

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

March 27, 2015

Summary

1. I describe an open source R package, **multimark**, for estimation of survival and abundance from capture-mark-recapture data consisting of multiple non-invasive marks. Non-invasive marks include natural pelt or skin patterns, scars, and genetic markers, and thus apply to any species that can be individually identified from visual or genetic sampling surveys. **multimark** provides a means for combining and jointly analyzing encounter histories from multiple non-invasive sources that otherwise cannot be reliably matched (e.g., left- and right-sided photos of bilaterally asymmetrical individuals).
2. **multimark** is currently capable of fitting open population Cormack-Jolly-Seber (CJS) and closed population abundance models with two mark types using Bayesian Markov chain Monte Carlo (MCMC) methods.
3. Some package features include: (i) general model specification using formulas already familiar to most R users, (ii) ability to include temporal, behavioural, cohort, and individual heterogeneity effects in detection and survival probabilities, (iii) Bayesian multimodel inference using reversible jump MCMC, and (iv) data simulation capabilities for power analyses and assessing model performance.
4. I demonstrate use of **multimark** using left- and right-sided encounter histories for bobcats (*Lynx rufus*) collected from remote single-camera stations in southern California. In this example, there is evidence of a behavioural effect (i.e., trap “happy” response) that is otherwise indiscernible using traditional single-sided analyses.
5. The package will be most useful to ecologists seeking to combine different sources of mark-recapture data that are difficult (or impossible) to reliably reconcile, particu-

larly with the sparse datasets typical of rare or elusive species for which non-invasive sampling techniques are most commonly employed.

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

“Non-invasive” capture-recapture sampling techniques are becoming standard practice for monitoring animal populations (e.g., Hammond, 1990; Lukacs & Burnham, 2005; O’Connell *et al.*, 2010). Examples of non-invasive marks enabling individual identification include natural pelt or skin patterns, scars, and genetic markers. Capture-recapture based on non-invasive marks has 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 has many advantages related to financial cost and animal welfare, it also poses new difficulties such as animal misidentification (Wright *et al.*, 2009; Yoshizaki *et al.*, 2009; Link *et al.*, 2010; Morrison *et al.*, 2011) and the potential for 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 natural 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 reliable means to match them (unless each mark type is simultaneously observed at least once for every encountered individual). Because the number of unique individuals encountered must be known for standard capture-recapture analyses of population abundance (or related demographic parameters), 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 the sparse datasets typical of rare and elusive populations for which non-invasive sampling techniques are most commonly employed.

Bonner & Holmberg (2013) and McClintock *et al.* (2013) recently developed methods for performing integrated analyses of capture-recapture data consisting of multiple non-invasive marks. However, to my knowledge, their approaches have yet to be applied by practitioners. This is certainly not due to a lack of appropriate data (e.g., Wilson *et al.*, 1999; Holmberg *et al.*, 2008; Madon *et al.*, 2011; Berrow *et al.*, 2012; Nair *et al.*, 2012), and is likely attributable to the mathematical and computational complexity of the models, as well as a lack of user-friendly software for implementing

42 them. This is the motivation for **multimark**, an R (R Core Team, 2013) package for
 43 the analysis of capture-recapture data consisting of multiple non-invasive marks.

44 After providing some additional background on capture-recapture with multiple
 45 marks, I briefly describe the models implemented in **multimark**. These currently
 46 include open population Cormack-Jolly-Seber (CJS) and closed population abun-
 47 dance models (e.g., Williams *et al.*, 2002) with two mark types. Using real and
 48 simulated data for illustration, I provide an overview of the workflow for the pack-
 49 age and a new analysis of left- and right-sided encounter histories for bobcats (*Lynx*
 50 *rufus*) collected from remote single-camera stations in southern California. Addi-
 51 tional information, including help files, data, examples, and package usage is avail-
 52 able by downloading the **multimark** package from CRAN (<http://cran.r-project.org>)
 53 or github (<https://github.com/bmcclintock/multimark>).

