

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** mark-recapture package

July 21, 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 ge-
5 netic markers that enable individual identification in lieu of physical capture, and
6 thus apply to any species that can be individually identified from visual or genetic
7 sampling surveys. **multimark** provides a means for combining and jointly analyz-
8 ing encounter histories from multiple non-invasive sources that otherwise cannot be
9 reliably matched (e.g. left- and right-sided photos of bilaterally asymmetrical indi-
10 viduals).

11 **2.** **multimark** is currently capable of fitting open population Cormack-Jolly-Seber
12 (CJS) and closed population abundance models with up to two mark types using
13 Bayesian Markov chain Monte Carlo (MCMC) methods. Although originally moti-
14 vated by the challenges posed by multiple non-invasive marks, **multimark** can also
15 be used for Bayesian analyses of conventional capture-recapture data consisting of a
16 single mark type.

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

24 **4.** I demonstrate use of `multimark` using left- and right-sided encounter histories for
25 bobcats (*Lynx rufus*) collected from remote single-camera stations in southern Cal-
26 ifornia. In this example, there is evidence of a behavioural effect (i.e. trap “happy”
27 response) that is otherwise indiscernible using traditional single-sided analyses.

28 **5.** The package will be most useful to ecologists seeking stronger inferences by
29 combining different sources of mark-recapture data that are difficult (or impossi-
30 ble) to reliably reconcile, particularly with the sparse datasets typical of rare or
31 elusive species for which non-invasive sampling techniques are most commonly em-
32 ployed. Addressing deficiencies in currently available software, `multimark` also pro-
33 vides a user-friendly interface for performing Bayesian multimodel inference using
34 capture-recapture data consisting of a single conventional mark or multiple non-
35 invasive marks.

36

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

40 **1 Introduction**

41 Capture-recapture methods historically relied on the physical capture, marking, and
42 recapturing of animals for estimating population abundance and related demographic
43 parameters such as survival (e.g. Williams *et al.* 2002). More recently, “non-invasive”
44 capture-recapture sampling techniques are becoming commonplace for monitoring
45 animal populations (e.g. Hammond 1990; Lukacs & Burnham 2005; O’Connell *et al.*
46 2010). Non-invasive marks can include natural pelt or skin patterns, scars, and

47 genetic markers that enable individual identification in the absence of physical cap-
48 ture. Capture-recapture methods based on non-invasive marks have been applied to
49 diverse taxa, including sharks (e.g. Holmberg *et al.* 2008), reptiles (e.g. Nair *et al.*
50 2012), ursids (e.g. Dreher *et al.* 2007), felids (e.g. Karanth & Nichols 1998; Ruell
51 *et al.* 2009), and marine mammals (e.g. Hammond 1990; Wilson *et al.* 1999; Madon
52 *et al.* 2011). While non-invasive capture-recapture methods have many advantages
53 related to financial cost and animal welfare, they also pose new difficulties such as
54 animal misidentification (Wright *et al.* 2009; Yoshizaki *et al.* 2009; Link *et al.* 2010;
55 Morrison *et al.* 2011) and the complexity of multiple types of marks (Corkrey *et al.*
56 2008; Madon *et al.* 2011; Bonner & Holmberg 2013; McClintock *et al.* 2013).

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

71 cohort effects, and, for capture-recapture models that condition on first encounter, a
72 forfeiting of information from histories with the (apparent) first encounter occurring
73 on the last sampling occasion. These limitations can be particularly important for
74 the sparse datasets typical of rare and elusive populations for which non-invasive
75 sampling techniques are most commonly employed.

