

multimark: an R package for analysis of
capture-recapture data consisting of multiple
“non-invasive” marks

Brett T. McClintock¹

National Marine Mammal Laboratory

Alaska Fisheries Science Center
NOAA National Marine Fisheries Service
Seattle, Washington, U.S.A.

¹*Email*: brett.mcclintock@noaa.gov

RUNNING HEAD: **multimark** capture-recapture package

July 31, 2015

1 Summary

2 **1.** I describe an open source R package, **multimark**, for estimation of survival and
3 abundance from capture-mark-recapture data consisting of multiple “non-invasive”
4 marks. Non-invasive marks include natural pelt or skin patterns, scars, and genetic
5 markers that enable individual identification in lieu of physical capture. **multimark**
6 provides a means for combining and jointly analyzing encounter histories from mul-
7 tiple non-invasive sources that otherwise cannot be reliably matched (e.g. left- and
8 right-sided photos of bilaterally asymmetrical individuals).

9 **2.** **multimark** is currently capable of fitting open population Cormack-Jolly-Seber
10 (CJS) and closed population abundance models with up to two mark types using
11 Bayesian Markov chain Monte Carlo (MCMC) methods. **multimark** can also be
12 used for Bayesian analyses of conventional capture-recapture data consisting of a
13 single mark type.

14 **3.** Some package features include: (i) general model specification using formulas al-
15 ready familiar to most R users, (ii) ability to include temporal, behavioural, age, co-
16 hort, and individual heterogeneity effects in detection and survival probabilities, (iii)
17 improved MCMC algorithm that is computationally faster and more efficient than
18 previously proposed methods, (iv) Bayesian multimodel inference using reversible
19 jump MCMC, and (v) data simulation capabilities for power analyses and assessing
20 model performance.

21 **4.** I demonstrate use of **multimark** using left- and right-sided encounter histories for
22 bobcats (*Lynx rufus*) collected from remote single-camera stations in southern Cal-
23 ifornia. In this example, there is evidence of a behavioural effect (i.e. trap “happy”

response) that is otherwise indiscernible using conventional single-sided analyses.

5. The package will be most useful to ecologists seeking stronger inferences by combining different sources of mark-recapture data that are difficult (or impossible) to reliably reconcile, particularly with the sparse datasets typical of rare or elusive species for which non-invasive sampling techniques are most commonly employed. Addressing deficiencies in currently available software, **multimark** also provides a user-friendly interface for performing Bayesian multimodel inference using capture-recapture data consisting of a single conventional mark or multiple non-invasive marks.

Key-words Bayesian multimodel inference, capture-recapture, Cormack-Jolly-Seber (CJS), latent multinomial, mark-recapture, Markov chain Monte Carlo (MCMC), multiple lists, population size

1 Introduction

Capture-recapture methods historically relied on the physical capture, marking, and recapturing of animals for estimating population abundance and related demographic parameters such as survival (e.g. Williams *et al.* 2002). More recently, “non-invasive” capture-recapture sampling techniques are becoming commonplace for monitoring animal populations (e.g. Hammond 1990; Lukacs & Burnham 2005; O’Connell *et al.* 2010). Non-invasive marks can include natural pelt or skin patterns, scars, and genetic markers that enable individual identification in the absence of physical capture. Capture-recapture methods based on non-invasive marks have been applied to diverse taxa, including sharks (e.g. Holmberg *et al.* 2008), reptiles (e.g. Nair *et al.*

2012), ursids (e.g. Dreher *et al.* 2007), felids (e.g. Karanth & Nichols 1998; Ruell
et al. 2009), and marine mammals (e.g. Hammond 1990; Wilson *et al.* 1999; Madon
et al. 2011). While non-invasive capture-recapture methods have many advantages
related to financial cost and animal welfare, they also pose new difficulties such as
animal misidentification (Wright *et al.* 2009; Yoshizaki *et al.* 2009; Link *et al.* 2010;
Morrison *et al.* 2011) and the complexity of multiple types of marks (Corkrey *et al.*
2008; Madon *et al.* 2011; Bonner & Holmberg 2013; McClintock *et al.* 2013).

Multiple marks can arise from sighting or camera surveys of species with natu-
ral mark patterns that are bilaterally asymmetrical (e.g. cetaceans, felids) or from
multiple sources of non-invasive capture-recapture data being collected concurrently
(e.g. faecal DNA sampling and visual surveys). With multiple marks, an encounter
history is produced for each individual and mark type, but there is typically no re-
liable means to match them (unless each mark type is simultaneously observed at
least once for every encountered individual). Because the number of unique individ-
uals encountered must be known for standard capture-recapture analyses, the typical
approach is to conduct separate analyses for each mark type and compare the results
(e.g. Wilson *et al.* 1999; Berrow *et al.* 2012; Nair *et al.* 2012). However, given that
sample sizes (and precision) may be considerably reduced, this is not as efficient as
conducting an integrated analysis utilizing encounter histories arising from all mark
types (McClintock *et al.* 2013). Additional costs of conducting separate analyses
for each mark type include a limited ability to explore models with behavioural or
cohort effects, and, for capture-recapture models that condition on first encounter, a
forfeiting of information from histories with the (apparent) first encounter occurring
on the last sampling occasion. These limitations can be particularly important for

71 the sparse datasets typical of rare and elusive populations for which non-invasive
72 sampling techniques are most commonly employed.

73 Based on the latent multinomial model of Link *et al.* (2010), Bonner & Holmberg
74 (2013) and McClintock *et al.* (2013) recently developed methods for performing inte-
75 grated analyses of capture-recapture data consisting of multiple non-invasive marks.
76 However, to my knowledge, their approaches have yet to be applied by practitioners.
77 This is certainly not due to a lack of appropriate data (e.g. Wilson *et al.* 1999; Holm-
78 berg *et al.* 2008; Madon *et al.* 2011; Berrow *et al.* 2012; Nair *et al.* 2012), and is likely
79 attributable to the mathematical and computational complexity of the models, as
80 well as a lack of user-friendly software for implementing them. Generalized software
81 for performing Bayesian multimodel inference with capture-recapture data has also
82 been lacking, thereby leaving these procedures largely inaccessible to non-statisticians
83 (e.g. Brooks *et al.* 2000; Durban & Elston 2005; King & Brooks 2008; Royle 2008;
84 McClintock *et al.* 2013). These software needs were the motivation for **multimark**,
85 an R (R Core Team 2013) package for Bayesian analysis of capture-recapture data
86 consisting of multiple non-invasive marks.

87 After providing some additional background on capture-recapture with multiple
88 marks, I briefly describe the models implemented in **multimark**. These currently in-
89 clude open population Cormack-Jolly-Seber (CJS) and closed population abundance
90 models (e.g. Williams *et al.* 2002) with up to two mark types. Although originally
91 motivated by the challenges posed by multiple non-invasive marks, **multimark** can
92 also be used for analyses of conventional capture-recapture data consisting of a single
93 mark type. Using real and simulated data for illustration, I provide an overview of
94 the workflow for the package and a new analysis of left- and right-sided encounter

histories for bobcats (*Lynx rufus*) collected from remote single-camera stations in southern California. Additional information, including help files, data, examples, and package usage is available by downloading the `multimark` package from CRAN (<http://cran.r-project.org>) or github (<https://github.com/bmcclintock/multimark>). This article describes `multimark` version 1.3.0.

