15
17-Jun	-	-	26-Jul	-	-	03-Sep	-	-	12-Oct	0	16
18-Jun	-	-	27-Jul	-	-	04-Sep	-	-	13-Oct	1	16
19-Jun	-	-	28-Jul	-	-	05-Sep	1	18	14-Oct	-	-
20-Jun	1	18	29-Jul	-	-	06-Sep	8	17	15-Oct	0	16
21-Jun	2	19	30-Jul	-	-	07-Sep	2	18	16-Oct	0	12
22-Jun	3	16	31-Jul	-	-	08-Sep	-	-	17-Oct	0	15
23-Jun	-	-	01-Aug	1	18	09-Sep	11	18	18-Oct	0	15
24-Jun	-	-	02-Aug	0	15	10-Sep	7	18	19-Oct	0	13
25-Jun	2	15	03-Aug	1	15	11-Sep	9	18	20-Oct	-	-
26-Jun	-	-	04-Aug	-	-	12-Sep	1	18	21-Oct	-	-
27-Jun	2	15	05-Aug	-	-	13-Sep	1	17	22-Oct	0	12
28-Jun	2	17	06-Aug	1	16	14-Sep	5	18	23-Oct	0	13
29-Jun	5	18	07-Aug	1	18	15-Sep	-	-	24-Oct	0	9
30-Jun	-	-	08-Aug	4	20	16-Sep	-	-	25-Oct	0	19
01-Jul	-	-	09-Aug	3	20	17-Sep	4	12			



TABLE 2

Removal sampling: counts of male great crested newts removed in 2010. A dash indicates that no sampling took place on that day. Temp denotes the minimum daily air temperature ($^{\circ}\text{C}$) recorded on each visit to the study site. Here, the number of removed individuals, $D = 741$. The nine missing covariate values were imputed using the average of the previous and next observed values.

Date	Count	Temp	Date	Count	Temp	Date	Count	Temp
18-Mar	65	7	13-Apr	0	2.8	08-May	1	7.1
19-Mar	115	5.2	14-Apr	0	6.5	10-May	0	9
20-Mar	185	12.4	15-Apr	0	1.1	11-May	0	-0.7
21-Mar	49	6.5	16-Apr	0	-	12-May	0	2.2
22-Mar	64	3.6	17-Apr	0	3.5	14-May	0	13.2
23-Mar	9	2.3	18-Apr	-	-	14-May	1	0.6
24-Mar	38	5.2	19-Apr	-	-	15-May	1	6.9
25-Mar	102	12.4	20-Apr	-	-	16-May	0	1.4
26-Mar	16	5.2	21-Apr	0	-1	17-May	4	1.4
27-Mar	7	2.7	22-Apr	0	-3.2	18-May	0	-
28-Mar	5	3.1	23-Apr	0	-4.2	19-May	1	2.2
29-Mar	8	10.8	24-Apr	1	7	20-May	1	10.8
30-Mar	22	-	25-Apr	0	6	21-May	0	9.3
31-Mar	0	3.1	26-Apr	16	4.2	22-May	-	-
01-Apr	2	-	27-Apr	2	4	23-May	-	-
02-Apr	-	-	28-Apr	1	7	24-May	0	-
03-Apr	-	-	29-Apr	3	4.3	25-May	0	-
04-Apr	-	-	30-Apr	3	7.7	26-May	0	18
05-Apr	-	-	01-May	4	5.1	27-May	0	10.3
06-Apr	-	-	02-May	1	8.2	28-May	0	-
07-Apr	3	3.7	03-May	0	4.2	29-May	0	15
08-Apr	2	2.1	04-May	0	2.5	30-May	0	11.5
09-Apr	0	1.3	05-May	1	2	31-May	0	-
10-Apr	4	1.8	06-May	1	4.9	01-Jun	0	5.1
11-Apr	0	3.2	07-May	0	3	02-Jun	0	3.9
12-Apr	0	6.1	08-May	0	3	03-Jun	0	3.9



REMOVAL DATA

MODELS

RESULTS

SUMMARY

FUTURE WORK

POPULATION CLOSURE ASSUMPTION

Existing removal models assume *population closure* for the duration of the study.

However, it is often the case when sampling amphibians and reptiles that the study period encompasses the reproduction period, and hence the population is renewed by the emergence of newborn individuals, violating the assumption of closure.

Additionally, when the study takes place at an unfenced breeding site and commences before the start of the breeding season, individuals can migrate into or out of the breeding site, thus again violating the assumption of population closure.

POPULATION RENEWAL AND EMIGRATION/DEATH

Motivated by these and similar data sets, we develop a model that allows for *renewal groups*: a, potentially, unknown number of groups of unknown size of individuals that emerge/arrive at unknown times.

We fit the model using a Bayesian approach and (if the number of renewal groups is unknown) use a reversible jump³ MCMC algorithm to fit the model and estimate the renewal pattern and the number of renewal groups.

We obtain the posterior distribution of the number of undetected individuals, N , while accounting for individuals that may die or permanently emigrate during the study period.

³Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82 711-732.

SET UP

Each individual in the population belongs to one of $T + 2$ categories.

If an individual belongs to category t , $t = 1, \dots, T$, then it was removed at sampling occasion t .

The N individuals that were never detected and not removed belong to category $T + 1$.

The remainder of the individuals left the study site either through death or emigration before being detected and before the study ended.

MODEL PARAMETERS

- $\beta_{t-1}, t = 1, \dots, T$: entry parameters^{4,5}. $\Omega = \sum_{t=1}^T \beta_{t-1}$.
- ϕ_t : survival probability.
- p_{t-b} : detection probability with p_0 denoting the probability of being detected upon arrival.

⁴Schwarz, C. J. and Arnason, A. N. 1996. A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* **52** 860-873.

⁵Pledger, 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 (Edited by D.L.Thomson, E.G.Cooch and M.J. Conroy)* **3** 349-363.

MODEL

The probability of being in category t is:

$$\gamma_t = \begin{cases} \sum_{b=1}^t \left[\beta_{b-1} \left\{ \prod_{k=b}^{t-1} \phi_k \right\} p_{t-b} \right], & t = 1, \dots, T \\ \sum_{b=1}^T \left[\beta_{b-1} \left\{ \prod_{k=b}^{T-1} \phi_k \right\} \left(1 - \sum_{k=b}^T p_{k-b} \right) \right] + (1 - \Omega), & t = T + 1 \\ 1 - \sum_{t=1}^{T+1} \gamma_t, & t = T + 2 \end{cases}$$

MODEL

If we denote the vector of model parameters by $\boldsymbol{\theta}$, the data are described by the multinomial distribution (conditioning on the event of surviving until detection or until the end of the study for undetected individuals)

$$\mathcal{P}_c(\mathbf{n}|\boldsymbol{\theta}) = \frac{(N + D)!}{\left(\prod_{t=1}^T n_t!\right) N!} \left\{ \prod_{t=1}^T \left(\frac{\gamma_t}{\sum_{t=1}^{T+1} \gamma_t} \right)^{n_t} \right\} \left(\frac{\gamma_{T+1}}{\sum_{t=1}^{T+1} \gamma_t} \right)^N .$$

MODEL

We allow for G renewal groups, each of unknown size and unknown arrival/emergence pattern, and each modelled by a normal pdf.

π_g , μ_g , and σ_g^2 , $g = 1, \dots, G$, are respectively, population fractions, mean arrival times, and variance of arrival times of the G renewal groups, with $\sum_{g=1}^G \pi_g = 1$.

The proportion of individuals with arrival time in the interval $(\tau_{b-1}, \tau_b]$ is given as

$$\beta_{b-1} = \begin{cases} \sum_{g=1}^G \pi_g P(\tau_{b-1} < X_g \leq \tau_b), & b = 2, \dots, T \\ \sum_{g=1}^G \pi_g P(X_g \leq \tau_1), & b = 1 \end{cases}$$

where $X_g \sim N(\mu_g, \sigma_g^2)$.

REMOVAL DATA

MODELS

RESULTS

SUMMARY

FUTURE WORK

COMMON LIZARDS

We logistically regress detection probability on maximum temperature and its square, and hence set

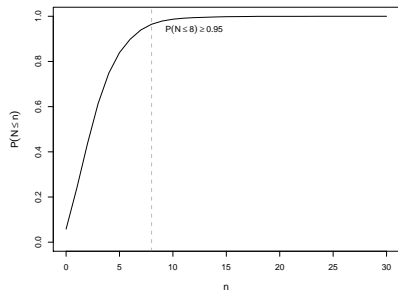
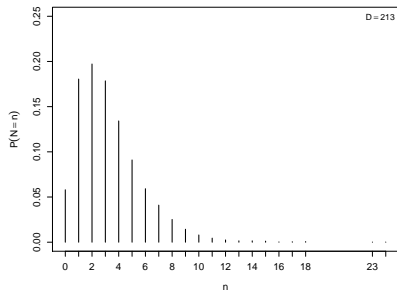
$$\log\left(\frac{\eta_t}{1 - \eta_t}\right) = \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2,$$

with x_t the maximum recorded temperature on sampling occasion t and $\alpha_0, \alpha_1, \alpha_2$ the regression coefficients, which gives

$$p_{t-b} = \eta_t \left\{ \prod_{k=b}^{t-1} (1 - \eta_k) \right\}^{\mathbb{I}(t > b)},$$

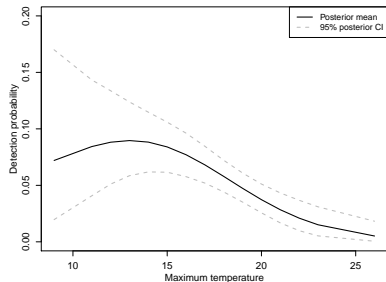
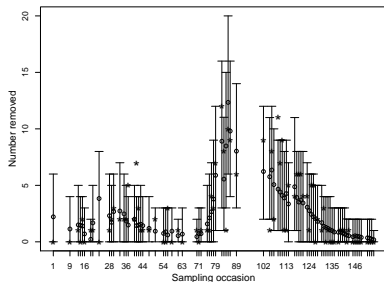
where the indicator variable $\mathbb{I}(t > b)$ is equal to 1 if $t > b$ and 0 otherwise.