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

90 After providing some additional background on capture-recapture with multiple
91 marks, I briefly describe the models implemented in **multimark**. These currently in-
92 clude open population Cormack-Jolly-Seber (CJS) and closed population abundance
93 models (e.g. Williams *et al.* 2002) with up to two mark types. Although originally
94 motivated by the challenges posed by multiple non-invasive marks, **multimark** can

also be used for analyses of conventional capture-recapture data consisting of a single mark type. Using real and simulated data for illustration, I provide an overview of 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 individuals 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

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

141 model of King *et al.* (2015). Given the latent encounter histories (\mathbf{Y}) that generated
 142 the observed encounter histories ($\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Y}}_{known}$), the likelihood for the CJS model
 143 with two mark types is

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

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

144 where $y_{i,t} = 0$ indicates a non-detection for individual i on occasion t , $y_{i,t} = 1$
 145 indicates a type 1 encounter, $y_{i,t} = 2$ indicates a type 2 encounter, $y_{i,t} = 3$ indi-
 146 cates a non-simultaneous type 1 and type 2 encounter, $y_{i,t} = 4$ indicates a simul-
 147 taneous type 1 and type 2 encounter, $C_i \in \{1, \dots, T\}$ is the time of first capture
 148 for individual i , $p_{i,t}$ is the detection probability for individual i during sampling
 149 occasion t , δ_m is the conditional probability of a type m encounter (given detec-
 150 tion), α is the conditional probability of a simultaneous type 1 and type 2 encounter
 151 (given both mark types detected), $\phi_{i,t-1}$ is the survival probability between times
 152 $t - 1$ and t , and $q_{i,t}$ is an indicator for whether individual i was alive ($q_{i,t} = 1$) or
 153 not ($q_{i,t} = 0$) during sampling occasion t . With $T = 3$, we for example have cell
 154 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-
 155 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, $\pi_i =$
 156 $(1 - p_{i,2}) \phi_{i,1} p_{i,3} (1 - \delta_1 - \delta_2) (1 - \alpha) \phi_{i,2}$ for history 103, and $p_{i,3} (1 - \delta_1 - \delta_2) \alpha \phi_{i,2}$
 157 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 both 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. 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 $y_{i,t} \in \{0, 1\}$. 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, $\pi_i = p_{i,1}\delta_2p_{i,2}(1 - \delta_1 - \delta_2)(1 - \alpha)$ 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.

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

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”	$\{0, 1, 2, 4\}$	$\alpha = 1$

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.

3 Workflow

3.1 Multiple non-invasive marks

3.1.1 Data formatting

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

`multimark` expects observed encounter history data to be a matrix with rows

203 corresponding to individuals and columns corresponding to sampling occasions. Be-
 204 cause the bobcat data were collected from single-camera stations, simultaneous left-
 205 and right-sided encounters were not possible; hence $\alpha = 0$ and the rows consist of
 206 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

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

209 The `multimark` function `processdata()` performs all additional data formatting.
 210 The basic inputs are the matrix of observed encounter histories (*Enc.Mat*) and the
 211 data type (*data.type*):

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

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

220 3.1.2 Model fitting

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

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

225 Equivalently, *Enc.Mat* and *data.type* can be provided in lieu of the *mms* argument.
 226 In this case, *processdata()* is called from within *multimarkClosed()*:

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

227 This creates a list, *bobcat.dot*, containing the MCMC output for the model
 228 (*bobcat.dot\$mcmc*). The MCMC output is of class *mcmc*, which should be familiar
 229 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)]	-1.3963	0.24837	0.0024837	0.01503
N	36.0808	5.78436	0.0578436	0.32252
delta_1	0.3677	0.08244	0.0008244	0.01217
delta_2	0.3867	0.08183	0.0008183	0.01172

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-1.9000	-1.5561	-1.3904	-1.2259	-0.9256
N	27.9750	32.0000	35.0000	39.0000	50.0000
delta_1	0.2127	0.3085	0.3661	0.4251	0.5308
delta_2	0.2293	0.3293	0.3857	0.4433	0.5457

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

pbeta[(Intercept)]	N	delta_1	delta_2
272.90838	321.66283	45.85059	48.75330

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

231 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
232 all quantities of interest.

