



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Capture-Recapture Abundance Estimation using a Semi-complete Data Likelihood Approach

**Citation for published version:**

King, R, McClintock, BT, Kidney, D & Borchers, DL 2016, 'Capture-Recapture Abundance Estimation using a Semi-complete Data Likelihood Approach', *Annals of Applied Statistics*, vol. 10, no. 1, pp. 264-285.  
<https://doi.org/10.1214/15-AOAS890>

**Digital Object Identifier (DOI):**

[10.1214/15-AOAS890](https://doi.org/10.1214/15-AOAS890)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Annals of Applied Statistics

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# CAPTURE-RECAPTURE ABUNDANCE ESTIMATION USING A SEMI-COMPLETE DATA LIKELIHOOD APPROACH

BY RUTH KING<sup>\*</sup>, BRETT T. MCCLINTOCK<sup>†</sup>, DARREN KIDNEY<sup>‡</sup> AND DAVID BORCHERS<sup>‡</sup>  
*University of Edinburgh<sup>\*</sup>, NOAA National Marine Mammal Laboratory<sup>†</sup> and University of St  
Andrews<sup>‡</sup>*

E-MAIL: [Ruth.King@ed.ac.uk](mailto:Ruth.King@ed.ac.uk)

Capture-recapture data are often collected when abundance estimation is of interest. In this manuscript we focus on abundance estimation of closed populations. In the presence of unobserved individual heterogeneity, specified on a continuous scale for the capture probabilities, the likelihood is not generally available in closed form, but expressible only as an analytically intractable integral. Model-fitting algorithms to estimate abundance most notably include a numerical approximation for the likelihood or use of a Bayesian data augmentation technique considering the complete data likelihood. We consider a Bayesian hybrid approach, defining a “semi-complete” data likelihood, composed of the product of a complete data likelihood component for individuals seen at least once within the study and a marginal data likelihood component for the individuals not seen within the study, approximated using numerical integration. This approach combines the advantages of the two different approaches, with the semi-complete likelihood component specified as a single integral (over the dimension of the individual heterogeneity component). In addition, the models can be fitted within BUGS/JAGS (commonly used for the Bayesian complete data likelihood approach) but with significantly improved computational efficiency compared to the commonly used super-population data augmentation approaches (between about 10 and 77 times more efficient in the two examples we consider). The semi-complete likelihood approach is flexible and applicable to a range of models, including spatially explicit capture-recapture models. The model-fitting approach is applied to two different datasets: the first relates to snowshoe hares where model  $M_h$  is applied and the second to gibbons where a spatially explicit capture-recapture model is applied.

**1. Introduction.** In order to estimate total abundance capture-recapture data are often collected on the population under study. Capture-recapture data collection methods involve partially observing the population at a series of capture events (or using a number of different sources), such that each individual observed within the study is uniquely identifiable. Assuming that marks are unique and cannot be lost, a capture history for each individual observed within the study can be constructed, detailing whether the given individual is observed or not at each capture event. Statistical models can be constructed and applied to capture-recapture data to estimate the number of individuals in the population that are not observed. We focus on closed population models, where it is assumed that there are no births/deaths/migrations in the population within the study period. Applications include estimating the number of injecting drug users ([King \*et al.\*, 2014](#); [Overstall \*et al.\*, 2014](#)), pages on the world wide web ([Fienberg \*et al.\*, 1999](#)), disease prevalence ([Manrique-Vallier and Fienberg, 2008](#)) and animal populations ([Borchers \*et al.\*, 2002](#)). We focus on statistical models for ecological data where individuals are observed at a series of capture events. For further discussion of ecological (closed) capture-recapture data, and the underlying assumptions, see for example, [Borchers \*et al.\* \(2002\)](#), [Williams \*et al.\* \(2002\)](#) and [McCrea and Morgan \(2014\)](#).

---

*Keywords and phrases:* BUGS; capture-recapture; closed populations; individual heterogeneity; JAGS; spatially explicit.

In general, the likelihood of capture-recapture data can be expressed in multinomial form, where the different multinomial cells correspond to each possible capture history and the cell entries to the number of individuals with the given capture history. The unknown parameters to be estimated in the likelihood function are the capture (or detection) probabilities and the total population size (or number of individuals in the population unobserved at any capture event). Otis *et al.* (1978) described three different possible effects on the capture probabilities corresponding to temporal ( $t$ ), behavioural ( $b$ ) and individual heterogeneity ( $h$ ) effects. We adopt the standard notation and describe the different models by  $M_a$ , such that  $a \subseteq \{t, b, h\}$ , corresponding to the combination of effects in the given model.

In this paper we focus on models that include individual heterogeneity (i.e.  $M_h$ -type models). Individual heterogeneity is often introduced by specifying the capture probabilities as a finite or infinite mixture. Finite mixture models lead to an explicit likelihood expression which can be maximised numerically to obtain the maximum likelihood estimates (MLEs) of the parameters of interest (Pledger, 2000). Infinite mixture models specify the individual heterogeneity as a random effects model. For the special case of a Beta-Binomial random effects component the likelihood is available in closed form (Dorazio and Royle, 2003; Morgan and Ridout, 2008). We will consider the more general case, with an arbitrary individual heterogeneity component leading to an analytically intractable likelihood. Previous approaches to fit such models to the data include (i) numerical integration to estimate the *marginal* (or observed) data likelihood (Coull and Agresti, 1999; Borchers and Efford, 2008; Gimenez and Choquet, 2010); and (ii) Bayesian data augmentation techniques, using a *complete* data likelihood approach (corresponding to the joint probability density function of the capture histories and individual effects), integrating out the individual heterogeneity component within a Markov chain Monte Carlo-type (MCMC) algorithm (Durban and Elston, 2005; Royle *et al.*, 2007, 2009; King and Brooks, 2008; King *et al.*, 2009; Royle and Dorazio, 2012). We combine these two approaches defining a *semi-complete* data likelihood constructed as the product of a complete data likelihood component for the individuals seen at least once in the study and a marginal data likelihood component for the unseen individuals. This combines the advantages of each of the individual approaches. We note that similar approaches have been previously proposed for specific applications, using bespoke computer codes. Most notably, Fienberg *et al.* (1999) propose a conditional MCMC algorithm for Rasch-type models, employing a block update of the total population size and individual heterogeneity terms; while Bonner and Schofield (2014) consider an additional Monte Carlo integration step within the MCMC algorithm applied to individual covariate models. We describe how the latter approach is a special case of our general semi-complete data likelihood approach in Section 3.3. Finally, we demonstrate how individual heterogeneity models can be efficiently fitted using BUGS/JAGS with general prior structures specified on all the model parameters (including the total population size) and provide the associated computer codes.