54 **2 Description**

55 **2.1 Background**

56 Capture-recapture data are typically represented by a collection of encounter histo-
 57 ries $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, where each element of $\mathbf{y}_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,T}\}$ indicates
 58 whether individual i was detected ($y_{i,t} = 1$) or not detected ($y_{i,t} = 0$) on each
 59 of $t = 1, \dots, T$ sampling occasions. Typical analyses then proceed by formulat-
 60 ing a likelihood conditional on the n unique individuals encountered (e.g., Williams
 61 *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
 62 $m \in \{1, 2\}$, 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
 63 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
 64 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, it may sometimes be possible to observe both marks simultaneously. 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 complicated (and 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.

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

87 2.2 Models

88 `multimark` currently includes open population Cormack-Jolly-Seber (CJS) and closed
 89 population abundance models (e.g., Williams *et al.*, 2002). These Bayesian imple-
 90 mentations are similar in spirit to the CJS model of Royle (2008) and the abundance
 91 model of King *et al.* (2015). Given the latent encounter histories (\mathbf{Y}) that generated
 92 the observed encounter histories ($\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Y}}_{known}$), the likelihood for the CJS model
 93 with two mark types is

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

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

where $C_i \in \{1, \dots, T\}$ is the time of first capture for individual i , $p_{i,t}$ is the detection probability for individual i during sampling occasion t , δ_m is the conditional probability of a type m encounter (given detection), α is the conditional probability of a simultaneous type 1 and type 2 encounter (given both mark types detected), $\phi_{i,t-1}$ is the survival probability between times $t-1$ and t , and $z_{i,t}$ is an indicator for whether individual i was alive ($z_{i,t} = 1$) or not ($z_{i,t} = 0$) during sampling occasion t . For added flexibility, p and ϕ are modeled using the probit link function:

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

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

94 where $\Phi()$ the cumulative distribution function of the standard normal density, \mathbf{x}_t^p
 95 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
 96 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
 97 effects. The probit link is implemented for CJS models in `multimark` because it
 98 facilitates a Gibbs sampler in the spirit of Albert & Chib (1993) and Laake *et al.*
 99 (2013).

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

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

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

where N is the population size, and p^* is the probability that a randomly selected individual is detected at least once. For added flexibility, 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

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

102 Although a Gibbs sampler has been proposed for closed population models using the
 103 probit link and a complete data likelihood (McClintock *et al.*, 2014), this does not

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”	$\{0, 1, 2, 4\}$	$\alpha = 1$

104 apply to the “semi-complete” data likelihood in Eq. 2 (hence the traditional logit
 105 link is used). The primary utility of **multimark** is finding the set of latent encounter
 106 histories that are feasible given the observed encounter histories (sensu Link *et al.*,
 107 2010; Bonner & Holmberg, 2013; McClintock *et al.*, 2013, 2014). Given a feasible set
 108 of latent encounter histories, fitting capture-recapture models such as Eqs. 1 or 2 is
 109 relatively straightforward.

110 3 Workflow

111 3.1 Data formatting

112 There are three types of multiple-mark data that can be analyzed with **multimark**.
 113 These are the “never”, “sometimes”, and “always” data types, and they are named
 114 based on their respective probabilities of a simultaneous type 1 and type 2 encounter
 115 (Table 2). An example of the “never” data type is provided with **multimark** and
 116 includes 23 left-sided ($\tilde{\mathbf{Y}}_1$) and 23 right-sided ($\tilde{\mathbf{Y}}_2$) encounter histories for bobcats
 117 (*Lynx rufus*) collected from remote single-camera stations in southern California over
 118 $T = 8$ sampling periods between July 2006 and January 2007 (McClintock *et al.*,
 119 2013; Alonso *et al.*, 2015).