234 Other common models for detection probability can be easily specified using linear
235 model formulas for *mod.p*, including shorthands for time variation (*mod.p*=*~time*),
236 temporal trends (*mod.p*=*~Time*), behavioural response to first capture (*mod.p*=*~c*),
237 and individual heterogeneity (*mod.p*=*~h*). Additive or interaction terms can be in-

cluded (e.g. $mod.p = \sim time + c + h$, $mod.p = \sim Time + I(Time^2)$, $mod.p = \sim time * c$). User-specified temporal covariates in detection probability can also be used:

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

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

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. $\delta_1 \neq \delta_2$), and $mod.delta = \sim 1$ (i.e. $\delta_1 = \delta_2$). The constraint $\delta_1 = \delta_2$ will often be reasonable when type 1 and type 2 encounters arise from a very similar process, such as with 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), then specifying $\delta_1 \neq \delta_2$ is likely a model deserving consideration.

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

The function *multimarkCJS()* works in exactly the same fashion, with the only notable difference being specification of models for ϕ (in addition to p and δ). Al-

258 though CJS-specific data are not included with `multimark`, data can be simulated

259 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,]     2     0     4     0     0     0     0
[2,]     1     0     0     0     0     0     0
[3,]     1     0     0     0     0     0     0
[4,]     4     3     0     0     0     0     0
[5,]     1     0     0     0     0     0     0
[6,]     3     2     0     0     2     2     4
```

```
> 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)]	-0.15716	0.12916	0.0004566	0.004168
phibeta[(Intercept)]	1.36677	0.27045	0.0009562	0.010230
alpha	0.41452	0.11586	0.0004096	0.002907
sigma2_zphi	0.02932	0.05608	0.0001983	0.002294
delta_1	0.19614	0.04581	0.0001620	0.001601
delta_2	0.61126	0.05195	0.0001837	0.001180

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-0.397592	-0.247046	-0.16172	-0.07123	0.1054
phibeta[(Intercept)]	0.922360	1.175204	1.33609	1.52690	1.9784
alpha	0.207287	0.331264	0.40812	0.49030	0.6577
sigma2_zphi	0.002495	0.006458	0.01266	0.02846	0.1662
delta_1	0.112033	0.163847	0.19424	0.22626	0.2905
delta_2	0.506597	0.576504	0.61231	0.64717	0.7098

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

266 **3.2 Single mark type**

267 For conventional capture-recapture data consisting of a single mark type, encounter
 268 histories are formatted the same way but now consist solely of 1’s (detections) and
 269 0’s (non-detections). The package currently includes the functions *markCJS()* and
 270 *markClosed()* for fitting conventional CJS and closed population models, respec-
 271 tively. These functions are essentially wrappers that “trick” *multimarkCJS()* and
 272 *multimarkClosed()* to fit models with a single mark type. The functions *simdat-*
 273 *aCJS()* and *simdataClosed()* can also be used to simulate encounter history data
 274 with a single mark type by setting the arguments $delta_1=1$ and $delta_2=0$. For
 275 example, to simulate CJS data and fit a model with constant detection probability
 276 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)

```

277 There are fewer arguments for *markCJS()* and *markClosed()* because there is only
 278 one mark type (e.g. the arguments *mms* and *mod.delta* are no longer needed), but
 279 the remaining arguments are specified exactly as for *multimarkCJS()* and *multimark-*
 280 *Closed()*.

281 3.3 Further analysis

282 While the *coda* package can be used to summarize, plot, and assess convergence
 283 of MCMC samples from *markClosed()*, *multimarkClosed()*, *markCJS()*, and *multi-*
 284 *markCJS()*, several additional functions are available for further analysis of model
 285 output. Because link functions are used for p and ϕ , the functions *getprobsClosed()*
 286 and *getprobsCJS()* provide estimates on the real scale. For example, we can com-
 287 pare the probabilities of capture (p) and recapture (c) when there is a behavioural
 288 response to first capture (i.e. *mod.p*= 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]	0.1406	0.05481	0.0005481	0.004992
c[2]	0.2510	0.05481	0.0005481	0.004678

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
p[1]	0.05674	0.09906	0.1338	0.1750	0.2673
c[2]	0.15111	0.21231	0.2487	0.2876	0.3642

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

291 Based on Barker & Link (2013), Bayesian multimodel inference using reversible
 292 jump MCMC is implemented through the functions *markClosed()*, *multimarkClosed()*,
 293 *markCJS()*, and *multimarkCJS()*. Using this approach, models are first run individ-
 294 ually and a Gibbs sampler explores the model space using the individual model
 295 MCMC output. All that must be provided to the multimodel inference functions is a
 296 list containing the output from at least two models. The models must have the same
 297 number and length of MCMC chains, and all model parameters must be monitored
 298 (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)
```