The paper proceeds as follows. Section 2 describes the general closed population model structure and associated notation. Section 3 describes previous model-fitting approaches and the new proposed semi-complete data likelihood approach. The implications of the BUGS/JAGS specification for the semi-complete data likelihood and previous Bayesian complete data likelihood approaches are compared in Section 4 and the approaches applied and compared for two real examples: the first example relates to snowshoe hares where model  $M_h$  is applied and the second to a dataset of gibbons where a spatially explicit capture-recapture model is applied. Finally in Section 5 we conclude with a discussion.

**2. Individual heterogeneity models.** We assume that within the capture-recapture study there is a series of  $T$  discrete capture occasions. Within the study a total of  $n$  distinct individuals are observed, with the total (unknown) population size denoted by  $N$ . For simplicity we arbitrarily number the observed individuals  $i = 1, \dots, n$  and the unobserved individuals  $i = n + 1, \dots, N$ . Let  $p_{it}$  denote the capture probability of individual  $i = 1, \dots, N$  at time  $t = 1, \dots, T$ . Further, for

101 standard capture-recapture data,  $\mathbf{x}_i = \{x_{it} : t = 1, \dots, T\}$  denotes the capture history of individual  
 102  $i = 1, \dots, N$ , such that

$$103 \quad x_{it} = \begin{cases} 0 & \text{individual } i \text{ is unobserved on occasion } t; \\ 1 & \text{individual } i \text{ is observed on occasion } t. \end{cases}$$

104 We consider individual heterogeneity specified such that

$$105 \quad p_{it} = g(\boldsymbol{\theta}, \boldsymbol{\epsilon}_i),$$

106 for some function  $g$ , where  $\boldsymbol{\theta}$  denotes the model parameters associated with the capture probabilities  
 107 (which may include, for example, temporal and/or behavioural effect terms, regression coefficients  
 108 for covariate values etc.) and  $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_i : i = 1, \dots, N\}$  such that  $\boldsymbol{\epsilon}_i \in \mathbb{S} \subset \mathbb{R}^k$ , corresponding to  
 109 the individual heterogeneity term for individual  $i = 1, \dots, N$ . Further, we assume an underlying  
 110 model for the individual heterogeneity, such that  $\boldsymbol{\epsilon}$  is a function of the parameters  $\boldsymbol{\eta}$ , and that the  
 111 individual heterogeneity terms,  $\boldsymbol{\epsilon}_i$ , are independent of each other conditional on  $\boldsymbol{\eta}$ . The associated  
 112 joint probability density function of the heterogeneity terms is given by  $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}|N, \boldsymbol{\eta}) = \prod_{i=1}^N f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_i|\boldsymbol{\eta})$ ,  
 113 using the conditional independence assumption (and dropping the dependence on  $N$  for the condi-  
 114 tional density function of the individual heterogeneity terms for individual  $i$ ). Further, to provide  
 115 a general framework for both observed and unobserved individual heterogeneity we additionally  
 116 write  $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}^{Obs}, \boldsymbol{\epsilon}^{Mis}\}$  where  $\boldsymbol{\epsilon}^{Obs}$  denotes the set of observed individual heterogeneity compo-  
 117 nents and  $\boldsymbol{\epsilon}^{Mis}$  the set of unobserved individual heterogeneity components. Similarly, we write  
 118  $\boldsymbol{\epsilon}_i = \{\boldsymbol{\epsilon}_i^{Obs}, \boldsymbol{\epsilon}_i^{Mis}\}$ , for  $i = 1, \dots, N$  with obvious notation. Finally, we assume that the capture  
 119 histories of the individuals are independent of each other given the capture probability model  
 120 parameters,  $\boldsymbol{\theta}$ , and individual heterogeneity terms,  $\boldsymbol{\epsilon}$ .

121 The *marginal* data likelihood can be expressed in the form,

$$\begin{aligned} 122 \quad f_m(\mathbf{x}, \boldsymbol{\epsilon}^{Obs}|N, \boldsymbol{\theta}, \boldsymbol{\eta}) &= \int_{\boldsymbol{\epsilon}_1^{Mis}} \dots \int_{\boldsymbol{\epsilon}_N^{Mis}} f_c(\mathbf{x}, \boldsymbol{\epsilon}|N, \boldsymbol{\theta}, \boldsymbol{\eta}) d\boldsymbol{\epsilon}_1^{Mis} \dots d\boldsymbol{\epsilon}_N^{Mis} \\ 123 \quad &= \int_{\boldsymbol{\epsilon}_1^{Mis}} \dots \int_{\boldsymbol{\epsilon}_N^{Mis}} f_{\mathbf{x}}(\mathbf{x}|N, \boldsymbol{\theta}, \boldsymbol{\epsilon}) f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}|N, \boldsymbol{\eta}) d\boldsymbol{\epsilon}_1^{Mis} \dots d\boldsymbol{\epsilon}_N^{Mis} \\ 124 \quad (2.1) \quad &\propto \frac{N!}{(N-n)!} \prod_{i=1}^N \int_{\boldsymbol{\epsilon}_i^{Mis}} f_{\mathbf{x}}(\mathbf{x}_i|\boldsymbol{\theta}, \boldsymbol{\epsilon}_i) f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_i|\boldsymbol{\eta}) d\boldsymbol{\epsilon}_i^{Mis}, \end{aligned}$$