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

121 corresponding to individuals and columns corresponding to sampling occasions. Be-
 122 cause the bobcat data were collected from single-camera stations, simultaneous left-
 123 and right-sided encounters were not possible; hence $\alpha = 0$ and the rows consist of
 124 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

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

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

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

130 This creates an object of class *multimarksetup* that includes everything needed for
 131 model fitting and further analysis. In particular, *processdata()* calculates all of the
 132 necessary ingredients for identifying the feasible set of latent encounter histories (for
 133 technical details see Bonner & Holmberg, 2013; McClintock *et al.*, 2013). There is
 134 also a feature enabling designation of individual encounter histories as known with
 135 certainty despite no simultaneous type 1 and type 2 detections (i.e., $y_{i,t} \neq 4 \forall t$),
 136 a situation that can arise from a previous physical capture or concurrent telemetry
 137 study (e.g., McClintock *et al.*, 2013). This feature can also be used to “trick” the
 138 package to perform analyses with traditional capture-recapture data with a single
 139 mark type.

140 3.2 Model fitting

141 The package currently includes functions *multimarkCJS()* and *multimarkClosed()*
 142 for fitting CJS and closed population models, respectively. Use of these functions is
 143 perhaps best explained by example. To fit a simple closed population model assuming
 144 constant detection probability using the default settings:

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

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

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

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

147 This creates a list, *bobcat.dot*, containing the MCMC output for the model
 148 (*bobcat.dot\$mcmc*). The MCMC output is of class *mcmc*, which should be familiar
 149 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)]	-1.2590	0.27604	0.0027604	0.02085
N	35.3603	5.43211	0.0543211	0.31004
delta_1	0.4212	0.08117	0.0008117	0.01216
delta_2	0.4356	0.07996	0.0007996	0.01265

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-1.8326	-1.4465	-1.2487	-1.0620	-0.7535
N	27.0000	32.0000	34.0000	38.0000	48.0000
delta_1	0.2562	0.3673	0.4244	0.4782	0.5703
delta_2	0.2721	0.3818	0.4399	0.4920	0.5792

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

	N	delta_1	delta_2
pbeta[(Intercept)]	175.26399	306.97405	44.54464
			39.95661

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

151 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
152 all quantities of interest.

154 Other common models for detection probability can be easily specified using
155 formulas for $mod.p$, including shorthands for time variation ($mod.p = \sim time$), temporal trends ($mod.p = \sim Time$), behavioural response to first capture ($mod.p = \sim c$),
156

157 and individual heterogeneity ($mod.p \sim h$). Additive or interaction terms can be in-
 158 cluded (e.g., $mod.p \sim time + c + h$, $mod.p \sim Time + I(Time^2)$, $mod.p \sim time * c$). User-
 159 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 <- multmarkClosed(mms=bobcatsetup,
+                                 mod.p=~cov1+h,
+                                 parms=c("pbeta","N","delta","sigma2_zp"))
```

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

162 There are currently two options for specifying models for δ , the default $mod.delta \sim type$
 163 (i.e., $\delta_1 \neq \delta_2$), and $mod.delta \sim 1$ (i.e., $\delta_1 = \delta_2$). There are many additional argu-
 164 ments for specifying the number (*nchains*) and length (*iter*) of chains, including
 165 burn-in and adaptive periods. For potential improvements in mixing, the number of
 166 “moves” used to update the feasible set of latent encounter histories at each iteration
 167 can be user specified (*maxnumbasis*; see Appendix S1). The default priors are “un-
 168 informative”, but user-specified priors can be used for each parameter. Initial values
 169 can be automatically generated or user-specified for each parameter.

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

```
> CJSdata <- simdataCJS(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)]	-0.14505	0.13195	0.0004665	0.004521
phibeta[(Intercept)]	1.39905	0.28513	0.0010081	0.011372
alpha	0.40530	0.11295	0.0003993	0.002603
sigma2_zphi	0.03712	0.09867	0.0003489	0.004207
delta_1	0.20192	0.04616	0.0001632	0.001583
delta_2	0.60387	0.05157	0.0001823	0.001129

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-0.389604	-0.237291	-0.15002	-0.05742	0.1245
phibeta[(Intercept)]	0.940375	1.200721	1.36546	1.55676	2.0510
alpha	0.204653	0.324695	0.39904	0.47879	0.6438
sigma2_zphi	0.002649	0.007147	0.01393	0.03141	0.2105
delta_1	0.116826	0.169463	0.20025	0.23221	0.2971
delta_2	0.501039	0.569305	0.60480	0.63940	0.7023

174 An additional feature for *multimarkCJS()* is simple specification of cohort effects for
 175 p (*mod.p*= $\sim age$) and ϕ (*mod.phi*= $\sim age$), which is useful for investigating structure
 176 related to age (or time of first capture).