299 The list bobcat.M includes RJMCMC output (bobcat.M\$rjmc) for parameters
 300 common to all models (which can be specified using the argument *monparms*) and

301 posterior model probabilities (`bobcat.M$pos.prob`). Other arguments for *multimod-*
 302 *elClosed()* and *multimodelCJS()* include prior model probabilities (*modprior*) and
 303 user-specified proposal distributions for moving between models.

304 4 Example

305 I will now provide results from a new closed population analysis of the bobcat data
 306 performed in `multimark`. Previous analyses of these data include McClintock *et al.*
 307 (2013), who performed an integrated analysis but for a limited model set that did not
 308 include behavioural or individual effects, and Alonso *et al.* (2015), who performed
 309 standard single-sided analyses that could not investigate behavioural responses to
 310 first capture. Using `multimark`, it is possible to conduct a more complete analysis
 311 using both left- and right-sided encounter histories that includes no effects, temporal
 312 effects, behavioural effects, and individual effects in detection probability. I also
 313 investigated two models for δ ($\delta_1 \neq \delta_2$ and $\delta_1 = \delta_2$) because it is reasonable to
 314 suspect that the conditional probabilities of left-sided (type 1) and right-sided (type
 315 2) encounters are similar.

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

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

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

328 Models including a positive behavioural response to first capture accounted for
329 0.51 of the posterior model weight, while models including $\delta_1 = \delta_2$ accounted for
330 0.78 of posterior model weight (Table 3). Model-averaged posterior modes were $N =$
331 35 (Highest Posterior Density Interval: 26-101; Fig. 1) for population abundance,
332 $p = 0.15$ (HPDI: 0.04-0.27) for capture probability, and $c = 0.21$ (HPDI: 0.07-0.33)
333 for recapture probability. With $\delta_1 = \delta_2 = 0.41$ (HPDI: 0.30-0.50) based on the
334 model with the highest posterior probability, both-sided encounters were relatively
335 infrequent for these data ($1 - \delta_1 - \delta_2 = 0.18$; HPDI: 0.00-0.39).

336 For comparison, I performed conventional left- and right-sided analyses for these
337 data using *markClosed()* and *multimodelClosed()*. Because models for δ and be-
338 havioural response do not apply, the candidate model set was limited to $mod.p = \sim 1$,
339 $mod.p = \sim time$, $mod.p = \sim h$, and $mod.p = \sim time + h$ for these single-sided analyses. As
340 before, the default "non-informative" priors were used, and the length and number
341 of chains, burn-in periods, and adaptive periods were also the same. For the left-side