125 using the multinomial distributional form of the capture-recapture data (omitting the constant  
 126 multinomial coefficients for simplicity), and conditional independence of the random effect terms.  
 127 The term  $f_c(\mathbf{x}, \boldsymbol{\epsilon}|N, \boldsymbol{\theta}, \boldsymbol{\eta})$  corresponds to the *complete* data likelihood (i.e. the joint probability den-  
 128 sity function of the capture histories and individual effects);  $f_{\mathbf{x}}(\mathbf{x}|N, \boldsymbol{\theta}, \boldsymbol{\epsilon})$  the conditional likelihood  
 129 of the capture histories (where the conditioning includes the individual heterogeneity terms); and  
 130  $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}|N, \boldsymbol{\eta})$  the joint probability density function of the individual heterogeneity terms. The term  
 131  $f_{\mathbf{x}}(\mathbf{x}_i|\boldsymbol{\theta}, \boldsymbol{\epsilon}_i)$  corresponds to the conditional likelihood of capture history for individual  $i = 1, \dots, N$ ;  
 132 and  $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_i|\boldsymbol{\eta})$  the conditional probability density function of the individual heterogeneity component  
 133 for individual  $i = 1, \dots, N$  (where in each case we drop the dependence on  $N$ ).

134 *Example 1 - Continuous individual covariates.* We consider the case with  $q$  time-invariant con-  
 135 tinuous individual covariates  $\boldsymbol{\epsilon} = \{\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N\}$  where  $\boldsymbol{\epsilon}_i \in \mathbb{S} \subseteq \mathbb{R}^q$  denotes the covariate values as-  
 136 sociated with individual  $i = 1, \dots, N$ . Since the covariate values are time-invariant, the associated  
 137 capture probabilities for each individual are also time-invariant, so that  $p_{it} = p_i$  for  $t = 1, \dots, T$ .  
 138 Assuming that the capture probabilities are linearly related to the covariate values via some link  
 139 function, we may specify,

$$140 \quad g^{-1}(p_i) = \alpha + \boldsymbol{\beta}^T \boldsymbol{\epsilon}_i,$$

141 so that  $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\beta}\}$ . Common choices for  $g^{-1}$  include the logit and probit functions. Additional  
 142 individual/temporal random effects can be included in the capture probabilities, but we omit  
 143 these here for simplicity (see Example 2). Further we specify a parametric model for the covariate  
 144 values, assuming that conditional on the additional covariate parameters  $\boldsymbol{\eta}$ , the covariate values  
 145 are independent.

146 Assuming that for each individual observed within the study the set of individual covariate  
 147 values is recorded, we have that  $\boldsymbol{\epsilon}^{Obs} = \{\epsilon_i : i = 1, \dots, n\}$  and  $\boldsymbol{\epsilon}^{Mis} = \{\epsilon_i : i = n + 1, \dots, N\}$ . More  
 148 generally, the covariate values may not be recorded for all observed individuals. For example, the  
 149 observation process may include sightings recorded from a distance (rather than physical captures)  
 150 so that the covariate may not be able to be obtained if a physical capture is necessary (for example if  
 151 the covariate corresponds to wingspan). In this case the set of unobserved individual heterogeneity  
 152 terms is extended to include the unknown covariate values for observed individuals.

153 The complete data likelihood is of the form,

$$154 \quad f_c(\mathbf{x}, \boldsymbol{\epsilon} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^N \left[ \prod_{t=1}^T p_{it}^{x_{it}} (1-p_{it})^{1-x_{it}} \right] \times f_\epsilon(\boldsymbol{\epsilon}_i | \boldsymbol{\eta})$$

$$155 \quad = \frac{N!}{(N-n)!} \prod_{i=1}^N p_i^{y_i} (1-p_i)^{T-y_i} \times f_\epsilon(\boldsymbol{\epsilon}_i | \boldsymbol{\eta}),$$

156 where  $p_i$  is of the above form and  $y_i = \sum_{t=1}^T x_{it}$  (denoting the total number of times individual  $i$  is  
 157 observed). The first term of the complete data likelihood corresponds to the conditional likelihood  
 158 (conditional on the individual covariate terms) and the second term to the individual covariate  
 159 component.

160 The marginal data likelihood integrates out the unobserved covariate values  $\boldsymbol{\epsilon}^{Mis}$ . For notational  
 161 simplicity we provide the marginal data likelihood for the special case where all covariate values  
 162 are known for individuals observed within the study (i.e.  $\boldsymbol{\epsilon}^{Obs} = \{\epsilon_i : i = 1, \dots, n\}$  and  $\boldsymbol{\epsilon}^{Mis} =$   
 163  $\{\epsilon_i : i = n + 1, \dots, N\}$ ):

$$164 \quad f_m(\mathbf{x}, \boldsymbol{\epsilon}^{Obs} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^n p_i^{y_i} (1-p_i)^{T-y_i} f_\epsilon(\boldsymbol{\epsilon}_i | \boldsymbol{\eta}) \times \prod_{i=n+1}^N \int_{\boldsymbol{\epsilon}_i} p_i^{y_i} (1-p_i)^{T-y_i} f_\epsilon(\boldsymbol{\epsilon}_i | \boldsymbol{\eta}) d\boldsymbol{\epsilon}_i$$

$$165 \quad = \frac{N!}{(N-n)!} \prod_{i=1}^n p_i^{y_i} (1-p_i)^{T-y_i} f_\epsilon(\boldsymbol{\epsilon}_i | \boldsymbol{\eta}) \times \left[ \int_{\boldsymbol{\epsilon}_0} (1-p_0)^T f_\epsilon(\boldsymbol{\epsilon}_0 | \boldsymbol{\eta}) d\boldsymbol{\epsilon}_0 \right]^{N-n},$$

166 where  $g^{-1}(p_0) = \alpha + \boldsymbol{\beta}^T \boldsymbol{\epsilon}_0$ . The extension to the case where observed individuals may also have  
 167 unknown covariate values is immediate.

168 We note, in general, the model can be extended to include time-varying individual covariates, us-  
 169 ing the time and individual dependent capture probability,  $p_{it}$ . This typically substantially increases  
 170 the number of unobserved covariate values, since if an individual is not observed, the corresponding  
 171 covariate value is necessarily also unknown. However, for closed populations, to satisfy the condi-  
 172 tion that the population is closed the study period is generally short in duration so that changes  
 173 in time-varying individual covariate values is likely to be limited.