2 Description

2.1 Background

Capture-recapture data are typically represented by a collection of encounter histories $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, where each element of $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,T})$ indicates whether individual i was detected ($y_{i,t} = 1$) or not detected ($y_{i,t} = 0$) on each of $t = 1, \dots, T$ sampling occasions. Typical analyses then proceed by formulating a likelihood conditional on the n unique individuals encountered (e.g. Williams *et al.* 2002). With two mark types, we instead have $\tilde{\mathbf{Y}}_m = \{\tilde{\mathbf{y}}_{m_1}, \tilde{\mathbf{y}}_{m_2}, \dots, \tilde{\mathbf{y}}_{m_{n_m}}\}$ for $m \in \{1, 2\}$, where each element of $\tilde{\mathbf{y}}_{m_i} = (\tilde{y}_{m_i,1}, \tilde{y}_{m_i,2}, \dots, \tilde{y}_{m_i,T})$ indicates individual i was detected ($\tilde{y}_{m_i,t} = m$) or not detected ($\tilde{y}_{m_i,t} = 0$), and n_m is the number of unique individuals encountered for mark type m . We focus on situations where it is difficult (or impossible) to reliably match individuals from $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$. In this case, although we know $n \leq n_1 + n_2$, n is nevertheless unknown and standard capture-recapture analysis methods cannot be reliably used for simultaneous inference using both sources of data.

Depending on the mark types and sampling design, it may sometimes be possible to observe both marks simultaneously within a sampling occasion. In this case, some of the encounter histories from $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$ can be matched to unique indi-

viduals with certainty. For example, suppose images were collected during vessel-based line transect surveys of surfacing whales, where mark type 1 corresponds to patch patterns on the left side and mark type 2 corresponds to patterns on the right side. If an individual happens to be photographed on both sides simultaneously on at least one sampling occasion, then the true encounter history for this individual would be known (i.e. left- and right-sided images could be matched). This results in an additional set of n_{known} observed encounter histories, $\tilde{\mathbf{Y}}_{known} = \{\tilde{\mathbf{y}}_{known_1}, \tilde{\mathbf{y}}_{known_2}, \dots, \tilde{\mathbf{y}}_{known_{n_{known}}}\}$, consisting of histories that are known with certainty (Table 1).

In essence, **multimark** facilitates the joint analysis of type 1 ($\tilde{\mathbf{Y}}_1$), type 2 ($\tilde{\mathbf{Y}}_2$), and known encounter histories ($\tilde{\mathbf{Y}}_{known}$) while accounting for uncertainty in the number of unique individuals encountered using extensions of the methodology proposed by Bonner & Holmberg (2013) and McClintock *et al.* (2013). While the mathematical and computational details are generally of little interest to ecologists, **multimark** performs these operations in the background and requires only simple data formatting and model specification formulas familiar to most R users.

2.2 Models

multimark currently includes open population Cormack-Jolly-Seber (CJS) and closed population abundance models (e.g. Williams *et al.* 2002). These Bayesian implementations are similar in spirit to the CJS model of Royle (2008) and the abundance model of King *et al.* (2015). Given the latent encounter histories (\mathbf{Y}) that generated the observed encounter histories ($\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Y}}_{known}$), the likelihood for the CJS model

140 with two mark types is

$$[\mathbf{Y} \mid \mathbf{p}, \boldsymbol{\delta}, \alpha, \boldsymbol{\phi}, \mathbf{Q}] \propto \prod_{i=1}^n \prod_{t=C_i+1}^T \pi_{i,t} \quad (1)$$

$$\pi_{i,t} = \begin{cases} (1 - p_{i,t}) \phi_{i,t-1} q_{i,t} + (1 - \phi_{i,t-1}) (1 - q_{i,t}) & \text{if } y_{i,t} = 0 \text{ and } q_{i,t-1} = 1 \\ p_{i,t} \delta_1 \phi_{i,t-1} & \text{if } y_{i,t} = 1 \\ p_{i,t} \delta_2 \phi_{i,t-1} & \text{if } y_{i,t} = 2 \\ p_{i,t} (1 - \delta_1 - \delta_2) (1 - \alpha) \phi_{i,t-1} & \text{if } y_{i,t} = 3 \\ p_{i,t} (1 - \delta_1 - \delta_2) \alpha \phi_{i,t-1} & \text{if } y_{i,t} = 4 \\ 1 & \text{otherwise} \end{cases}$$

141 where $y_{i,t} = 0$ indicates a non-detection for individual i on occasion t , $y_{i,t} = 1$ in-
 142 dicates a type 1 encounter, $y_{i,t} = 2$ indicates a type 2 encounter, $y_{i,t} = 3$ indicates
 143 a non-simultaneous type 1 and type 2 encounter, $y_{i,t} = 4$ indicates a simultane-
 144 ous type 1 and type 2 encounter, $C_i \in \{1, \dots, T\}$ is the time of first capture for
 145 individual i , $p_{i,t}$ is the detection probability for individual i during sampling occa-
 146 sion t , δ_m is the conditional probability of a type m encounter (given detection),
 147 α is the conditional probability of a simultaneous type 1 and type 2 encounter
 148 (given both mark types detected), $\phi_{i,t-1}$ is the survival probability between times
 149 $t - 1$ and t , and $q_{i,t}$ is an indicator for whether individual i was alive ($q_{i,t} = 1$) or
 150 not ($q_{i,t} = 0$) during sampling occasion t . With $T = 3$, we for example have cell
 151 probabilities $\pi_i = \prod_{t=C_i+1}^T \pi_{i,t} = (1 - p_{i,3}) \phi_{i,2} q_{i,3} + (1 - \phi_{i,2}) (1 - q_{i,3})$ for latent en-
 152 counter history 020, $\pi_i = p_{i,2} \delta_1 \phi_{i,1} p_{i,3} \delta_2 \phi_{i,2}$ for latent encounter history 412, $\pi_i =$
 153 $(1 - p_{i,2}) \phi_{i,1} p_{i,3} (1 - \delta_1 - \delta_2) (1 - \alpha) \phi_{i,2}$ for history 103, and $p_{i,3} (1 - \delta_1 - \delta_2) \alpha \phi_{i,2}$
 154 for history 034.

For added flexibility, p and ϕ are modeled using the probit link function:

$$\Phi(p_{i,t}) = \mathbf{x}_t^{p'} \boldsymbol{\beta}^p + z_i^p$$

$$\Phi(\phi_{i,t}) = \mathbf{x}_t^{\phi'} \boldsymbol{\beta}^{\phi} + z_i^{\phi}$$

where $\Phi()$ the cumulative distribution function of the standard normal density, \mathbf{x}_t^p and \mathbf{x}_t^{ϕ} are row t of the design matrices for p and ϕ , $\boldsymbol{\beta}^p$ and $\boldsymbol{\beta}^{\phi}$ are the corresponding regression coefficients, and $z_i^p \sim \mathcal{N}(0, \sigma_{z^p}^2)$ and $z_i^{\phi} \sim \mathcal{N}(0, \sigma_{z^{\phi}}^2)$ are individual-level effects that respectively allow for individual heterogeneity in detection and survival probability. Thus, while exploring the feasible set of latent encounter histories (\mathbf{Y}), the parameters and latent variables to be estimated by `multimark` include $\boldsymbol{\beta}^p$, $\boldsymbol{\beta}^{\phi}$, $\boldsymbol{\delta}$, α , \mathbf{Q} , \mathbf{z}^p , \mathbf{z}^{ϕ} , $\sigma_{z^p}^2$, and $\sigma_{z^{\phi}}^2$.