177 3.3 Further analysis

178 While the `coda` package can be used to summarize, plot, and assess convergence
 179 of MCMC samples from *multimarkClosed()* and *multimarkCJS()*, several additional
 180 functions are available for further analysis. Because link functions are used for p and
 181 ϕ , the functions *getprobsClosed()* and *getprobsCJS()* provide estimates on the real
 182 scale:

```
> 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]	0.1366	0.05521	0.0005521	0.006385
c[2]	0.2651	0.05115	0.0005115	0.003045

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
p[1]	0.04261	0.09569	0.1302	0.1716	0.2609
c[2]	0.17281	0.22827	0.2632	0.2989	0.3705

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

Based on Barker & Link (2013), Bayesian multimodel inference using reversible jump MCMC is implemented through the functions *multimodelClosed()* and *multimodelCJS()*. Using this approach, models are first run individually and a Gibbs sampler explores the model space using the individual model MCMC output. All that must be provided to the multimodel inference functions is an object of class *multimarksetup* and 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(mms=bobcatsetup,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

203 standard single-sided analyses that could not investigate behavioural effects to first
 204 capture. Using **multimark**, it is possible to conduct a more complete analysis using
 205 both left- and right-sided encounter histories that includes no effects, temporal effects,
 206 behavioural effects, and individual effects in detection probability. I also investigated
 207 two models for δ ($\delta_1 \neq \delta_2$ and $\delta_1 = \delta_2$) because it is reasonable to suspect that the
 208 conditional probabilities of left- and right-sided encounters are similar.

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

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

211 With 2 chains each consisting of 2 million iterations (with thinning every 20 iterations
 212 to reduce memory requirements), the simplest models required 12 mins on a machine
 213 running Windows 7 (3.4GHz Intel Core i7, 16GB RAM), while the more complicated
 214 models including time variation required at most 2 hrs. These relatively fast run times
 215 are attributable to **multimark**’s parallel processing of MCMC algorithms written in
 216 the C programming language (Kernighan & Ritchie, 1988). Bayesian multimodel
 217 inference was performed with *multimodelClosed()* using the default equal prior model
 218 weights, where 300000 iterations for each chain required 10.5 hrs. The longer run time
 219 for *multimodelClosed()* owes to the number of models and the RJMCMC algorithm

220 being written in R with no parallel processing.

221 Models including a positive behavioural response to first capture accounted for
222 0.51 of the posterior model weight, while models including $\delta_1 = \delta_2$ accounted for 0.78
223 of posterior model weight (Table 3). Model-averaged posterior modes were $N = 34$
224 (HPDI: 25-99; Fig. 1) for population abundance, $p = 0.15$ (HPDI: 0.04-0.28) for