174 *Example 2 -  $M_h$ -type models.* For  $M_h$ -type models the individual heterogeneity corresponds to  
 175 an unobserved individual random effect component (so that  $\boldsymbol{\epsilon}^{Obs} = \emptyset$  and  $\boldsymbol{\epsilon}^{Mis} = \boldsymbol{\epsilon}$ ). For example,  
 176 for model  $M_h$  we may set  $\boldsymbol{\theta} = \{\alpha\}$  and  $\boldsymbol{\eta} = \{\sigma^2\}$  such that,

$$177 \quad \epsilon_i \sim N(0, \sigma^2),$$

178 for  $i = 1, \dots, N$ , where  $\sigma^2$  denotes the individual random effect variance and  $\mathbb{S} = \mathbb{R}$ . For this  
 179 model, the capture probabilities are again independent of time  $t$ , so we can write  $p_{it} = p_i$  for all  
 180  $t = 1, \dots, T$ , with

$$181 \quad g^{-1}(p_i) = \alpha + \epsilon_i,$$

182 for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . Common choices for  $g^{-1}$  include the logit and probit functions.  
 183 The extension to incorporate additional time and/or behavioural effects is immediate (i.e. models  
 184  $M_{th}$ ,  $M_{bh}$  and  $M_{tbh}$ ; see for example [King and Brooks, 2008](#)).

185 The complete data likelihood for model  $M_h$  can be written in the form,

$$186 \quad f_c(\mathbf{x}, \boldsymbol{\epsilon} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^N p_i^{y_i} (1-p_i)^{T-y_i} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right),$$

187 where  $p_i$  is of the above form and  $y_i = \sum_{t=1}^T x_{it}$ . Once again, the first term of the complete data  
 188 likelihood corresponds to the conditional likelihood (conditional on the individual random effect  
 189 terms) and the second term to the individual effect component.

190 The marginal data likelihood integrates out the  $\boldsymbol{\epsilon}$  terms and (dropping the term  $\boldsymbol{\epsilon}^{Obs}$  since no  
 191 individual heterogeneity terms are observed, i.e.  $\boldsymbol{\epsilon}^{Obs} = \emptyset$ ) can be efficiently expressed as,

$$192 \quad f_m(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{k=0}^T \left[ \int_{\epsilon_k \in \mathbb{S}} (p_k)^k (1-p_k)^{T-k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_k^2}{2\sigma^2}\right) d\epsilon_k \right]^{n_k},$$

193 where  $n_k = \sum_{i=1}^N I(y_i = k)$  and denotes the number of individuals observed  $k$  times within the  
 194 study, for  $k = 0, \dots, T$  (so that  $n_0$  is unobserved and  $N = n_0 + n$ ) and  $g^{-1}(p_k) = \alpha + \epsilon_k$ .

195 *Example 3 - SECR models.* For traditional spatially explicit capture-recapture models,  $\mathbb{S} \subset \mathbb{R}^2$   
 196 and the individual heterogeneity corresponds to the unobserved activity centre of the individual  
 197 (so that  $\boldsymbol{\epsilon}^{Obs} = \emptyset$  and  $\boldsymbol{\epsilon}^{Mis} = \boldsymbol{\epsilon}$ ). The range of possible models is greater for SECR than non-spatial  
 198 capture-recapture as SECR models involve multiple traps or detectors at different locations on each  
 199 occasion and take account of the location(s) of observations within occasions. To this end we define  
 200  $\mathbf{u}_j = (u_{j1}, u_{j2}) \in \mathbb{R}^2$  to be the Cartesian coordinates of trap  $j$ , for  $j = 1, \dots, J$ . We consider the  
 201 likelihood for a study with binary detection data within occasion, such that

$$202 \quad x_{ijt} = \begin{cases} 0 & \text{individual } i \text{ is unobserved by detector } j \text{ on occasion } t; \\ 1 & \text{individual } i \text{ is observed by detector } j \text{ on occasion } t. \end{cases}$$

203 We consider the case where individuals can be observed by more than one detector at each occa-  
 204 sion and we assume that observations by different detectors within occasions (as well as between  
 205 occasions) are independent. In this context,  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}) \in \mathbb{R}^2$  ( $i = 1, \dots, N$ ) denote the Carte-  
 206 sian coordinates of the activity centres of the  $N$  individuals in  $\mathbb{S} \subset \mathbb{R}^2$ . It is usually assumed that  
 207 these are independently uniformly distributed in  $\mathbb{S}$  and do not change between occasions, so that  
 208  $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} | N, \boldsymbol{\eta}) = \prod_{i=1}^N f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_i | \boldsymbol{\eta}) = A^{-N}$ , where  $A$  is the area of  $\mathbb{S}$ . The probability of individual  $i$  being  
 209 observed by detector  $j$  at capture occasion  $t$ , denoted  $p_{ijt}$  is assumed to depend on only the dis-  
 210 tance of the detector from the activity centre of individual  $i$ , so that  $p_{ijt} = g(\boldsymbol{\theta}, \|\mathbf{u}_j - \boldsymbol{\epsilon}_i\|)$ , where  
 211  $\|\mathbf{u}_j - \boldsymbol{\epsilon}_i\|$  is the vector norm  $\sqrt{\sum_{k=1}^2 (u_{jk} - \epsilon_{ik})^2}$ . The half-normal form is a common choice for  $g$ .  
 212 For example, assuming that the capture probabilities are time-independent, we may specify,

$$213 \quad p_{ijt} = p_{ij} = p_0 \exp\left(-\frac{\|\mathbf{u}_j - \boldsymbol{\epsilon}_i\|^2}{2\sigma^2}\right)$$

214 with  $\boldsymbol{\theta} = \{p_0, \sigma^2\}$ .

215 The complete data likelihood can be written as

$$216 \quad f_c(\mathbf{x}, \epsilon|N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^N \left[ \prod_{t=1}^T \prod_{j=1}^J p_{ijt}^{x_{ijt}} (1-p_{ijt})^{1-x_{ijt}} \times f_\epsilon(\epsilon_i|\boldsymbol{\eta}) \right],$$

217 where  $p_{ijt}$  is of the above form. The first term in the product over individuals corresponds to the  
 218 conditional likelihood associated with individual  $i$  (conditional on the individual random effect  
 219 terms) and the second term to the corresponding individual effect component.