The probit link is implemented for CJS models in `multimark` because it facilitates a Gibbs sampler in the spirit of Albert & Chib (1993) and Laake *et al.* (2013). The probit link is very similar to the logit link, but the logit link has slightly fatter tails and is interpretable in terms of log-odds. I note that this model reduces to that of Laake *et al.* (2013) for conventional capture-recapture data with a single mark type when $\delta_1 = 1$ and $\delta_2 = 0$ for $y_{i,t} \in \{0, 1\}$.

Similarly, the likelihood for the closed population abundance model with two mark types is

$$[\mathbf{Y} \mid \mathbf{p}, \boldsymbol{\delta}, \alpha, N] \propto \frac{1}{(p^*)^n} \prod_{i=1}^n \prod_{t=1}^T \pi_{i,t} \times \text{Binomial}(n; N, p^*) \quad (2)$$

$$\pi_{i,t} = \begin{cases} (1 - p_{i,t}) & \text{if } y_{i,t} = 0 \\ p_{i,t} \delta_1 & \text{if } y_{i,t} = 1 \\ p_{i,t} \delta_2 & \text{if } y_{i,t} = 2 \\ p_{i,t} (1 - \delta_1 - \delta_2) (1 - \alpha) & \text{if } y_{i,t} = 3 \\ p_{i,t} (1 - \delta_1 - \delta_2) \alpha & \text{if } y_{i,t} = 4 \\ 1 & \text{otherwise} \end{cases}$$

where N is the population size, and p^* is the probability that a randomly selected

individual is detected at least once. Returning to Table 1 with $T = 2$, we for example have cell probabilities $\pi_i = \prod_{t=1}^T \pi_{i,t} = (1 - p_{i,1}) p_{i,2} \delta_1$ for latent encounter history 01, $\pi_i = p_{i,1} \delta_2 (1 - p_{i,2})$ for history 20, $\pi_i = p_{i,1} \delta_2 p_{i,2} (1 - \delta_1 - \delta_2) (1 - \alpha)$ for history 23, and $\pi_i = p_{i,1} (1 - \delta_1 - \delta_2) \alpha p_{i,2} (1 - \delta_1 - \delta_2) (1 - \alpha)$ for history 43. As before, this model reduces to that for conventional capture-recapture data with a single mark type when $\delta_1 = 1$ and $\delta_2 = 0$ for $y_{i,t} \in \{0, 1\}$.

For closed population models, p is modeled using the logit link function:

$$\text{logit}(p_{i,t}) = \mathbf{x}_t^{p'} \boldsymbol{\beta}^p + z_i^p$$

such that

$$p^* = 1 - \int_{-\infty}^{\infty} \prod_{t=1}^T \left(1 - \frac{1}{1 + \exp(-(\mathbf{x}_t^{p'} \boldsymbol{\beta}^p + z^p))} \right) \mathcal{N}(z^p; 0, \sigma_{z^p}^2) dz^p$$

is the probability of being detected at least once after accounting for individual heterogeneity in p (note that $p^* = 1 - \prod_{t=1}^T (1 - \text{logit}^{-1}(\mathbf{x}_t^{p'} \boldsymbol{\beta}^p))$ when $\sigma_{z^p}^2 = 0$). The parameters and latent variables to be estimated therefore include $\boldsymbol{\beta}^p$, $\boldsymbol{\delta}$, α , N , \mathbf{z}^p , and $\sigma_{z^p}^2$. Although a Gibbs sampler has been proposed for closed population models using the probit link and a complete data likelihood (McClintock *et al.* 2014), this does not apply to the “semi-complete” data likelihood in Eq. 2 (hence the traditional logit link is used). The primary utility of **multimark** is finding the set of latent encounter histories that are feasible given the observed encounter histories (sensu Link *et al.* 2010; Bonner & Holmberg 2013; McClintock *et al.* 2013, 2014). Given a feasible set of latent encounter histories (\mathbf{Y}), fitting capture-recapture models such as Eqs. 1 or 2 is relatively straightforward.

188 3 Workflow

189 3.1 Multiple non-invasive marks

190 3.1.1 Data formatting

191 There are three types of multiple-mark data that can be analyzed with `multimark`.
192 These are the “never”, “sometimes”, and “always” data types, and they are named
193 based on their respective probabilities of a simultaneous type 1 and type 2 encounter
194 (Table 2). An example of the “never” data type is provided with `multimark` and
195 includes 23 left-sided ($\tilde{\mathbf{Y}}_1$) and 23 right-sided ($\tilde{\mathbf{Y}}_2$) encounter histories for bobcats
196 (*Lynx rufus*) collected from remote single-camera stations in southern California over
197 $T = 8$ sampling periods between July 2006 and January 2007 (McClintock *et al.* 2013;
198 Alonso *et al.* 2015).

199 `multimark` expects observed encounter history data to be a matrix with rows
200 corresponding to individuals and columns corresponding to sampling occasions. Be-
201 cause the bobcat data were collected from single-camera stations, simultaneous left-
202 and right-sided encounters were not possible; hence $\alpha = 0$ and the rows consist of
203 either 0’s and 1’s or 0’s and 2’s:

```
> library(multimark)
> data(bobcat)
> head(bobcat)
```

	occ1	occ2	occ3	occ4	occ5	occ6	occ7	occ8
ID2	0	0	0	0	0	1	1	0
ID3	0	0	1	0	1	0	0	0
ID4	0	0	0	0	1	0	0	0
ID6	1	0	0	0	0	0	0	0
ID7	0	0	1	0	0	0	0	1
ID8	0	1	0	0	0	0	0	0

```
> tail(bobcat)
```

	occ1	occ2	occ3	occ4	occ5	occ6	occ7	occ8
ID49	0	0	2	0	0	0	0	0
ID50	0	0	2	0	0	0	0	0
ID51	0	0	0	2	0	0	0	0
ID52	0	0	0	0	2	0	0	0
ID53	0	0	0	0	0	2	0	0
ID54	0	0	0	0	0	0	2	0

204 The ordering of the rows is unimportant; the package automatically recognizes which
 205 histories belong to $\tilde{\mathbf{Y}}_1$, $\tilde{\mathbf{Y}}_2$, and, if applicable, $\tilde{\mathbf{Y}}_{known}$.

206 The `multimark` function `processdata()` performs all additional data formatting.
 207 The basic inputs are the matrix of observed encounter histories (`Enc.Mat`) and the
 208 data type (`data.type`):

```
> bobcatsetup <- processdata(Enc.Mat=bobcat,data.type="never")
```

209 This creates an object of class `multimarksetup` that includes everything needed for
 210 model fitting and further analysis. In particular, `processdata()` calculates all of the
 211 necessary ingredients for identifying the feasible set of latent encounter histories (for
 212 technical details see Bonner & Holmberg 2013; McClintock *et al.* 2013). There is
 213 also a feature enabling designation of individual encounter histories as known with
 214 certainty despite no simultaneous type 1 and type 2 detections (i.e. $y_{i,t} \neq 4 \forall t$),
 215 a situation that can arise from a previous physical capture or concurrent telemetry
 216 study (e.g. McClintock *et al.* 2013).