225 capture probability, and $c = 0.21$ (HPDI: 0.07-0.34) for recapture probability. With
226 $\delta_1 = \delta_2 = 0.41$ (HPDI: 0.31-0.5) based on the model with the highest posterior
227 probability, both-sided encounters were relatively infrequent for these data ($1 - \delta_1 -$
228 $\delta_2 = 0.17$).

229 5 Discussion

230 I have described some of the key features of **multimark**, a new R package for the
231 analysis of capture-recapture data with multiple non-invasive marks. The package
232 adds to the growing toolbox of freely-available software for the analysis of non-spatial
233 (e.g., White & Burnham, 1999; Choquet *et al.*, 2009; Laake, 2013; Laake *et al.*, 2013)
234 and spatial (e.g., Gopalaswamy *et al.*, 2012; Efford, 2015) capture-recapture data,
235 but it is the first to combine otherwise irreconcilable encounter histories arising from
236 multiple mark types. Although initially developed for integrated analyses of left- and
237 right-sided images for bilaterally asymmetrical species, the package can be used to
238 jointly analyze data arising from any two types of marks. For example, **multimark**
239 could be used to integrate an analysis of encounter histories arising from genetic
240 (e.g., hair or faecal) and visual (e.g., photo-ID) detections (sensu Madon *et al.* 2011;
241 but see Bonner 2013).

242 **multimark** currently includes open population CJS and closed population models,

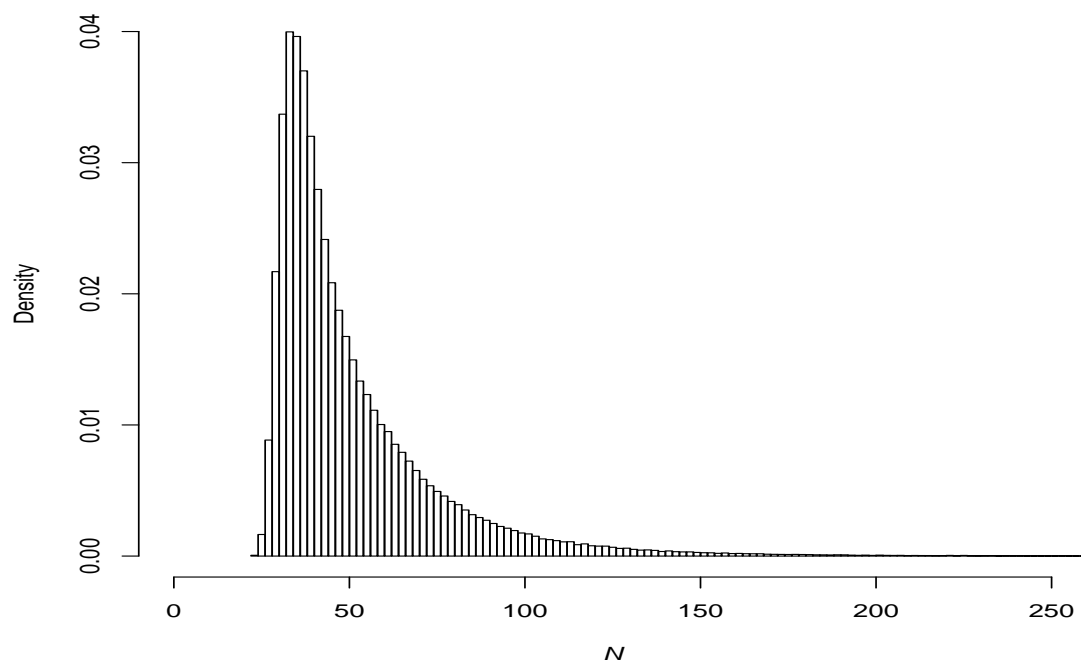


Figure 1. Model-averaged posterior distribution of population abundance (N) for the bobcat data.

Model	PMM	N	HPDI	ESS	GRB
p($\sim c$)delta(~ 1)	0.30	37	27-89	38769	1.00
p(~ 1)delta(~ 1)	0.22	33	27-46	51133	1.00
p($\sim h$)delta(~ 1)	0.15	48	28-113	12795	1.00
p($\sim c + h$)delta(~ 1)	0.09	49	28-144	19192	1.00
p($\sim c$)delta($\sim \text{type}$)	0.09	38	27-89	37853	1.00
p(~ 1)delta($\sim \text{type}$)	0.07	33	27-46	52354	1.00
p($\sim h$)delta($\sim \text{type}$)	0.04	46	28-113	13214	1.00
p($\sim c + h$)delta($\sim \text{type}$)	0.03	49	29-143	19525	1.00
p($\sim \text{time} + h$)delta(~ 1)	0.00	48	28-115	14781	1.00
p($\sim c + \text{time} + h$)delta(~ 1)	0.00	45	27-115	20273	1.00
p($\sim \text{time}$)delta(~ 1)	0.00	32	25-44	42386	1.00
p($\sim c + \text{time}$)delta(~ 1)	0.00	33	24-76	36202	1.00
p($\sim \text{time} + h$)delta($\sim \text{type}$)	0.00	47	28-115	12824	1.00
p($\sim \text{time}$)delta($\sim \text{type}$)	0.00	33	26-44	42781	1.00
p($\sim c + \text{time} + h$)delta($\sim \text{type}$)	0.00	44	28-115	20761	1.00
p($\sim c + \text{time}$)delta($\sim \text{type}$)	0.00	33	24-76	35808	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 $\delta_1 \neq \delta_2$ ($\sim \text{type}$).

with functions for derived parameters (e.g., ϕ , p) and Bayesian multimodel inference. 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 condi-

tioning on the observed encounter histories when identifying the feasible set of latent encounter histories. To facilitate better mixing, **multimark** extends 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. 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

275 the US Government.

276 References