220 The marginal data likelihood integrates out the  $\epsilon_i$  terms and can be expressed as,

$$221 \quad f_m(\mathbf{x}|N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^N \int_{\epsilon_i^{Mis}} \prod_{t=1}^T \prod_{j=1}^J p_{ijt}^{x_{ijt}} (1-p_{ijt})^{1-x_{ijt}} \times f_\epsilon(\epsilon_i|\boldsymbol{\eta}) d\epsilon_i^{Mis},$$

222 once more omitting the term  $\epsilon^{Mis} = \emptyset$ .

223 **2.1. Model fitting.** In the presence of individual heterogeneity leading to an analytically in-  
 224 tractable marginal data likelihood a range of different approaches have been proposed. These in-  
 225 clude a (classical) numerical integration approach, approximating the marginal data likelihood and  
 226 a (Bayesian) data augmentation approach using the complete data likelihood. For the particular  
 227 application to  $M_h$ -type models and SECR, see for example [Coull and Agresti \(1999\)](#); [Borchers and](#)  
 228 [Efford \(2008\)](#); [Gimenez and Choquet \(2010\)](#) (for a classical numerical integration approach) and  
 229 [Durban and Elston \(2005\)](#); [Royle et al. \(2007, 2009\)](#); [King and Brooks \(2008\)](#); [Royle and Dorazio](#)  
 230 [\(2012\)](#) (for Bayesian data augmentation approaches). We briefly describe the approaches in turn.

231 **2.1.1. Marginal data likelihood.** For a general individual heterogeneity model, the marginal data  
 232 likelihood may not be available in closed form (exceptions exist where the heterogeneity component  
 233 is described as a finite mixture model or infinite Beta distribution). In this case, the corresponding  
 234 likelihood is given in equation (2.1) as a product of integrals. For computational efficiency, we  
 235 are able to combine like terms in the likelihood corresponding to each unique encounter history  
 236 (corresponding to the combined capture history and observed individual heterogeneity values).  
 237 Notationally, let  $\Omega$  denote the set of possible encounter histories;  $\mathbf{x}_\omega$  the capture history for  $\omega \in \Omega$ ;  
 238  $\epsilon_\omega$  the individual heterogeneity terms for encounter history  $\omega \in \Omega$ ;  $\epsilon_\omega^{Mis}$  the unobserved individual  
 239 heterogeneity terms for encounter history  $\omega \in \Omega$  and  $n_\omega$  the number of individuals with encounter  
 240 history  $\omega$ . The marginal data likelihood can be expressed as,

$$241 \quad f_m(\mathbf{x}, \epsilon^{Obs}|N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{\omega \in \Omega} \left[ \int_{\epsilon_\omega^{Mis}} f_{\mathbf{x}}(\mathbf{x}_\omega|N, \boldsymbol{\theta}, \epsilon_\omega) f_\epsilon(\epsilon_\omega|N, \boldsymbol{\eta}) d\epsilon_\omega^{Mis} \right]^{n_\omega}.$$

242 Thus, this likelihood requires the estimation of a series of integrals each of dimension (at most)  
 243  $\dim(\mathbb{S})$ , where typically  $\dim(\mathbb{S})$  is small. For example, in the presence of  $q$  time invariant continuous  
 244 covariates,  $\dim(\mathbb{S}) = q$ , for model  $M_h$ ,  $\dim(\mathbb{S}) = 1$  and for the standard SECR model  $\dim(\mathbb{S}) =$   
 245  $2$  (see Examples 1-3 above). The number of integrals in the marginal data likelihood is equal  
 246 to the number of unique observed encounter histories plus one (corresponding to the encounter  
 247 history of not being observed). Each integral can, in general, be approximated using standard  
 248 integration techniques, such as Gauss-Hermite quadrature, grid-based approaches etc. Thus the  
 249 computational efficiency of this approach will be dependent on  $\dim(\mathbb{S})$  and the number of unique  
 250 encounter histories observed. For closed population models,  $\dim(\mathbb{S})$  is typically very small. This  
 251 (approximate) likelihood can be estimated using standard optimisation techniques to obtain the  
 252 associated MLEs of the model parameters.

253 2.1.2. *Complete data likelihood.* The Bayesian complete data likelihood approach specifies the  
 254 unobserved individual heterogeneity terms,  $\epsilon^{Mis}$ , as auxiliary variables (or additional parameters).  
 255 The joint posterior distribution of the parameters and auxiliary variables is then formed and given  
 256 by,

$$\begin{aligned}
 257 \quad \pi(N, \boldsymbol{\theta}, \boldsymbol{\eta}, \epsilon^{Mis} | \mathbf{x}, \epsilon^{Obs}) &\propto f_c(\mathbf{x}, \epsilon | N, \boldsymbol{\theta}, \boldsymbol{\eta}) p(N, \boldsymbol{\theta}, \boldsymbol{\eta}) \\
 258 \quad &= f_{\mathbf{x}}(\mathbf{x} | N, \boldsymbol{\theta}, \epsilon) f_{\epsilon}(\epsilon | N, \boldsymbol{\eta}) p(N, \boldsymbol{\theta}, \boldsymbol{\eta}),
 \end{aligned}$$

259 where  $f_c(\mathbf{x}, \epsilon | N, \boldsymbol{\theta}, \boldsymbol{\eta})$  denotes the complete data likelihood;  $f_{\mathbf{x}}(\mathbf{x} | N, \boldsymbol{\theta}, \epsilon)$  the conditional likelihood  
 260 of the observed data (conditional on the full set of individual heterogeneity terms);  $f_{\epsilon}(\epsilon | N, \boldsymbol{\eta})$  the  
 261 individual heterogeneity component; and  $p(N, \boldsymbol{\theta}, \boldsymbol{\eta})$  the prior density specified on  $N$ ,  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ . The  
 262 posterior density of only the model parameters,  $\pi(N, \boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{x}, \epsilon^{Obs})$ , is obtained by integrating out  
 263 over the auxiliary variables,  $\epsilon^{Mis}$ . However, the integration is analytically intractable so that an  
 264 MCMC approach is typically implemented, whereby we construct a Markov chain with stationary  
 265 distribution equal to the joint posterior distribution,  $\pi(N, \boldsymbol{\theta}, \boldsymbol{\eta}, \epsilon^{Mis} | \mathbf{x}, \epsilon^{Obs})$ , and subsequently  
 266 estimates of the marginal posterior summary statistics of interest are obtained.

