

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

May 11, 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 that enable individual identification in lieu of physical capture, 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, particularly 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 commonplace for monitoring animal populations (e.g. Hammond 1990; Lukacs & Burnham 2005; O’Connell *et al.* 2010). Examples of non-invasive marks include natural pelt or skin patterns, scars, and genetic markers that enable individual identification in the absence of physical capture. Capture-recapture methods based on non-invasive marks have been applied to diverse taxa, including sharks (e.g. Holmberg *et al.* 2008), reptiles (e.g. Nair *et al.* 2012), ursids (e.g. Dreher *et al.* 2007), felids (e.g. Karanth & Nichols 1998; Ruell *et al.* 2009), and marine mammals (e.g. Hammond 1990; Wilson *et al.* 1999; Madon *et al.* 2011). While non-invasive capture-recapture methods have many advantages related to financial cost and animal welfare, they also pose new difficulties such as animal misidentification (Wright *et al.* 2009; Yoshizaki *et al.* 2009; Link *et al.* 2010; Morrison *et al.* 2011) and the complexity of multiple types of marks (Corkrey *et al.* 2008; Madon *et al.* 2011; Bonner & Holmberg 2013; McClintock *et al.* 2013).

Multiple marks can arise from sighting or camera surveys of species with natu-

17 ral mark patterns that are bilaterally asymmetrical (e.g. cetaceans, felids) or from
18 multiple sources of non-invasive capture-recapture data being collected concurrently
19 (e.g. faecal DNA sampling and visual surveys). With multiple marks, an encounter
20 history is produced for each individual and mark type, but there is typically no reli-
21 able means to match them (unless each mark type is simultaneously observed at least
22 once for every encountered individual). Because the number of unique individuals
23 encountered must be known for standard capture-recapture analyses of population
24 abundance (or related demographic parameters), the typical approach is to conduct
25 separate analyses for each mark type and compare the results (e.g. Wilson *et al.*
26 1999; Berrow *et al.* 2012; Nair *et al.* 2012). However, given that sample sizes (and
27 precision) may be considerably reduced, this is not as efficient as conducting an inte-
28 grated analysis utilizing encounter histories arising from all mark types (McClintock
29 *et al.* 2013). Additional costs of conducting separate analyses for each mark type
30 include a limited ability to explore models with behavioural or cohort effects, and, for
31 capture-recapture models that condition on first encounter, a forfeiting of informa-
32 tion from histories with the (apparent) first encounter occurring on the last sampling
33 occasion. These limitations can be particularly important for the sparse datasets
34 typical of rare and elusive populations for which non-invasive sampling techniques
35 are most commonly employed.

36 Bonner & Holmberg (2013) and McClintock *et al.* (2013) recently developed meth-
37 ods for performing integrated analyses of capture-recapture data consisting of mul-
38 tiple non-invasive marks. However, to my knowledge, their approaches have yet to
39 be applied by practitioners. This is certainly not due to a lack of appropriate data
40 (e.g. Wilson *et al.* 1999; Holmberg *et al.* 2008; Madon *et al.* 2011; Berrow *et al.* 2012;

41 Nair *et al.* 2012), and is likely attributable to the mathematical and computational
 42 complexity of the models, as well as a lack of user-friendly software for implementing
 43 them. This is the motivation for **multimark**, an R (R Core Team 2013) package for
 44 the analysis of capture-recapture data consisting of multiple non-invasive marks.

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

55 **2 Description**

56 **2.1 Background**

57 Capture-recapture data are typically represented by a collection of encounter histo-
 58 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
 59 whether individual i was detected ($y_{i,t} = 1$) or not detected ($y_{i,t} = 0$) on each
 60 of $t = 1, \dots, T$ sampling occasions. Typical analyses then proceed by formulat-
 61 ing a likelihood conditional on the n unique individuals encountered (e.g. Williams
 62 *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
 63 $m \in \{1, 2\}$, 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

64 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
 65 unique individuals encountered for mark type m . We focus on situations where it is
 66 difficult (or impossible) to reliably match individuals from $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$. In this case,
 67 although we know $n \leq n_1 + n_2$, n is nevertheless unknown and standard capture-
 68 recapture analysis methods cannot be reliably used for simultaneous inference using
 69 both sources of data.

70 Depending on the mark types, it may sometimes be possible to observe both marks
 71 simulateneously. In this case, some of the encounter histories from $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$ can
 72 be matched to unique individuals with certainty. For example, suppose images were
 73 collected during vessel-based line transect surveys of surfacing whales, where mark
 74 type 1 corresponds to patch patterns on the left side and mark type 2 corresponds
 75 to patterns on the right side. If an individual happens to be photographed on both
 76 sides simultaneously on at least one sampling occasion, then the true encounter
 77 history for this individual would be known (i.e. left- and right-sided images could be
 78 matched). This results in an additional set of n_{known} observed encounter histories,
 79 $\tilde{\mathbf{Y}}_{known} = \{\tilde{\mathbf{y}}_{known1}, \tilde{\mathbf{y}}_{known2}, \dots, \tilde{\mathbf{y}}_{knownn_{known}}\}$, consisting of histories that are known
 80 with certainty (Table 1).