- 277 Albert, J.H. & Chib, S. (1993) Bayesian analysis of binary and polychotomous re-
278 sponse data. *Journal of the American Statistical Association*, **88**, 669–679.
- 279 Alonso, R.S., McClintock, B.T., Lyren, L.M., Boydston, E.E. & Crooks, K.R. (2015)
280 Mark-recapture and mark-resight methods for estimating abundance with remote
281 cameras: a carnivore case study. *PLoS ONE*, **10**, e0123032.
- 282 Barker, R.J. & Link, W.A. (2013) Bayesian multimodel inference by RJMCMC: A
283 Gibbs sampling approach. *The American Statistician*, **67**, 150–156.
- 284 Berrow, S., O’Brien, J., Groth, L., Foley, A. & Voigt, K. (2012) Abundance estimate
285 of bottlenose dolphins (*tursiops truncatus*) in the Lower River Shannon candidate
286 Special Area of Conservation, Ireland. *Aquatic Mammals*, **38**, 136–144.
- 287 Bonner, S.J. & Holmberg, J. (2013) Mark-recapture with multiple, non-invasive
288 marks. *Biometrics*, **69**, 766–775.
- 289 Bonner, S. (2013) Response to: a new method for estimating animal abundance
290 with two sources of data in capture–recapture studies. *Methods in Ecology and*
291 *Evolution*, **4**, 585–588.
- 292 Choquet, R., Rouan, L. & Pradel, R. (2009) Program E-SURGE: a software appli-
293 cation for fitting multievent models. *Modeling demographic processes in marked*
294 *populations*, pp. 845–865. Springer.

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

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

303 Efford, M. (2015) *secr: Spatially explicit capture-recapture models*. R package version
 304 2.9.4.

305 Gopalaswamy, A.M., Royle, J.A., Hines, J.E., Singh, P., Jathanna, D., Kumar, N.
 306 & Karanth, K.U. (2012) Program SPACECAP: software for estimating animal
 307 density using spatially explicit capture–recapture models. *Methods in Ecology and*
 308 *Evolution*, **3**, 1067–1072.

309 Hammond, P. (1990) Capturing whales on film—estimating cetacean population pa-
 310 rameters from individual recognition data. *Mammal Review*, **20**, 17–22.

311 Holmberg, J., Norman, B. & Arzoumanian, Z. (2008) Robust, comparable population
 312 metrics through collaborative photo-monitoring of whale sharks *rhincodon typus*.
 313 *Ecological Applications*, **18**, 222–233.

314 Karanth, K.U. & Nichols, J.D. (1998) Estimation of tiger densities in India using
 315 photographic captures and recaptures. *Ecology*, **79**, 2852–2862.

- 316 Kernighan, B.W. & Ritchie, D.M. (1988) *The C Programming Language*. Prentice
317 Hall, Englewood Cliffs, New Jersey, USA, second edition.
- 318 King, R., McClintock, B.T., Kidney, D. & Borchers, D.L. (2015) Capture-recapture
319 abundance estimation using a semi-complete data likelihood approach. *The Annals
320 of Applied Statistics*.
- 321 Laake, J.L. (2013) *RMark: An R Interface for Analysis of Capture-Recapture Data
322 with MARK*. AFSC Processed Rep 2013-01, 25p. Alaska Fish. Sci. Cent., NOAA,
323 Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.
- 324 Laake, J.L., Johnson, D.S. & Conn, P.B. (2013) marked: an R package for maxi-
325 mum likelihood and markov chain monte carlo analysis of capture-recapture data.