COMMON LIZARDS



Common lizard data. (left) Posterior distribution, and (right) cumulative posterior distribution of N . Also shown in (right) is the value of N which corresponds to the 95% quantile of the posterior distribution.

COMMON LIZARDS



Common lizard data. (left) Assessment of model fit. The vertical bars show the width of the 95% quantile interval of values generated from 1000 randomly sampled parts of the chain, the circles show the mean simulated values and the stars the observed values. The gaps in the data correspond to days when sampling was not conducted and (right) Summaries of the posterior distribution of detection probability as a function of maximum daily recorded temperature. A logistic-quadratic function has been fitted.

GREAT CRESTED NEWTS

Detection probability assumed to vary with minimum air temperature at sampling occasion t , x_t , such that

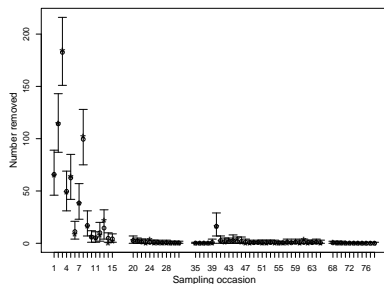
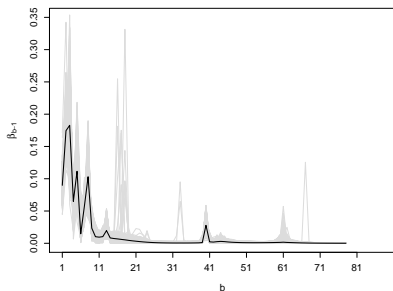
$$\log \left(\frac{\eta_t}{1 - \eta_t} \right) = \alpha_0 + \alpha_1 x_t.$$

Missing values for the covariate were imputed using the average of minimum temperatures recorded on the two adjacent sampling occasions.

In this case the number of renewal groups is unknown so we implement an RJMCMC algorithm for normal mixture models⁶

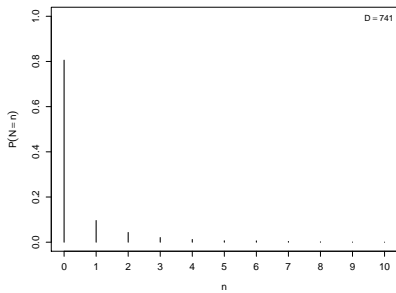
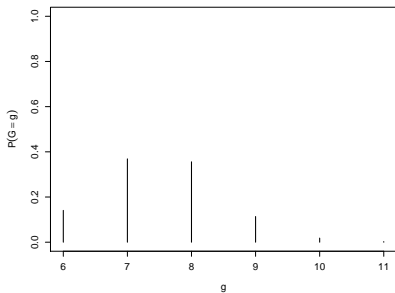
⁶Matechou, E., Nicholls, G., Morgan, B. J. T., Collazo, J. A. and Lyons, J. E. (2015). Bayesian analysis of Jolly-Seber type models; incorporating heterogeneity in arrival and departure. [arXiv:1512.05170](https://arxiv.org/abs/1512.05170).

GREAT CRESTED NEWTS



Male great crested newt data. (left) Estimated entry parameters obtained at each iteration of the RJMCMC algorithm, gray lines, together with the mean of all iterations, shown by the black line. The values on the x-axis correspond to sampling occasions. (right) Assessment of model fit. The observed counts, stars, lie within the 95% quantile intervals of counts simulated from the model, bars, using randomly selected iterations of the algorithm.

GREAT CRESTED NEWTS



Male great crested newt data. (left) Posterior distribution for the number of arrival groups, G (right), and the number of individuals remaining at the site after the end of the study, N .

REMOVAL DATA

MODELS

RESULTS

SUMMARY

FUTURE WORK

We have proposed a new model for the case of removal data when the assumption of population closure is violated because of:

- a) individuals emerging/arriving sometime after the start of the study, either through birth and/or immigration, and renewing the population and
- b) individuals departing from the study site before being detected through death and/or emigration.

The model allows for either a fixed known number of renewal groups or an unknown number, and estimates their sizes as well as the means and variances of arrival times of the individuals in these groups.

The model responds to a practical need, as the data described are commonly encountered.

REMOVAL DATA

MODELS

RESULTS

SUMMARY

FUTURE WORK

Incorporating heterogeneity in detection.

Currently working on providing a general framework and models for removal data together with R code.

Ming Zhou has been fitting models for data collected under a robust design using the HMM framework, allowing for temporary emigration.

New removal approaches for reptile and amphibian populations

M. Zhou, R. S. McCrea, E. Matechou, D. J. Cole & R. A. Griffiths
WEDNESDAY 29 JUNE 10.30-10.45, Session Code: D.5

Thank you!