217 3.1.2 Model fitting

218 The package currently includes functions `multimarkCJS()` and `multimarkClosed()`
 219 for fitting CJS and closed population models, respectively, with two mark types.
 220 Use of these functions is perhaps best explained by example. To fit a simple closed
 221 population model assuming constant detection probability using the default settings:

```
> bobcat.dot <- multimarkClosed(mms=bobcatsetup,
+                               mod.p=~1)
```

222 Equivalently, *Enc.Mat* and *data.type* can be provided in lieu of the *mms* argument.

223 In this case, *processdata()* is called from within *multimarkClosed()*:

```
> bobcat.dot <- multimarkClosed(Enc.Mat=bobcat,data.type="never",
+                               mod.p=~1)
```

224 This creates a list, *bobcat.dot*, containing the MCMC output for the model

225 (*bobcat.dot\$mcmc*). The MCMC output is of class *mcmc*, which should be familiar

226 to users of the R package *coda* (Plummer *et al.* 2006):

```
> summary(bobcat.dot$mcmc)
```

```
Iterations = 2001:12000
```

```
Thinning interval = 1
```

```
Number of chains = 1
```

```
Sample size per chain = 10000
```

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
pbeta[(Intercept)]	-1.3302	0.23847	0.0023847	0.012847
N	35.6166	5.20282	0.0520282	0.277289
delta_1	0.3949	0.07296	0.0007296	0.007221
delta_2	0.4112	0.07269	0.0007269	0.006086

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-1.7987	-1.4982	-1.3330	-1.1614	-0.8783
N	28.0000	32.0000	35.0000	39.0000	48.0000
delta_1	0.2540	0.3444	0.3940	0.4457	0.5360
delta_2	0.2707	0.3605	0.4113	0.4611	0.5524

```
> coda::effectiveSize(bobcat.dot$mcmc)
```

<code>pbeta[(Intercept)]</code>	<code>N</code>	<code>delta_1</code>	<code>delta_2</code>
344.5661	352.0563	102.0897	142.6469

227 Here we can see posterior summaries for the default monitored parameters (β^p , N , δ_1 , δ_2).

228 Based on the effective sample sizes, it's clear that the default chain length is inadequate for this example; a typical "rule of thumb" is effective sample sizes > 4000 for
 229 all quantities of interest.
 230

231 Other common models for detection probability can be easily specified using linear
 232 model formulas for *mod.p*, including shorthands for time variation (*mod.p*= $\sim time$),
 233 temporal trends (*mod.p*= $\sim Time$), behavioural response to first capture (*mod.p*= $\sim c$),
 234 and individual heterogeneity (*mod.p*= $\sim h$). Additive or interaction terms can be included (e.g. *mod.p*= $\sim time+c+h$, *mod.p*= $\sim Time+I(Time^2)$, *mod.p*= $\sim time*c$). User-
 235 specified temporal covariates in detection probability can also be used:
 236

```

> dummy <- rnorm(ncol(bobcat)) # some fake temporal covariates
> bobcatsetup <- processdata(Enc.Mat=bobcat,data.type="never",
+                           covs=data.frame(cov1=dummy))
> bobcat.dummy_h <- multmarkClosed(mms=bobcatsetup,
+                                 mod.p= $\sim cov1+h$ ,
+                                 parms=c("pbeta","N","delta","sigma2_zp"))

```

237 The *covs* argument is a data frame used to enter discrete- or continuous-valued
 238 temporal covariates, and *parms* specifies the parameters to monitor.

239 There are currently two options for specifying models for the conditional probabilities of type 1 and type 2 encounters (δ), the default *mod.delta*= $\sim type$ (i.e. $\delta_1 \neq \delta_2$),
 240 and *mod.delta*= ~ 1 (i.e. $\delta_1 = \delta_2$). The constraint $\delta_1 = \delta_2$ will often be reasonable
 241 when type 1 and type 2 encounters arise from a very similar process, such as with
 242 left- and right-sided images (see *Example*) . However, when type 1 and type 2 encounters arise from very different processes (e.g. faecal DNA and visual surveys),
 244

245 then specifying $\delta_1 \neq \delta_2$ is likely a model deserving consideration.

246 There are many additional arguments for specifying the number (*nchains*) and
 247 length (*iter*) of chains, including burn-in and adaptive periods. For potential im-
 248 provements in mixing, the number of “moves” used to update the feasible set of
 249 latent encounter histories at each iteration can be user specified (*maxnumbasis*; see
 250 Appendix S1). The default priors are “uninformative”, but user-specified priors can
 251 be used for each parameter. Initial values can be automatically generated or user
 252 specified for each parameter.

253 The function *multimarkCJS()* works in exactly the same fashion, with the only
 254 notable difference being specification of models for ϕ (in addition to p and δ). Al-
 255 though CJS-specific data are not included with *multimark*, data can be simulated
 256 using the *simdataCJS()* function (or *simdataClosed()* for closed populations):

```
> CJSdata <- simdataCJS(N=100,noccas=7,pbeta=-0.25,phibeta=1,delta_1=0.2,
+                       delta_2=0.5,alpha=0.5,sigma2_zphi=0.25,data.type="sometimes")
> Enc.Mat <- CJSdata$Enc.Mat
> head(Enc.Mat)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	1	0	0	0	0	0	0
[2,]	1	0	0	0	0	0	0
[3,]	0	0	1	0	0	0	0
[4,]	2	0	2	3	4	0	2
[5,]	4	1	0	0	0	0	0
[6,]	4	3	0	0	0	0	0

```
> CJSsetup <- processdata(Enc.Mat=Enc.Mat,data.type="sometimes")
> CJS.dot.h <- multimarkCJS(mms=CJSsetup,
+                           mod.p=~1,mod.delta=~type,mod.phi=~h,
+                           parms=c("pbeta","delta","alpha","phibeta","sigma2_zphi"),
+                           nchains=2,iter=45000,burnin=5000)
> summary(CJS.dot.h$mcmc)
```