81 In essence, **multimark** facilitates the joint analysis of type 1 ($\tilde{\mathbf{Y}}_1$), type 2 ($\tilde{\mathbf{Y}}_2$),
 82 and known encounter histories ($\tilde{\mathbf{Y}}_{known}$) while accounting for uncertainty in the num-
 83 ber of unique individuals encountered using extensions of the methodology proposed
 84 by Bonner & Holmberg (2013) and McClintock *et al.* (2013). While the mathematical
 85 and computational details are complicated (and generally of little interest to ecol-
 86 ogists), **multimark** performs these operations in the background and requires only
 87 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

88 2.2 Models

89 `multimark` currently includes open population Cormack-Jolly-Seber (CJS) and closed
 90 population abundance models (e.g. Williams *et al.* 2002). These Bayesian implemen-
 91 tations are similar in spirit to the CJS model of Royle (2008) and the abundance
 92 model of King *et al.* (2015). Given the latent encounter histories (\mathbf{Y}) that generated
 93 the observed encounter histories ($\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Y}}_{known}$), the likelihood for the CJS model
 94 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:

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

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

95 where $\Phi()$ the cumulative distribution function of the standard normal density, \mathbf{x}_t^p
 96 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
 97 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
 98 effects. The probit link is implemented for CJS models in `multimark` because it
 99 facilitates a Gibbs sampler in the spirit of Albert & Chib (1993) and Laake *et al.*
 100 (2013).

101 Similarly, the likelihood for the closed population abundance model with two
 102 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. 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

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

103 Although a Gibbs sampler has been proposed for closed population models using the
 104 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”	$\{0, 1, 2, 4\}$	$\alpha = 1$

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

111 3 Workflow

112 3.1 Data formatting

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

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

122 corresponding to individuals and columns corresponding to sampling occasions. Be-
 123 cause the bobcat data were collected from single-camera stations, simultaneous left-
 124 and right-sided encounters were not possible; hence $\alpha = 0$ and the rows consist of
 125 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

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

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

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

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

141 3.2 Model fitting

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

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

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

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

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

148 This creates a list, *bobcat.dot*, containing the MCMC output for the model
 149 (*bobcat.dot\$mcmc*). The MCMC output is of class *mcmc*, which should be familiar
 150 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)]	-1.4607	0.24430	0.0024430	0.011345
N	36.7901	6.00502	0.0600502	0.308523
delta_1	0.3512	0.06969	0.0006969	0.003857
delta_2	0.3703	0.07016	0.0007016	0.003403

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-1.9684	-1.6154	-1.4524	-1.2990	-0.9968
N	28.0000	33.0000	36.0000	40.0000	51.0000
delta_1	0.2173	0.3041	0.3495	0.3983	0.4911
delta_2	0.2329	0.3226	0.3697	0.4173	0.5097

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

	N	delta_1	delta_2
pbeta[(Intercept)]	463.7188	378.8376	326.4041
			425.0732

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

152 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
153 all quantities of interest.

155 Other common models for detection probability can be easily specified using
156 formulas for $mod.p$, including shorthands for time variation ($mod.p=\sim time$), temporal trends ($mod.p=\sim Time$), behavioural response to first capture ($mod.p=\sim c$),
157