267 An additional computational model fitting difficulty arises since  $\epsilon = \{\epsilon_1, \dots, \epsilon_N\}$  and hence  $\epsilon$   
 268 is itself a function of the unknown parameter,  $N$ . To address this issue [King and Brooks \(2008\)](#)  
 269 describe a reversible jump (RJ) MCMC algorithm for  $M_h$ -type models that is able to explore the  
 270 joint posterior distribution, where the number of parameters is able to vary within the constructed  
 271 Markov chain. This involved writing bespoke computer code. Alternatively, [Durban and Elston](#)  
 272 [\(2005\)](#); [Royle et al. \(2007, 2009\)](#); [Royle and Dorazio \(2012\)](#) use data augmentation techniques that  
 273 can be fitted in BUGS/JAGS. The underlying idea is to specify a super-population of size  $M$ , with  
 274 associated individual random effect terms  $\epsilon_i$  for  $i = 1, \dots, M$ . The encounter histories for individu-  
 275 als  $n + 1, \dots, M$  correspond to not being observed within the study. Within the MCMC algorithm,  
 276 the random effect term for each individual in this super-population is imputed in addition to a  
 277 binary indicator variable,  $z_i$  for  $i = 1, \dots, M$ , identifying which members of the super-population  
 278 are members of the target population of interest (by definition  $z_i = 1$  for  $i = 1, \dots, n$ , i.e. for  
 279 all individuals observed at least once within the study). This binary indicator variable has been  
 280 implemented using two different techniques each with different consequences. [Durban and Elston](#)  
 281 [\(2005\)](#) specify the binary variables, such that  $z_1, \dots, z_N = 1$  and  $z_{N+1}, \dots, z_M = 0$  (i.e. the indi-  
 282 cator variables are ordered); whereas [Royle et al. \(2007, 2009\)](#) do not induce any such structure  
 283 on the indicator variables relating to unobserved individuals, setting  $z_i = 1$  for  $i = 1, \dots, n$  and  
 284 modelling each indicator variable  $z_i$  for  $i = n + 1, \dots, M$ . The estimate of  $N$  is obtained as the sum  
 285 of non-zero indicator variables, i.e.  $N = \sum_{i=1}^M z_i$ . In other words [Durban and Elston \(2005\)](#) define  
 286 the indicator variables, conditional on  $N$ , whereas [Royle et al. \(2007, 2009\)](#) define  $N$ , conditional on  
 287 the indicator variables. For ease of reference we refer to the complete data likelihood data approach  
 288 of [Durban and Elston \(2005\)](#) as CD:DE (complete data: Durban and Elston) and of [Royle et al.](#)  
 289 [\(2007, 2009\)](#); [Royle and Dorazio \(2012\)](#) as CD:R (complete data: Royle).

290 Several issues arise with regard to these super-population data augmentation approaches. For  
 291 both approaches  $M$  needs to be specified and corresponds to an upper bound for the total population  
 292 size. This necessarily leads to a trade-off between the size specified for  $M$  and the computational  
 293 speed of the code. The larger the value of  $M$ , the greater the computational time due to the  
 294 imputation of the random effect term (and binary indicator variable for CD:R) for each individual in  
 295 the super-population. Too small a value for  $M$  will lead to a truncation of the posterior distribution  
 296 and biased inference. In addition, for CD:R, since  $N$  is derived as a deterministic function of the  
 297 indicator variables, it has a more limited prior specification (see Section 3.2 for further discussion  
 298 regarding prior specification). Alternatively for the approach of CD:DE, due to the more restricted  
 299 nature of the indicator variable specification, mixing issues can arise. To aid in the efficiency of  
 300 the computational algorithm [Durban and Elston \(2005\)](#) advocate the use of a pseudo-prior for



301 the corresponding random effect terms for individuals not in the population (i.e. for  $\epsilon_i$  for all  
 302  $i = N + 1, \dots, M$ ). The pseudo-prior is obtained from an initial MCMC run, using the estimated  
 303 posterior distribution for the random effect of an unobserved individual. For further discussion of  
 304 data augmentation techniques (particularly focusing on CD:R), see for example, [Link \(2013\)](#) and  
 305 [Schofield and Barker \(2014\)](#).

306 In general, without any prior information, the choice of analysis (classical marginal data likeli-  
 307 hood or Bayesian complete data likelihood) may be data dependent. In general, for a given dataset,  
 308 there is a computational trade-off between these different approaches. The marginal data likelihood  
 309 requires the numerical approximation of the integrals over the individual random effects; the com-  
 310 plete data likelihood is fast to evaluate but the individual random effects need to be updated  
 311 within the MCMC algorithm (using either RJMCMC or a super-population approach). To avoid  
 312 the use of explicitly approximating multiple integrals or the need to use a super-population or trans-  
 313 dimensional algorithm, we propose a hybrid semi-complete data likelihood approach. This involves  
 314 numerical integration for that part of the likelihood corresponding to unobserved individuals (as  
 315 in the marginal likelihood approach), while for the observed individuals any unobserved individual  
 316 heterogeneity terms are treated as auxiliary variables within a data augmentation approach (as in  
 317 the complete data likelihood approach). In this case, the number of auxiliary variables is known  
 318 so that the dimension of the parameter space is known and fixed. Standard BUGS/JAGS soft-  
 319 ware readily accommodates this approach, which involves approximation of only a single integral  
 320 of dimension  $\dim(\mathcal{S})$ . We describe this approach in more detail next.