```

Iterations = 5001:45000
Thinning interval = 1
Number of chains = 2
Sample size per chain = 40000

```

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
pbeta[(Intercept)]	-0.23929	0.12483	0.0004413	0.0034456
phibeta[(Intercept)]	1.39411	0.30023	0.0010615	0.0123494
alpha	0.51730	0.11716	0.0004142	0.0026694
sigma2_zphi	0.04021	0.09187	0.0003248	0.0039210
delta_1	0.20576	0.04557	0.0001611	0.0010992
delta_2	0.59324	0.05317	0.0001880	0.0009821

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-0.471868	-0.326159	-0.24406	-0.15725	0.01808
phibeta[(Intercept)]	0.916923	1.186706	1.35644	1.55790	2.08917
alpha	0.295278	0.434658	0.51570	0.59789	0.74769
sigma2_zphi	0.002618	0.007179	0.01413	0.03391	0.26382
delta_1	0.122343	0.173677	0.20383	0.23541	0.30069
delta_2	0.486954	0.557618	0.59420	0.62968	0.69478

257 An additional feature for *multimarkCJS()* is simple specification of “age” and cohort
 258 effects for p ($mod.p = \sim age$ and $mod.p = \sim cohort$) and ϕ ($mod.phi = \sim age$ and $mod.phi = \sim cohort$),
 259 which can be useful for investigating structure related to time since first capture and
 260 time of initial capture, respectively. These variables by default include a level for
 261 each unique age or cohort, but they can be binned to reduce the number of levels
 262 using additional arguments.

263 3.2 Single mark type

264 For conventional capture-recapture data consisting of a single mark type, encounter
265 histories are formatted the same way but now consist solely of 1's (detections) and
266 0's (non-detections). The package currently includes the functions *markCJS()* and
267 *markClosed()* for fitting conventional CJS and closed population models, respec-
268 tively. These functions are essentially wrappers that “trick” *multimarkCJS()* and
269 *multimarkClosed()* to fit models with a single mark type. The functions *simdat-*
270 *aCJS()* and *simdataClosed()* can also be used to simulate encounter history data
271 with a single mark type by setting the arguments *delta_1=1* and *delta_2=0*. For
272 example, to simulate CJS data and fit a model with constant detection probability
273 and individual heterogeneity in survival:

```
> singleCJSdata <- simdataCJS(delta_1=1,delta_2=0,  
+                             pbeta=-0.25,phibeta=1,sigma2_zphi=0.25)  
> Enc.Mat <- singleCJSdata$Enc.Mat  
> singleCJS.dot.h <- markCJS(Enc.Mat=Enc.Mat,mod.p=~1,mod.phi=~h,  
+                             parms=c("pbeta","phibeta","sigma2_zphi"),  
+                             nchains=2,iter=45000,burnin=5000)
```

274 There are fewer arguments for *markCJS()* and *markClosed()* because there is only
275 one mark type (e.g. the arguments *mms* and *mod.delta* are no longer necessary),
276 but the remaining arguments are specified exactly as for *multimarkCJS()* and *mul-*
277 *timarkClosed()*.

278 3.3 Further analysis

279 While the *coda* package can be used to summarize, plot, and assess convergence
280 of MCMC samples from *markClosed()*, *multimarkClosed()*, *markCJS()*, and *multi-*
281 *markCJS()*, several additional functions are available for further analysis of model

282 output. Because link functions are used for p and ϕ , the functions *getprobsClosed()*
 283 and *getprobsCJS()* provide estimates on the real scale. For example, we can com-
 284 pare the probabilities of capture (p) and recapture (c) when there is a behavioural
 285 response to first capture (i.e. $\text{mod.p}=\sim c$):

```
> bobcat.c <- multimarkClosed(mms=bobcatsetup,mod.p=~c)
> pc <- getprobsClosed(bobcat.c)
> summary(pc[,c("p[1]", "c[2]")])
```

```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000
```

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
p[1]	0.137	0.05407	0.0005407	0.005286
c[2]	0.259	0.05148	0.0005148	0.002981

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
p[1]	0.0485	0.09735	0.1321	0.1716	0.2538
c[2]	0.1619	0.22273	0.2569	0.2931	0.3639

286 Here, p[1] and c[2] refer to the probabilities of capture and recapture at times $t = 1$
 287 and $t = 2$, respectively.

288 Based on Barker & Link (2013), Bayesian multimodel inference using reversible
 289 jump MCMC is implemented through the functions *markClosed()*, *multimarkClosed()*,
 290 *markCJS()*, and *multimarkCJS()*. Using this approach, models are first run individ-
 291 ually and a Gibbs sampler explores the model space using the individual model
 292 MCMC output. All that must be provided to the multimodel inference functions is a

list containing the output from at least two models. The models must have the same number and length of MCMC chains, and all model parameters must be monitored (this is accomplished by setting *parms*="all"):