326 *Methods in Ecology and Evolution*, **4**, 885–890.
- 327 Link, W.A., Yoshizaki, J., Bailey, L.L. & Pollock, K.H. (2010) Uncovering a latent
328 multinomial: analysis of mark-recapture data with misidentification. *Biometrics*,
329 **66**, 178–185.
- 330 Lukacs, P.M. & Burnham, K.P. (2005) Review of capture-recapture methods appli-
331 cable to noninvasive genetic sampling. *Molecular ecology*, **14**, 3909–3919.
- 332 Madon, B., Gimenez, O., McArdle, B., Scott Baker, C. & Garrigue, C. (2011) A new
333 method for estimating animal abundance with two sources of data in capture-
334 recapture studies. *Methods in Ecology and Evolution*, **2**, 390–400.
- 335 McClintock, B.T., Bailey, L.L., Dreher, B.P. & Link, W.A. (2014) Probit models for
336 capture-recapture data subject to imperfect detection, individual heterogeneity
337 and misidentification. *The Annals of Applied Statistics*, **8**, 2461–2484.

338 McClintock, B.T., Conn, P.B., Alonso, R.S. & Crooks, K.R. (2013) Integrated mod-
339 eling of bilateral photo-identification data in mark-recapture analyses. *Ecology*,
340 **94**, 1464–1471.

341 Morrison, T.A., Yoshizaki, J., Nichols, J.D. & Bolger, D.T. (2011) Estimating sur-
342 vival in photographic capture-recapture studies: overcoming misidentification er-
343 ror. *Methods in Ecology and Evolution*, **2**, 454–463.

344 Nair, T., Thorbjarnarson, J.B., Aust, P. & Krishnaswamy, J. (2012) Rigorous ghar-
345 ial population estimation in the Chambal: implications for conservation and man-
346 agement of a globally threatened crocodilian. *Journal of Applied Ecology*, **49**,
347 1046–1054.

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

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

352 R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R
353 Foundation for Statistical Computing, Vienna, Austria.

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

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

- 358 Ruell, E.W., Riley, S.P.D., Douglas, M.R., Pollinger, J.P. & Crooks, K.R. (2009)
 359 Estimating bobcat population sizes and densities in a fragmented urban landscape
 360 using noninvasive capture-recapture sampling. *Journal of Mammalogy*, **90**, 129–
 361 135.
- 362 White, G.C. & Burnham, K.P. (1999) Program MARK: Survival estimation from
 363 populations of marked animals. *Bird Study*, **46 Supplement**, 120–138.
- 364 Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of*
 365 *Animal Populations*. Academic Press, San Diego, CA, USA.
- 366 Wilson, B., Hammond, P.S. & Thompson, P.M. (1999) Estimating size and assessing
 367 trends in a coastal bottlenose dolphin population. *Ecological Applications*, **9**, 288–
 368 300.
- 369 Wright, J.A., Barker, R.J., Schofield, M.R., Frantz, A.C., Byrom, A.E. & Gleeson,
 370 D.M. (2009) Incorporating genotype uncertainty into mark-recapture-type models
 371 for estimating abundance using DNA samples. *Biometrics*, **65**, 833–840.
- 372 Yoshizaki, J., Pollock, K.H., Brownie, C. & Webster, R.A. (2009) Modeling misiden-
 373 tification errors in capture-recapture studies using photographic identification of
 374 evolving marks. *Ecology*, **90**, 3–9.