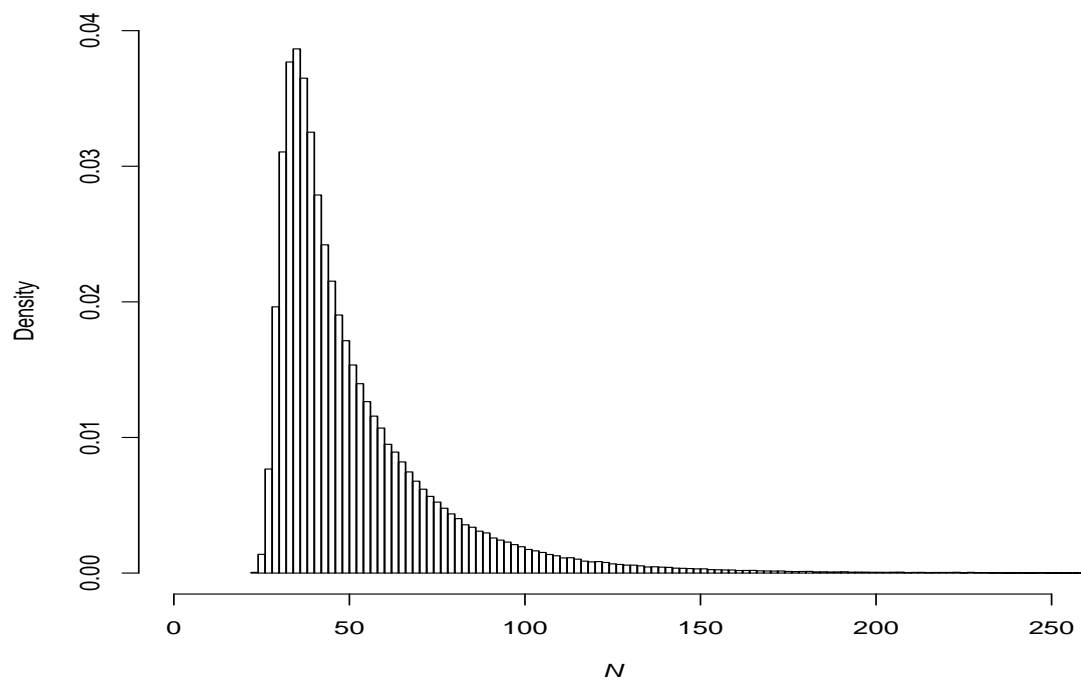


Figure 1. Model-averaged posterior distribution of population abundance (N) for the bobcat data.

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($\sim type$)	0.09	38	27-90	35054	1.00
p(~ 1)delta($\sim type$)	0.06	33	26-46	53961	1.00
p($\sim h$)delta($\sim type$)	0.05	48	29-113	12099	1.00
p($\sim c + h$)delta($\sim type$)	0.03	51	29-146	17276	1.00
p($\sim time + h$)delta(~ 1)	0.00	47	28-115	14414	1.00
p($\sim c + time + h$)delta(~ 1)	0.00	45	28-116	21473	1.00
p($\sim time$)delta(~ 1)	0.00	33	26-45	47781	1.00
p($\sim c + time$)delta(~ 1)	0.00	33	25-78	35169	1.00
p($\sim time + h$)delta($\sim type$)	0.00	50	29-118	13882	1.00
p($\sim c + time + h$)delta($\sim type$)	0.00	46	27-115	21337	1.00
p($\sim time$)delta($\sim type$)	0.00	33	26-45	49425	1.00
p($\sim c + time$)delta($\sim type$)	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 ($\sim time$), 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 $\delta_1 \neq \delta_2$ ($\sim type$).

analysis, the constant detection probability model accounted for 0.95 of the posterior model weight, while the individual heterogeneity model accounted for 0.04 of posterior model weight. Model-averaged posterior modes were $N = 32$ (HPDI: 24-52) for population abundance and $p = 0.12$ (HPDI: 0.07-0.19) for capture probability. For the right-side analysis, the constant detection probability model accounted for 0.6 of the posterior model weight and the individual heterogeneity model accounted for 0.39 of posterior model weight. Model-averaged posterior modes were $N = 33$ (HPDI: 23-85) for population abundance and $p = 0.12$ (HPDI: 0.04-0.19) for capture