```
> bobcat.dot <- multimarkClosed(mms=bobcatsetup,mod.p=~1,parms="all")
> bobcat.c <- multimarkClosed(mms=bobcatsetup,mod.p=~c,parms="all")
> bobcat.time <- multimarkClosed(mms=bobcatsetup,mod.p=~time,parms="all")
> bobcat.h <- multimarkClosed(mms=bobcatsetup,mod.p=~h,parms="all")
> modlist <- list(mod1=bobcat.dot,mod2=bobcat.c,
+               mod3=bobcat.time,mod4=bobcat.h)
> bobcat.M <- multimodelClosed(modlist=modlist)
```

The list `bobcat.M` includes RJMCMC output (`bobcat.M$rjmc`) for parameters common to all models (which can be specified using the argument *monparms*) and posterior model probabilities (`bobcat.M$pos.prob`). Other arguments for *multimodelClosed()* and *multimodelCJS()* include prior model probabilities (*modprior*) and user-specified proposal distributions for moving between models.

4 Example

I will now provide results from a new closed population analysis of the bobcat data performed in `multimark`. Previous analyses of these data include McClintock *et al.* (2013), who performed an integrated analysis but for a limited model set that did not include behavioural or individual effects, and Alonso *et al.* (2015), who performed standard single-sided analyses that could not investigate behavioural responses to first capture. Using `multimark`, it is possible to conduct a more complete analysis using both left- and right-sided encounter histories that includes no effects, temporal effects, behavioural effects, and individual effects in detection probability. I also investigated two models for δ ($\delta_1 \neq \delta_2$ and $\delta_1 = \delta_2$) because it is reasonable to

311 suspect that the conditional probabilities of left-sided (type 1) and right-sided (type
312 2) encounters are similar.

313 Fitting all possible additive combinations yielded 16 models using the default
314 “non-informative” priors for *multimarkClosed()*:

$$\begin{aligned}\beta^p &\sim \mathcal{N}(0, 1.75) \\ \boldsymbol{\delta} &\sim \begin{cases} \text{Beta}(1, 1) & \text{if } \delta_1 = \delta_2 \\ \text{Dirichlet}(1, 1, 1) & \text{if } \delta_1 \neq \delta_2 \end{cases} \\ z_i^p &\sim \mathcal{N}(0, \sigma_{z^p}^2) \\ \sigma_{z^p} &\sim \text{half-Cauchy}(25) \\ N &\propto \frac{1}{N}\end{aligned}$$

315 With 2 chains each consisting of 2 million iterations (with thinning every 20 iterations
316 to reduce memory requirements), the simplest models required 12 mins on a computer
317 running Windows 7 (3.4GHz Intel Core i7, 16GB RAM), while the more complicated
318 models including time variation required at most 2 hrs. These relatively fast run times
319 are attributable to **multimark**’s parallel processing of MCMC algorithms written in
320 the C programming language (Kernighan & Ritchie 1988). Bayesian multimodel
321 inference was performed with *multimodelClosed()* using the default equal prior model
322 weights, where 300000 iterations for each chain required 2.6 hrs. The longer run time
323 for *multimodelClosed()* owes to the number of models and the RJMCMC algorithm
324 being written entirely in R.

325 Models including a positive behavioural response to first capture accounted for
326 0.51 of the posterior model weight, while models including $\delta_1 = \delta_2$ accounted for
327 0.78 of posterior model weight (Table 3). Model-averaged posterior modes were $N =$

328 35 (Highest Posterior Density Interval: 26-101; Fig. 1) for population abundance,
 329 $p = 0.15$ (HPDI: 0.04-0.27) for capture probability, and $c = 0.21$ (HPDI: 0.07-0.33)
 330 for recapture probability. With $\delta_1 = \delta_2 = 0.41$ (HPDI: 0.30-0.50) based on the
 331 model with the highest posterior probability, both-sided encounters were relatively
 332 infrequent for these data ($1 - \delta_1 - \delta_2 = 0.18$; HPDI: 0.00-0.39).

333 For comparison, I performed conventional left- and right-sided analyses for these
 334 data using *markClosed()* and *multimodelClosed()*. Because models for δ and be-
 335 havioural response do not apply, the candidate model set was limited to *mod.p=~1*,
 336 *mod.p=~time*, *mod.p=~h*, and *mod.p=~time+h* for these single-sided analyses. As
 337 before, the default “non-informative” priors were used, and the length and number
 338 of chains, burn-in periods, and adaptive periods were also the same. For the left-side
 339 analysis, the constant detection probability model accounted for 0.95 of the posterior
 340 model weight, while the individual heterogeneity model accounted for 0.04 of pos-
 341 terior model weight. Model-averaged posterior modes were $N = 32$ (HPDI: 24-52)
 342 for population abundance and $p = 0.12$ (HPDI: 0.07-0.19) for capture probability.
 343 For the right-side analysis, the constant detection probability model accounted for
 344 0.6 of the posterior model weight and the individual heterogeneity model accounted
 345 for 0.39 of posterior model weight. Model-averaged posterior modes were $N = 33$
 346 (HPDI: 23-85) for population abundance and $p = 0.12$ (HPDI: 0.04-0.19) for capture
 347 probability. These conflicting results demonstrate the unenviable position one can of-
 348 ten find oneself when conducting separate analyses for different mark types from the
 349 same population. One may be tempted to choose the “most precise” estimate based
 350 on the left-side analysis, but the integrated analysis suggests this would considerably
 351 underestimate the uncertainty about N . Choosing the “more conservative” right-

sided results or averaging the N estimates from the left- and right-sided analyses would also underestimate the uncertainty about N based on the integrated analysis. This discrepancy is likely attributable to the potential behavioural response to first capture identified by the integrated analysis.

5 Discussion

I have described some of the key features of **multimark**, a new R package for the analysis of capture-recapture data consisting of a single conventional mark or multiple non-invasive marks. The package currently includes open population CJS and closed population models, with functions for derived parameters (e.g. ϕ , p) and multimodel inference. It adds to the growing toolbox of freely-available software for the analysis of non-spatial (e.g. White & Burnham 1999; Choquet *et al.* 2009; Laake 2013; Laake *et al.* 2013) and spatial (e.g. Gopalaswamy *et al.* 2012; Efford 2015) capture-recapture data, but it is the first to combine otherwise irreconcilable encounter histories arising from multiple mark types. Although initially developed for integrated analyses of left- and right-sided images for bilaterally asymmetrical species, the package can be used to jointly analyze data arising from any two types of marks. For example, **multimark** could be used to integrate an analysis of encounter histories arising from genetic (e.g. hair or faecal) and visual (e.g. photo-ID) detections (*sensu* Madon *et al.* 2011; but see Bonner 2013). **multimark** is also the first capture-recapture software to implement generalized Bayesian multimodel inference based on the RJMCMC algorithm proposed by Barker & Link (2013).