321 **3. Semi-complete data likelihood.** We propose a *semi-complete* data likelihood approach,  
 322 combining the complete data likelihood for the individuals that are observed within the study (i.e.  
 323 individuals  $i = 1, \dots, n$ ), with a marginal data likelihood for the individuals that are not observed  
 324 within the study (i.e. individuals  $i = n + 1, \dots, N$ ). The semi-complete likelihood is expressed in  
 325 the form,

$$326 \quad f_s(\mathbf{x}, \boldsymbol{\epsilon}_{1:n} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) = f_{\mathbf{x}^*}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\epsilon}_{1:n}) f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_{1:n} | N, \boldsymbol{\eta})$$

327 where  $\boldsymbol{\epsilon}_{1:n} = \{\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n\}$ ;  $f_{\mathbf{x}^*}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\epsilon}_{1:n})$  denotes the likelihood of the capture histories condi-  
 328 tional on the model parameters ( $N$ ,  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$ ) and individual heterogeneity terms for the observed  
 329 individuals only ( $\boldsymbol{\epsilon}_{1:n}$ ); and  $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_{1:n} | N, \boldsymbol{\eta})$  the joint probability density function of the individ-  
 330 ual heterogeneity component for the observed individuals. Further, we have the following condi-  
 331 tional likelihood functions:  $f_{\mathbf{x}^*}(\mathbf{x}_{1:n} | N, \boldsymbol{\theta}, \boldsymbol{\epsilon}_{1:n})$  for the capture histories of the observed individuals  
 332 only, conditional on the model parameters and individual heterogeneity terms for the observed  
 333 individuals (dropping the dependence on  $\boldsymbol{\eta}$  since these are conditionally independent given  $\boldsymbol{\epsilon}_{1:n}$ );  
 334  $f_{\mathbf{x}^*}(\mathbf{x}_{n+1:N} | N, \boldsymbol{\theta}, \boldsymbol{\eta})$  for the capture histories of the unobserved individuals, conditional on the model  
 335 parameters; and  $f_{\mathbf{x}^*}(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\eta})$  for the capture history for unobserved individual  $i = n + 1, \dots, N$ ,  
 336 given the capture probability and individual heterogeneity model parameters (in the latter two  
 337 cases dropping the conditioning on  $\boldsymbol{\epsilon}_{1:n}$ ). Then, letting  $\mathbf{x}_{a:b} = \{\mathbf{x}_a, \dots, \mathbf{x}_b\}$ , we can express the  
 338 conditional likelihood in the form:

$$339 \quad \begin{aligned} f_{\mathbf{x}^*}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\epsilon}_{1:n}) &= f_{\mathbf{x}^*}(\mathbf{x}_{1:n} | N, \boldsymbol{\theta}, \boldsymbol{\epsilon}_{1:n}) f_{\mathbf{x}^*}(\mathbf{x}_{n+1:N} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) \\ 340 &\propto \frac{N!}{(N-n)!} \prod_{i=1}^n f_x(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\epsilon}_i) \times \prod_{i=n+1}^N f_{\mathbf{x}^*}(\mathbf{x}_i | \boldsymbol{\theta}) \\ 341 \quad (3.1) &= \prod_{i=1}^n f_x(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\epsilon}_i) \times \frac{N!}{(N-n)!} (1-p^*)^{N-n}, \end{aligned}$$

where  $1 - p^*$  denotes the probability of not being observed within the study (or conversely  $p^*$  denotes the probability of being seen at least once within the study) such that,

$$(3.2) \quad 1 - p^* = \int_{\boldsymbol{\epsilon}_\omega \in \mathbb{S}} f_x(\boldsymbol{\omega} = \mathbf{0} | \boldsymbol{\theta}, \boldsymbol{\epsilon}_\omega) f_\epsilon(\boldsymbol{\epsilon}_\omega | \boldsymbol{\eta}) d\boldsymbol{\epsilon}_\omega,$$

and  $\boldsymbol{\omega} = \mathbf{0}$  denotes the encounter history of a single individual who is unobserved within the study;  $f_x(\boldsymbol{\omega} = \mathbf{0} | \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\epsilon}_\omega)$  the conditional likelihood function associated with an individual not observed within the study and  $f_\epsilon(\boldsymbol{\epsilon}_\omega | \boldsymbol{\theta}, \boldsymbol{\eta})$  the probability density function of the associated individual heterogeneity terms for an individual not observed within the study. The product in equation (3.1) corresponds to the likelihood of the encounter histories, for an individual observed at least once within the study, conditional on the individual heterogeneity terms. The latter term corresponds to the contribution to the likelihood relating to the unobserved individuals.

An alternative (equivalent) model specification is given by

$$(3.3) \quad f_{\mathbf{x}^*}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\epsilon}_{1:n}) \propto \frac{1}{(p^*)^n} \prod_{i=1}^n f_x(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\epsilon}_i) \times \frac{N!}{(N-n)!} (p^*)^n (1-p^*)^{N-n},$$

where  $p^*$  is as above. The first term corresponds to the conditional likelihood of the observed capture histories, *given* that each of these individuals has been observed within the study and the corresponding individual heterogeneity terms. The second term corresponds to the Binomial probability of observing the number of individuals in the study, *given* the total population size.

We note that the semi-complete likelihood reduces to a single integral (over the dimension of the individual heterogeneity terms, i.e.  $\dim(\mathbb{S})$ ). This is in contrast to the marginal data likelihood which is a product of integrals (see Section 2.1.1), where the number of additional integrals corresponds to the number of unique encounter histories observed.

**3.1. Bayesian implementation.** Notationally, we let  $\boldsymbol{\epsilon}_{1:n}^{Obs}$  and  $\boldsymbol{\epsilon}_{1:n}^{Mis}$  denote the set of observed and unobserved individual heterogeneity terms for the observed individuals, respectively. The joint posterior distribution for the model parameters and unobserved individual heterogeneity terms for the observed individuals is given by,

$$(3.4) \quad \begin{aligned} \pi(N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\epsilon}_{1:n}^{Mis} | \mathbf{x}, \boldsymbol{\epsilon}_{1:n}^{Obs}) &\propto f_s(\mathbf{x}, \boldsymbol{\epsilon}_{1:n} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) p(N, \boldsymbol{\theta}, \boldsymbol{\eta}) \\ &= f_{\mathbf{x}^*}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\epsilon}_{1:n}) f_\epsilon(\boldsymbol{\epsilon}_{1:n} | N, \boldsymbol{\eta}) p(N, \boldsymbol{\theta}, \boldsymbol{\eta}), \end{aligned}$$

where  $f_s(\mathbf{x}, \boldsymbol{\epsilon}_{1:n} | N, \boldsymbol{\theta}, \boldsymbol{\eta})$  is the semi-complete data likelihood. Note that, as is typically the case, we assume that the priors specified for the total population size and model parameters are independent, so that  $p(N, \boldsymbol{\theta}, \boldsymbol{\eta}) = p(N)p(\boldsymbol{\theta})p(\boldsymbol{\eta})$ .