350 probability. These conflicting results demonstrate the unenviable position one can of-
351 ten find oneself when conducting separate analyses for different mark types from the
352 same population. One may be tempted to choose the “most precise” estimate based
353 on the left-side analysis, but the integrated analysis suggests this would considerably
354 underestimate the uncertainty about N . Choosing the “more conservative” right-
355 sided results or averaging the N estimates from the left- and right-sided analyses
356 would also underestimate the uncertainty about N based on the integrated analysis.
357 This discrepancy is likely attributable to the potential behavioural response to first
358 capture identified by the integrated analysis.

359 5 Discussion

360 I have described some of the key features of **multimark**, a new R package for the
361 analysis of capture-recapture data consisting of a single conventional mark or multiple
362 non-invasive marks. The package currently includes open population CJS and closed
363 population models, with functions for derived parameters (e.g. ϕ , p) and multimodel
364 inference. It adds to the growing toolbox of freely-available software for the analysis
365 of non-spatial (e.g. White & Burnham 1999; Choquet *et al.* 2009; Laake 2013; Laake
366 *et al.* 2013) and spatial (e.g. Gopalaswamy *et al.* 2012; Efford 2015) capture-recapture
367 data, but it is the first to combine otherwise irreconcilable encounter histories arising
368 from multiple mark types. Although initially developed for integrated analyses of
369 left- and right-sided images for bilaterally asymmetrical species, the package can be
370 used to jointly analyze data arising from any two types of marks. For example,
371 **multimark** could be used to integrate an analysis of encounter histories arising from
372 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 attributable to its use of “semi-complete” data likelihoods (King *et al.* 2015), parallel processing, and MCMC algorithms written in C (instead of R). Because parallel processing relies on the **parallel** package (R Core Team 2013), first-time Windows and OS X users can expect a firewall pop-up dialog box asking if an R process should accept incoming connections. Memory requirements are minimized by conditioning on the observed encounter histories when identifying the feasible set of latent encounter histories. To facilitate better mixing, **multimark** improves the MCMC algorithms proposed by Bonner & Holmberg (2013) and McClintock *et al.* (2013, 2014) by avoiding latent encounter history proposals with negative frequencies in a manner that requires no proposal tuning (see Appendix S1 for details).

Many potentially desirable extensions to **multimark** are possible. These include a broader suite of capture-recapture models, such as multi-state and robust design models (e.g. Williams *et al.* 2002). In addition to individual-level heterogeneity, “random effect” distributions for temporal or user-specified covariates could also be incorporated (e.g. Laake *et al.* 2013). More general modelling formulae for δ and α would allow additional hypotheses related to detection to be explored. The package could also be extended to accomodate >2 mark types and additional link functions. Although many individual covariates tend to be difficult (or impossible) to observe with non-invasive sampling, some (e.g. sex) may be easily discernable for each mark

type. For these cases, it would be relatively straightforward to extend `multimark` to accommodate individual covariates. Other extensions include spatially-explicit models (e.g. Royle 2015) and allowing for partial overlap in the sampling periods for each mark type. Practitioners interested in such extensions are encouraged to contact the author.

Acknowledgments

P. Boveng, P. Conn, and J. Laake for helpful comments on earlier drafts. D. Johnson for helpful discussions. The findings and conclusions in the manuscript are those of the author(s) and do not necessarily represent the views of the National Marine Fisheries Service, NOAA. Any use of trade, product, or firm names does not imply an endorsement by the US Government.

References

- Albert, J.H. & Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669–679.
- Alonso, R.S., McClintock, B.T., Lyren, L.M., Boydston, E.E. & Crooks, K.R. (2015) Mark-recapture and mark-resight methods for estimating abundance with remote cameras: a carnivore case study. *PLoS ONE*, **10**, e0123032.
- Barker, R.J. & Link, W.A. (2013) Bayesian multimodel inference by RJMCMC: A Gibbs sampling approach. *The American Statistician*, **67**, 150–156.
- Berrow, S., O’Brien, J., Groth, L., Foley, A. & Voigt, K. (2012) Abundance estimate of bottlenose dolphins (*tursiops truncatus*) in the Lower River Shannon candidate Special Area of Conservation, Ireland. *Aquatic Mammals*, **38**, 136–144.