Relative to previous applications using multiple marks (Bonner & Holmberg 2013; McClintock *et al.* 2013), the relatively fast computation times of the package are at-

375 tributable to its use of “semi-complete” data likelihoods (King *et al.* 2015), parallel
376 processing, and MCMC algorithms written in C (instead of R). Because parallel pro-
377 cessing relies on the **parallel** package (R Core Team 2013), first-time Windows and
378 OS X users can expect a firewall pop-up dialog box asking if an R process should
379 accept incoming connections. Memory requirements are minimized by condition-
380 ing on the observed encounter histories when identifying the feasible set of latent
381 encounter histories. To facilitate better mixing, **multimark** improves the MCMC
382 algorithms proposed by Bonner & Holmberg (2013) and McClintock *et al.* (2013,
383 2014) by avoiding latent encounter history proposals with negative frequencies in a
384 manner that requires no proposal tuning (see Appendix S1 for details).

385 Many potentially desirable extensions to **multimark** are possible. These include
386 a broader suite of capture-recapture models, such as multi-state and robust design
387 models (e.g. Williams *et al.* 2002). In addition to individual-level heterogeneity,
388 “random effect” distributions for temporal or user-specified covariates could also be
389 incorporated (e.g. Laake *et al.* 2013). More general modelling formulae for δ and α
390 would allow additional hypotheses related to detection to be explored. The package
391 could also be extended to accomodate >2 mark types and additional link functions.
392 Although many individual covariates tend to be difficult (or impossible) to observe
393 with non-invasive sampling, some (e.g. sex) may be easily discernable for each mark
394 type. For these cases, it would be relatively straightforward to extend **multimark**
395 to accommodate individual covariates. Other extensions include spatially-explicit
396 models (e.g. Royle 2015) and allowing for partial overlap in the sampling periods
397 for each mark type. Practitioners interested in such extensions are encouraged to
398 contact the author.

399 Acknowledgments

400 I thank Associate Editor B. Anholt, G. Péron, and E. Cooch for reviewing the
401 manuscript, P. Boveng, P. Conn, and J. Laake for helpful comments on earlier
402 drafts, and D. Johnson for helpful discussions. The findings and conclusions in the
403 manuscript are those of the author(s) and do not necessarily represent the views of
404 the National Marine Fisheries Service, NOAA. Any use of trade, product, or firm
405 names does not imply an endorsement by the US Government.

406 References

- 407 Albert, J.H. & Chib, S. (1993) Bayesian analysis of binary and polychotomous re-
408 sponse data. *Journal of the American Statistical Association*, **88**, 669–679.
- 409 Alonso, R.S., McClintock, B.T., Lyren, L.M., Boydston, E.E. & Crooks, K.R. (2015)
410 Mark-recapture and mark-resight methods for estimating abundance with remote
411 cameras: a carnivore case study. *PLoS ONE*, **10**, e0123032.
- 412 Barker, R.J. & Link, W.A. (2013) Bayesian multimodel inference by RJMCMC: A
413 Gibbs sampling approach. *The American Statistician*, **67**, 150–156.
- 414 Berrow, S., O’Brien, J., Groth, L., Foley, A. & Voigt, K. (2012) Abundance estimate
415 of bottlenose dolphins (*tursiops truncatus*) in the Lower River Shannon candidate
416 Special Area of Conservation, Ireland. *Aquatic Mammals*, **38**, 136–144.
- 417 Bonner, S.J. & Holmberg, J. (2013) Mark-recapture with multiple, non-invasive
418 marks. *Biometrics*, **69**, 766–775.
- 419 Bonner, S. (2013) Response to: a new method for estimating animal abundance

420 with two sources of data in capture–recapture studies. *Methods in Ecology and*
421 *Evolution*, **4**, 585–588.

422 Brooks, S.P., Catchpole, E.A. & Morgan, B.J.T. (2000) Bayesian animal survival
423 estimation. *Statistical Science*, **15**, 357–376.

424 Choquet, R., Rouan, L. & Pradel, R. (2009) Program E-SURGE: a software applica-
425 tion for fitting multievent models. D.L. Thomson, E.G. Cooch & M.J. Conroy, eds.,
426 *Modeling demographic processes in marked populations*, pp. 845–865. Springer.

427 Corkrey, R., Brooks, S., Lusseau, D., Parsons, K., Durban, J.W., Hammond, P.S.
428 & Thompson, P.M. (2008) A Bayesian capture-recapture population model with
429 simultaneous estimation of heterogeneity. *Journal of the American Statistical As-*
430 *sociation*, **103**, 948–960.

431 Dreher, B.P., Winterstein, S.R., Scribner, K.T., Lukacs, P.M., Etter, D.R., Rosa,
432 G.J.M., Lopez, V.A., Libants, S. & Filcek, K.B. (2007) Noninvasive estimation
433 of black bear abundance incorporating genotyping errors and harvested bears.
434 *Journal of Wildlife Management*, **71**, 2684–2693.

435 Durban, J.W. & Elston, D.A. (2005) Mark-recapture with occasion and individual
436 effects: abundance estimation through Bayesian model selection in a fixed dimen-
437 sional parameter space. *Journal of Agricultural, Biological, and Environmental*
438 *Statistics*, **10**, 291–305.

439 Efford, M. (2015) *secr: Spatially explicit capture-recapture models*. R package version
440 2.9.4.

- 441 Gopalaswamy, A.M., Royle, J.A., Hines, J.E., Singh, P., Jathanna, D., Kumar, N.
 442 & Karanth, K.U. (2012) Program SPACECAP: software for estimating animal
 443 density using spatially explicit capture–recapture models. *Methods in Ecology and*
 444 *Evolution*, **3**, 1067–1072.
- 445 Hammond, P. (1990) Capturing whales on film—estimating cetacean population pa-
 446 rameters from individual recognition data. *Mammal Review*, **20**, 17–22.
- 447 Holmberg, J., Norman, B. & Arzoumanian, Z. (2008) Robust, comparable population
 448 metrics through collaborative photo-monitoring of whale sharks *rhincodon typus*.
 449 *Ecological Applications*, **18**, 222–233.
- 450 Karanth, K.U. & Nichols, J.D. (1998) Estimation of tiger densities in India using
 451 photographic captures and recaptures. *Ecology*, **79**, 2852–2862.
- 452 Kernighan, B.W. & Ritchie, D.M. (1988) *The C Programming Language*. Prentice
 453 Hall, Englewood Cliffs, New Jersey, USA, second edition.
- 454 King, R. & Brooks, S.P. (2008) On the Bayesian estimation of a closed population size
 455 in the presence of heterogeneity and model uncertainty. *Biometrics*, **64**, 816–824.
- 456 King, R., McClintock, B.T., Kidney, D. & Borchers, D.L. (2015) Capture-recapture
 457 abundance estimation using a semi-complete data likelihood approach. *The Annals*
 458 *of Applied Statistics*.
- 459 Laake, J.L. (2013) *RMark: An R Interface for Analysis of Capture-Recapture Data*
 460 *with MARK*. AFSC Processed Rep 2013-01, 25p. Alaska Fish. Sci. Cent., NOAA,
 461 Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