158 and individual heterogeneity ($mod.p=\sim h$). Additive or interaction terms can be in-
 159 cluded (e.g. $mod.p=\sim time+c+h$, $mod.p=\sim Time+I(Time^2)$, $mod.p=\sim time*c$). User-
 160 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"))
```

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

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

171 The function *multimarkCJS()* works in exactly the same fashion, with the only
 172 notable difference being specification of models for ϕ (in addition to p and δ). Al-
 173 though CJS-specific data are not included with *multimark*, data can be simulated
 174 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)]	-0.1506	0.13261	0.0004688	0.004015
phibeta[(Intercept)]	1.3628	0.28253	0.0009989	0.011343
alpha	0.4120	0.11487	0.0004061	0.002768
sigma2_zphi	0.0358	0.08388	0.0002965	0.004013
delta_1	0.1966	0.04696	0.0001660	0.001721
delta_2	0.6094	0.05155	0.0001823	0.001148

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
pbeta[(Intercept)]	-0.399585	-0.242602	-0.15405	-0.06165	0.1178
phibeta[(Intercept)]	0.910577	1.165018	1.32782	1.52288	2.0049
alpha	0.206888	0.329362	0.40570	0.48827	0.6517
sigma2_zphi	0.002717	0.006915	0.01313	0.02964	0.2260
delta_1	0.110136	0.163964	0.19486	0.22763	0.2930
delta_2	0.506525	0.574743	0.61029	0.64484	0.7074

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

178 3.3 Further analysis

179 While the `coda` package can be used to summarize, plot, and assess convergence
 180 of MCMC samples from *multimarkClosed()* and *multimarkCJS()*, several additional
 181 functions are available for further analysis. Because link functions are used for p and
 182 ϕ , the functions *getprobsClosed()* and *getprobsCJS()* provide estimates on the real
 183 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]	0.1408	0.05535	0.0005535	0.005415
c[2]	0.2586	0.05406	0.0005406	0.003720

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
p[1]	0.0473	0.09948	0.1379	0.1749	0.2630
c[2]	0.1587	0.22002	0.2568	0.2941	0.3705

184 Here, `p[1]` and `c[2]` refer to the probabilities of capture and recapture at times $t = 1$
 185 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

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

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

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

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

221 being written entirely in R.

222 Models including a positive behavioural response to first capture accounted for
223 0.51 of the posterior model weight, while models including $\delta_1 = \delta_2$ accounted for 0.78
224 of posterior model weight (Table 3). Model-averaged posterior modes were $N = 35$
225 (HPDI: 26-101; Fig. 1) for population abundance, $p = 0.15$ (HPDI: 0.04-0.27) for
226 capture probability, and $c = 0.21$ (HPDI: 0.07-0.33) for recapture probability. With
227 $\delta_1 = \delta_2 = 0.41$ (HPDI: 0.30-0.50) based on the model with the highest posterior
228 probability, both-sided encounters were relatively infrequent for these data ($1 - \delta_1 -$
229 $\delta_2 = 0.18$; HPDI: 0.00-0.39).

230 5 Discussion

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

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

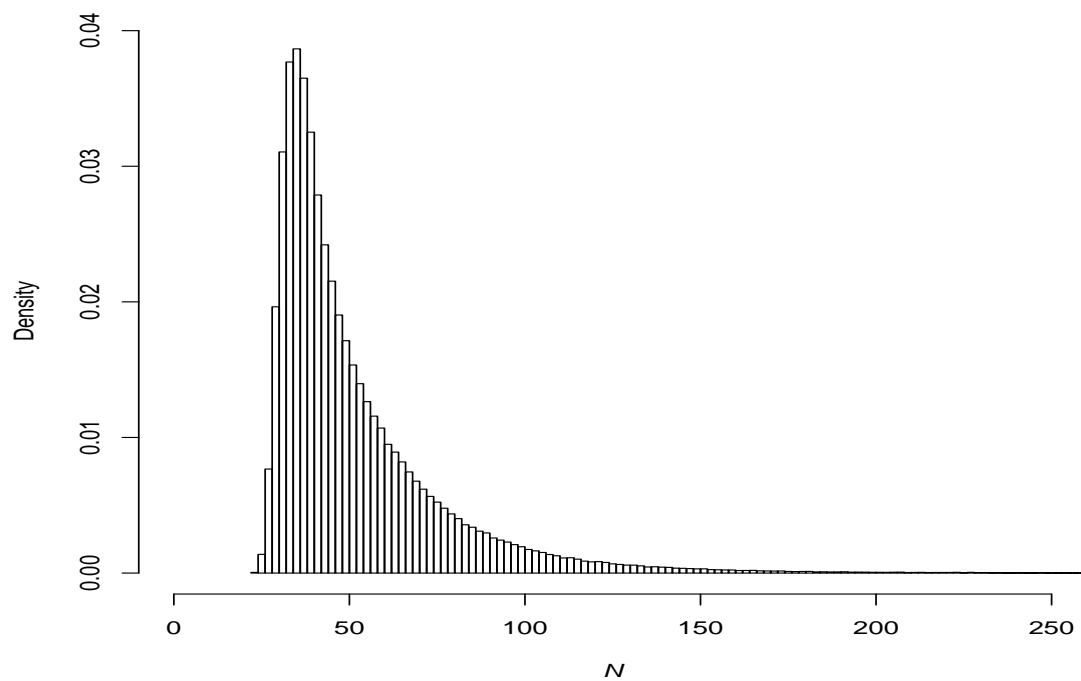


Figure 1. Model-averaged posterior distribution of population abundance (N) for the bobcat data.

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

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

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 condition-

ing 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. 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

276 an endorsement by the US Government.

277 References

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

- 296 Corkrey, R., Brooks, S., Lusseau, D., Parsons, K., Durban, J.W., Hammond, P.S.
297 & Thompson, P.M. (2008) A Bayesian capture-recapture population model with
298 simultaneous estimation of heterogeneity. *Journal of the American Statistical As-*