- 419 Bonner, S.J. & Holmberg, J. (2013) Mark-recapture with multiple, non-invasive
420 marks. *Biometrics*, **69**, 766–775.
- 421 Bonner, S. (2013) Response to: a new method for estimating animal abundance
422 with two sources of data in capture–recapture studies. *Methods in Ecology and*
423 *Evolution*, **4**, 585–588.
- 424 Brooks, S.P., Catchpole, E.A. & Morgan, B.J.T. (2000) Bayesian animal survival
425 estimation. *Statistical Science*, **15**, 357–376.
- 426 Choquet, R., Rouan, L. & Pradel, R. (2009) Program E-SURGE: a software applica-
427 tion for fitting multievent models. D.L. Thomson, E.G. Cooch & M.J. Conroy, eds.,
428 *Modeling demographic processes in marked populations*, pp. 845–865. Springer.
- 429 Corkrey, R., Brooks, S., Lusseau, D., Parsons, K., Durban, J.W., Hammond, P.S.
430 & Thompson, P.M. (2008) A Bayesian capture-recapture population model with
431 simultaneous estimation of heterogeneity. *Journal of the American Statistical As-*
432 *sociation*, **103**, 948–960.
- 433 Dreher, B.P., Winterstein, S.R., Scribner, K.T., Lukacs, P.M., Etter, D.R., Rosa,
434 G.J.M., Lopez, V.A., Libants, S. & Filcek, K.B. (2007) Noninvasive estimation
435 of black bear abundance incorporating genotyping errors and harvested bears.
436 *Journal of Wildlife Management*, **71**, 2684–2693.
- 437 Durban, J.W. & Elston, D.A. (2005) Mark-recapture with occasion and individual
438 effects: abundance estimation through Bayesian model selection in a fixed dimen-
439 sional parameter space. *Journal of Agricultural, Biological, and Environmental*
440 *Statistics*, **10**, 291–305.

- 441 Efford, M. (2015) *secr: Spatially explicit capture-recapture models*. R package version
442 2.9.4.
- 443 Gopalaswamy, A.M., Royle, J.A., Hines, J.E., Singh, P., Jathanna, D., Kumar, N.
444 & Karanth, K.U. (2012) Program SPACECAP: software for estimating animal
445 density using spatially explicit capture–recapture models. *Methods in Ecology and*
446 *Evolution*, **3**, 1067–1072.
- 447 Hammond, P. (1990) Capturing whales on film—estimating cetacean population pa-
448 rameters from individual recognition data. *Mammal Review*, **20**, 17–22.
- 449 Holmberg, J., Norman, B. & Arzoumanian, Z. (2008) Robust, comparable population
450 metrics through collaborative photo-monitoring of whale sharks *rhincodon typus*.
451 *Ecological Applications*, **18**, 222–233.
- 452 Karanth, K.U. & Nichols, J.D. (1998) Estimation of tiger densities in India using
453 photographic captures and recaptures. *Ecology*, **79**, 2852–2862.
- 454 Kernighan, B.W. & Ritchie, D.M. (1988) *The C Programming Language*. Prentice
455 Hall, Englewood Cliffs, New Jersey, USA, second edition.
- 456 King, R. & Brooks, S.P. (2008) On the Bayesian estimation of a closed population size
457 in the presence of heterogeneity and model uncertainty. *Biometrics*, **64**, 816–824.
- 458 King, R., McClintock, B.T., Kidney, D. & Borchers, D.L. (2015) Capture-recapture
459 abundance estimation using a semi-complete data likelihood approach. *The Annals*
460 *of Applied Statistics*.