- Laake, J.L., Johnson, D.S. & Conn, P.B. (2013) marked: an R package for maximum likelihood and markov chain monte carlo analysis of capture–recapture data. *Methods in Ecology and Evolution*, **4**, 885–890.
- Link, W.A., Yoshizaki, J., Bailey, L.L. & Pollock, K.H. (2010) Uncovering a latent multinomial: analysis of mark-recapture data with misidentification. *Biometrics*, **66**, 178–185.
- Lukacs, P.M. & Burnham, K.P. (2005) Review of capture–recapture methods applicable to noninvasive genetic sampling. *Molecular ecology*, **14**, 3909–3919.
- Madon, B., Gimenez, O., McArdle, B., Scott Baker, C. & Garrigue, C. (2011) A new method for estimating animal abundance with two sources of data in capture–recapture studies. *Methods in Ecology and Evolution*, **2**, 390–400.
- McClintock, B.T., Bailey, L.L., Dreher, B.P. & Link, W.A. (2014) Probit models for capture-recapture data subject to imperfect detection, individual heterogeneity and misidentification. *The Annals of Applied Statistics*, **8**, 2461–2484.
- McClintock, B.T., Conn, P.B., Alonso, R.S. & Crooks, K.R. (2013) Integrated modeling of bilateral photo-identification data in mark-recapture analyses. *Ecology*, **94**, 1464–1471.
- Morrison, T.A., Yoshizaki, J., Nichols, J.D. & Bolger, D.T. (2011) Estimating survival in photographic capture-recapture studies: overcoming misidentification error. *Methods in Ecology and Evolution*, **2**, 454–463.
- Nair, T., Thorbjarnarson, J.B., Aust, P. & Krishnaswamy, J. (2012) Rigorous gharial population estimation in the Chambal: implications for conservation and man-

484 agement of a globally threatened crocodilian. *Journal of Applied Ecology*, **49**,
485 1046–1054.

486 O’Connell, A.F., Nichols, J.D. & Karanth, K.U. (2010) *Camera traps in animal*
487 *ecology: methods and analyses*. Springer Science & Business Media.

488 Plummer, M., Best, N., Cowles, K. & Vines, K. (2006) CODA: convergence diagnosis
489 and output analysis for MCMC. *R News*, **6**, 7–11.

490 R Core Team (2013) *R: A Language and Environment for Statistical Comput-*
491 *ing*. R Foundation for Statistical Computing, Vienna, Austria. [http://www.R-](http://www.R-project.org/)
492 [project.org/](http://www.R-project.org/).

493 Royle, J.A. (2008) Modeling individual effects in the Cormack-Jolly-Seber model: a
494 state-space formulation. *Biometrics*, **64**, 364–370.

495 Royle, J.A. (2015) Spatial capture-recapture with partial identity. *arXiv preprint*
496 *arXiv:150306873*.

497 Ruell, E.W., Riley, S.P.D., Douglas, M.R., Pollinger, J.P. & Crooks, K.R. (2009)
498 Estimating bobcat population sizes and densities in a fragmented urban landscape
499 using noninvasive capture-recapture sampling. *Journal of Mammalogy*, **90**, 129–
500 135.

501 White, G.C. & Burnham, K.P. (1999) Program MARK: Survival estimation from
502 populations of marked animals. *Bird Study*, **46**, S120–S138.

503 Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of*
504 *Animal Populations*. Academic Press, San Diego, CA, USA.

- 505 Wilson, B., Hammond, P.S. & Thompson, P.M. (1999) Estimating size and assessing
506 trends in a coastal bottlenose dolphin population. *Ecological Applications*, **9**, 288–
507 300.
- 508 Wright, J.A., Barker, R.J., Schofield, M.R., Frantz, A.C., Byrom, A.E. & Gleeson,
509 D.M. (2009) Incorporating genotype uncertainty into mark-recapture-type models
510 for estimating abundance using DNA samples. *Biometrics*, **65**, 833–840.
- 511 Yoshizaki, J., Pollock, K.H., Brownie, C. & Webster, R.A. (2009) Modeling misiden-
512 tification errors in capture-recapture studies using photographic identification of
513 evolving marks. *Ecology*, **90**, 3–9.

Table 1. Latent encounter histories \mathbf{y} and the recorded histories $(\tilde{\mathbf{y}}_1, \tilde{\mathbf{y}}_2, \tilde{\mathbf{y}}_{known})$ they generate for $T = 2$ sampling occasions and two mark types, where $\mathbf{y} = (y_1, y_2)$ for $y_t \in \{0, 1, 2, 3, 4\}$. Latent encounter histories are indexed by $j = 1 + \sum_{t=1}^T y_t 5^{T-t}$, where the encounter types indicate non-detection ($y_t = 0$), type 1 encounter ($y_t = 1$), type 2 encounter ($y_t = 2$), non-simultaneous type 1 and type 2 encounter ($y_t = 3$), and simultaneous type 1 and type 2 encounter ($y_t = 4$). If simultaneous encounters are possible, these result in some \mathbf{y} being completely observable (as indicated by $\tilde{\mathbf{y}}_{known}$).

j	\mathbf{y}	$\tilde{\mathbf{y}}_1$	$\tilde{\mathbf{y}}_2$	$\tilde{\mathbf{y}}_{known}$
1	00
2	01	01
3	02	..	02	..
4	03	01	02	..
5	04	04
6	10	10
7	11	11
8	12	10	02	..
9	13	11	02	..
10	14	14
11	20	..	20	..
12	21	01	20	..
13	22	..	22	..
14	23	01	22	..
15	24	24
16	30	10	20	..
17	31	11	20	..
18	32	10	22	..
19	33	11	22	..
20	34	34
21	40	40
22	41	41
23	42	42
24	43	43
25	44	44

Table 2. Summary of three different types of multiple-mark data. The data differ in terms of the latent encounter types (y_t) that are possible based on the conditional probability of a simultaneous type 1 and type 2 encounter, $\alpha = \Pr(y_t = 4 | y_t = 3 \text{ or } y_t = 4)$.

Data type	y_t	Constraints
“never”	$\{0, 1, 2, 3\}$	$\alpha = 0$
“sometimes”	$\{0, 1, 2, 3, 4\}$	$0 < \alpha < 1$
“always”	$\{0, 1, 2, 4\}$	$\alpha = 1$

Model	PMM	N	HPDI	ESS	GRB
p($\sim c$)delta(~ 1)	0.30	38	27-91	38944	1.00
p(~ 1)delta(~ 1)	0.22	33	26-46	54696	1.00
p($\sim h$)delta(~ 1)	0.16	46	29-114	11685	1.00
p($\sim c + h$)delta(~ 1)	0.09	50	29-145	18544	1.00
p($\sim c$)delta($\sim \text{type}$)	0.09	38	27-90	35054	1.00
p(~ 1)delta($\sim \text{type}$)	0.06	33	26-46	53961	1.00
p($\sim h$)delta($\sim \text{type}$)	0.05	48	29-113	12099	1.00
p($\sim c + h$)delta($\sim \text{type}$)	0.03	51	29-146	17276	1.00
p($\sim \text{time} + h$)delta(~ 1)	0.00	47	28-115	14414	1.00
p($\sim c + \text{time} + h$)delta(~ 1)	0.00	45	28-116	21473	1.00
p($\sim \text{time}$)delta(~ 1)	0.00	33	26-45	47781	1.00
p($\sim c + \text{time}$)delta(~ 1)	0.00	33	25-78	35169	1.00
p($\sim \text{time} + h$)delta($\sim \text{type}$)	0.00	50	29-118	13882	1.00
p($\sim c + \text{time} + h$)delta($\sim \text{type}$)	0.00	46	27-115	21337	1.00
p($\sim \text{time}$)delta($\sim \text{type}$)	0.00	33	26-45	49425	1.00
p($\sim c + \text{time}$)delta($\sim \text{type}$)	0.00	32	25-78	35360	1.00

Table 3. Posterior model probabilities (PMM) and abundance estimates for the bobcat data. Summaries include posterior modes (N), 95% highest posterior density intervals (HPDI), effective sample sizes (ESS), and Gelman-Rubin-Brooks diagnostics (GRB) for N . Models for detection probability (p) included no effects (~ 1), behavioural effects ($\sim c$), time effects ($\sim \text{time}$), and individual effects ($\sim h$). Models for the conditional probability of a left- or right-sided encounter (delta) included $\delta_1 = \delta_2$ (~ 1) and $\delta_1 \neq \delta_2$ ($\sim \text{type}$).

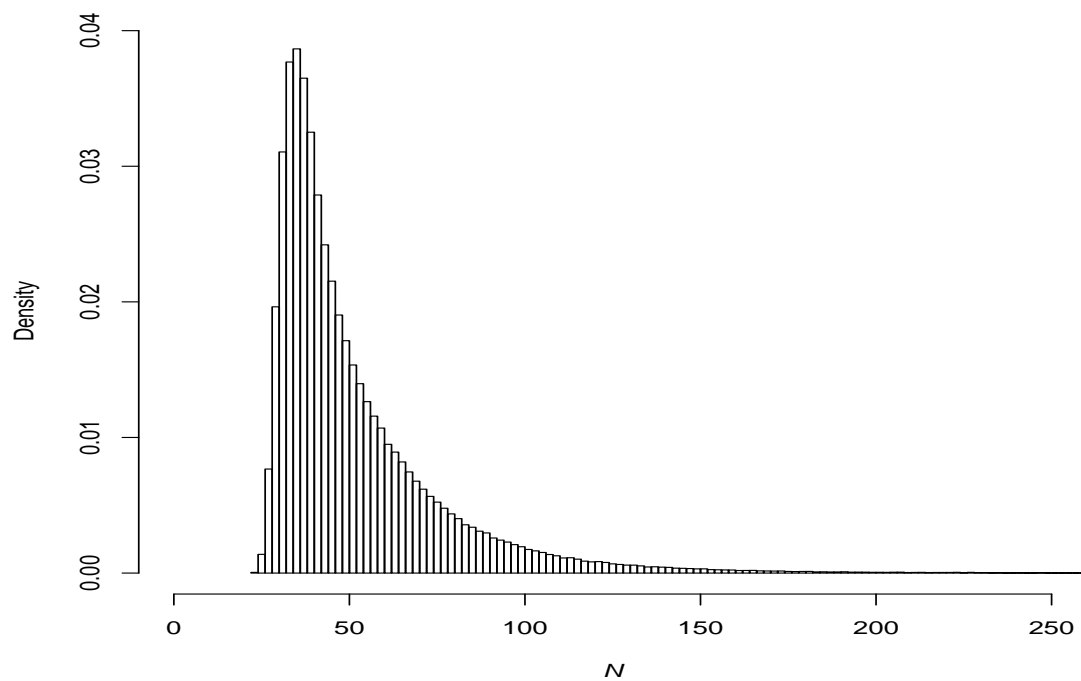


Figure 1. Model-averaged posterior distribution of population abundance (N) for the bobcat data.