299 *sociation*, **103**, 948–960.
- 300 Dreher, B.P., Winterstein, S.R., Scribner, K.T., Lukacs, P.M., Etter, D.R., Rosa,
301 G.J.M., Lopez, V.A., Libants, S. & Filcek, K.B. (2007) Noninvasive estimation
302 of black bear abundance incorporating genotyping errors and harvested bears.
303 *Journal of Wildlife Management*, **71**, 2684–2693.
- 304 Efford, M. (2015) *secr: Spatially explicit capture-recapture models*. R package version
305 2.9.4.
- 306 Gopalaswamy, A.M., Royle, J.A., Hines, J.E., Singh, P., Jathanna, D., Kumar, N.
307 & Karanth, K.U. (2012) Program SPACECAP: software for estimating animal
308 density using spatially explicit capture–recapture models. *Methods in Ecology and*
309 *Evolution*, **3**, 1067–1072.
- 310 Hammond, P. (1990) Capturing whales on film—estimating cetacean population pa-
311 rameters from individual recognition data. *Mammal Review*, **20**, 17–22.
- 312 Holmberg, J., Norman, B. & Arzoumanian, Z. (2008) Robust, comparable population
313 metrics through collaborative photo-monitoring of whale sharks *rhincodon typus*.
314 *Ecological Applications*, **18**, 222–233.
- 315 Karanth, K.U. & Nichols, J.D. (1998) Estimation of tiger densities in India using
316 photographic captures and recaptures. *Ecology*, **79**, 2852–2862.

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

- 339 McClintock, B.T., Conn, P.B., Alonso, R.S. & Crooks, K.R. (2013) Integrated mod-
 340 eling of bilateral photo-identification data in mark-recapture analyses. *Ecology*,
 341 **94**, 1464–1471.
- 342 Morrison, T.A., Yoshizaki, J., Nichols, J.D. & Bolger, D.T. (2011) Estimating sur-
 343 vival in photographic capture-recapture studies: overcoming misidentification er-
 344 ror. *Methods in Ecology and Evolution*, **2**, 454–463.
- 345 Nair, T., Thorbjarnarson, J.B., Aust, P. & Krishnaswamy, J. (2012) Rigorous ghar-
 346 ial population estimation in the Chambal: implications for conservation and man-
 347 agement of a globally threatened crocodilian. *Journal of Applied Ecology*, **49**,
 348 1046–1054.
- 349 O’Connell, A.F., Nichols, J.D. & Karanth, K.U. (2010) *Camera traps in animal*
 350 *ecology: methods and analyses*. Springer Science & Business Media.
- 351 Plummer, M., Best, N., Cowles, K. & Vines, K. (2006) CODA: convergence diagnosis
 352 and output analysis for MCMC. *R News*, **6**, 7–11.
- 353 R Core Team (2013) *R: A Language and Environment for Statistical Comput-*
 354 *ing*. R Foundation for Statistical Computing, Vienna, Austria. [http://www.R-](http://www.R-project.org/)
 355 [project.org/](http://www.R-project.org/).
- 356 Royle, J.A. (2008) Modeling individual effects in the Cormack-Jolly-Seber model: a
 357 state-space formulation. *Biometrics*, **64**, 364–370.
- 358 Royle, J.A. (2015) Spatial capture-recapture with partial identity. *arXiv preprint*
 359 *arXiv:150306873*.

- 360 Ruell, E.W., Riley, S.P.D., Douglas, M.R., Pollinger, J.P. & Crooks, K.R. (2009)
361 Estimating bobcat population sizes and densities in a fragmented urban landscape
362 using noninvasive capture-recapture sampling. *Journal of Mammalogy*, **90**, 129–
363 135.
- 364 White, G.C. & Burnham, K.P. (1999) Program MARK: Survival estimation from
365 populations of marked animals. *Bird Study*, **46**, S120–S138.
- 366 Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of*
367 *Animal Populations*. Academic Press, San Diego, CA, USA.
- 368 Wilson, B., Hammond, P.S. & Thompson, P.M. (1999) Estimating size and assessing
369 trends in a coastal bottlenose dolphin population. *Ecological Applications*, **9**, 288–
370 300.
- 371 Wright, J.A., Barker, R.J., Schofield, M.R., Frantz, A.C., Byrom, A.E. & Gleeson,
372 D.M. (2009) Incorporating genotype uncertainty into mark-recapture-type models
373 for estimating abundance using DNA samples. *Biometrics*, **65**, 833–840.
- 374 Yoshizaki, J., Pollock, K.H., Brownie, C. & Webster, R.A. (2009) Modeling misiden-
375 tification errors in capture-recapture studies using photographic identification of
376 evolving marks. *Ecology*, **90**, 3–9.