- 461 Laake, J.L. (2013) *RMark: An R Interface for Analysis of Capture-Recapture Data*
 462 *with MARK*. AFSC Processed Rep 2013-01, 25p. Alaska Fish. Sci. Cent., NOAA,
 463 Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.
- 464 Laake, J.L., Johnson, D.S. & Conn, P.B. (2013) marked: an R package for maxi-
 465 mum likelihood and markov chain monte carlo analysis of capture-recapture data.
 466 *Methods in Ecology and Evolution*, **4**, 885–890.
- 467 Link, W.A., Yoshizaki, J., Bailey, L.L. & Pollock, K.H. (2010) Uncovering a latent
 468 multinomial: analysis of mark-recapture data with misidentification. *Biometrics*,
 469 **66**, 178–185.
- 470 Lukacs, P.M. & Burnham, K.P. (2005) Review of capture-recapture methods appli-
 471 cable to noninvasive genetic sampling. *Molecular ecology*, **14**, 3909–3919.
- 472 Madon, B., Gimenez, O., McArdle, B., Scott Baker, C. & Garrigue, C. (2011) A new
 473 method for estimating animal abundance with two sources of data in capture-
 474 recapture studies. *Methods in Ecology and Evolution*, **2**, 390–400.
- 475 McClintock, B.T., Bailey, L.L., Dreher, B.P. & Link, W.A. (2014) Probit models for
 476 capture-recapture data subject to imperfect detection, individual heterogeneity
 477 and misidentification. *The Annals of Applied Statistics*, **8**, 2461–2484.
- 478 McClintock, B.T., Conn, P.B., Alonso, R.S. & Crooks, K.R. (2013) Integrated mod-
 479 eling of bilateral photo-identification data in mark-recapture analyses. *Ecology*,
 480 **94**, 1464–1471.
- 481 Morrison, T.A., Yoshizaki, J., Nichols, J.D. & Bolger, D.T. (2011) Estimating sur-

482 vival in photographic capture-recapture studies: overcoming misidentification er-
 483 ror. *Methods in Ecology and Evolution*, **2**, 454–463.

484 Nair, T., Thorbjarnarson, J.B., Aust, P. & Krishnaswamy, J. (2012) Rigorous ghar-
 485 ial population estimation in the Chambal: implications for conservation and man-
 486 agement of a globally threatened crocodilian. *Journal of Applied Ecology*, **49**,
 487 1046–1054.

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

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

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

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

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

499 Ruell, E.W., Riley, S.P.D., Douglas, M.R., Pollinger, J.P. & Crooks, K.R. (2009)
 500 Estimating bobcat population sizes and densities in a fragmented urban landscape
 501 using noninvasive capture-recapture sampling. *Journal of Mammalogy*, **90**, 129–
 502 135.

- 503 White, G.C. & Burnham, K.P. (1999) Program MARK: Survival estimation from
504 populations of marked animals. *Bird Study*, **46**, S120–S138.
- 505 Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of*
506 *Animal Populations*. Academic Press, San Diego, CA, USA.
- 507 Wilson, B., Hammond, P.S. & Thompson, P.M. (1999) Estimating size and assessing
508 trends in a coastal bottlenose dolphin population. *Ecological Applications*, **9**, 288–
509 300.
- 510 Wright, J.A., Barker, R.J., Schofield, M.R., Frantz, A.C., Byrom, A.E. & Gleeson,
511 D.M. (2009) Incorporating genotype uncertainty into mark-recapture-type models
512 for estimating abundance using DNA samples. *Biometrics*, **65**, 833–840.
- 513 Yoshizaki, J., Pollock, K.H., Brownie, C. & Webster, R.A. (2009) Modeling misiden-
514 tification errors in capture-recapture studies using photographic identification of
515 evolving marks. *Ecology*, **90**, 3–9.