We use a standard Bayesian data augmentation approach for obtaining inference on the posterior distribution of interest,  $\pi(N, \boldsymbol{\theta}, \boldsymbol{\eta} | \mathbf{x}, \boldsymbol{\epsilon}_{1:n}^{Obs})$ . The number of auxiliary variables needed within this Bayesian data augmentation approach, using the semi-complete likelihood, is fixed and simply equal to  $|\boldsymbol{\epsilon}_{1:n}^{Mis}|$  (i.e. the auxiliary variables correspond to the number of unobserved individual heterogeneity terms of observed individuals). This is in contrast to the use of the joint posterior distribution of the model parameters and all unobserved individual heterogeneity terms,  $\boldsymbol{\epsilon}^{Mis}$ , given in equation (2.2), since  $\boldsymbol{\epsilon}^{Mis} = \{\boldsymbol{\epsilon}_{1:n}^{Mis}, \boldsymbol{\epsilon}_{n+1:N}\}$  where  $N$  is a parameter to be estimated. A number of different approaches have been proposed to fit individual heterogeneity models. These include trans-dimensional algorithms using reversible jump MCMC (King and Brooks, 2008), a joint posterior conditional MCMC algorithm (Fienberg *et al.*, 1999) for Rasch-type ( $M_{th}$ ) models and super-population data augmentation techniques. The first two approaches require bespoke code, while the super-population data augmentation approaches can be implemented within BUGS/JAGS (Durban and Elston, 2005; Royle *et al.*, 2007; Royle and Dorazio, 2012) but require the specification

of an upper bound  $M$  and imputation of the  $(M - n)$  individual random effect terms  $\epsilon_{n+1:M}$  (and dependent on the exact coding approach,  $M$  binary indicator variables).

Using the semi-complete data likelihood and corresponding posterior distribution given in equation (3.4), including only the heterogeneity terms for the observed individuals, permits standard (non-trans-dimensional) MCMC updating algorithms (such as the Metropolis-Hastings algorithm) to obtain inference on the parameters  $\theta$ ,  $\eta$  and  $N$ . However, the semi-complete data likelihood removes the necessity to impute the terms  $\epsilon_{n+1:M}$  and the need to specify an upper bound on the total population size, in general (see Section 3.2). Consequently, the models can be immediately fitted within BUGS/JAGS packages (see Section 4 for further discussion and King *et al.* (2015) for example JAGS code), with an explicit prior distribution specified on  $N$ . The trade-off of using the posterior distribution with semi-complete data likelihood, given in equation (3.4), is that the integral in equation (3.2) needs to be explicitly (numerically) estimated. However, in general this will be of very low dimension (often only one or two dimensions) for closed population models and so computationally fast and able to be accurately estimated (for example using Gaussian quadrature). We compare the complete data likelihood and semi-complete data likelihood approaches in Section 4 using JAGS for two different applications.

**3.2. Prior specification for  $N$ .** We briefly discuss possible prior distributions that are commonly specified on  $N$  and the corresponding Bayesian (and BUGS/JAGS) implementation. For the Bayesian data augmentation approach of Royle *et al.* (2007) (approach CD:R), the prior on  $N$  is only defined implicitly, given the prior specification on the indicator function relating to the probability that an individual in the super-population is a member of the population of interest, denoted  $\psi$ . The most common form of induced prior on  $N$  is the Uniform prior. However, Link (2013) showed that the uninformative prior  $\psi \sim U[0, 1]$  which induces the discrete uniform prior on  $N$  can lead to undesirable properties. Link (2013) therefore recommended the prior  $\psi \sim \text{Beta}(0.001, 1)$  which is easily implemented in BUGS/JAGS and induces an approximate Jeffreys' prior on  $N$ . More generally, specifying the prior  $\psi \sim \text{Beta}(a, b)$  induces the prior  $N \sim \text{Beta} - \text{Binomial}(M, a, b)$ , where  $M$  is the super-population upper bound. This is a fairly flexible prior structure, but the computational limitations with regard to specifying a suitable value of  $M$  remain.

For the complete data likelihood approach of Durban and Elston (2005) (approach CD:DE) and the semi-complete data likelihood approach an explicit prior is directly specified on  $N$ . Thus, any arbitrary distribution (specified on the set of non-negative integers) can be specified on the total population size. For example, Jeffreys' prior is a commonly used uninformative prior, given by  $p(N) \propto N^{-1}$  (see for example, Madigan and York (1997); King and Brooks (2008)). We note that specifying Jeffreys' prior, and using the semi-complete data likelihood expression given in equation (3.3) leads to a standard posterior conditional distribution for  $N$ , i.e.,

$$(N - n) | \mathbf{x}, \theta, \eta \sim \text{Neg} - \text{Bin}(n, p^*),$$

for  $p^*$  given in equation (3.1)<sup>1</sup>. Consequently, for Jeffreys' prior, the Gibbs sampler can be implemented for updating  $N$  within the MCMC algorithm. In general, if the prior or posterior conditional distribution for  $N$  is of (closed or) standard form this also simplifies the specification of the model in BUGS/JAGS, since this prior or posterior conditional distribution can be explicitly specified in the model component (see King *et al.* (2015) for sample JAGS code for the above Negative-Binomial posterior conditional distribution case). See also Fienberg *et al.* (1999) for further discussion.

<sup>1</sup>We use the form of the Negative Binomial distribution such that for  $X \sim \text{Neg} - \text{Bin}(n, q)$  the probability mass function is given by,

$$f(x) = \frac{(x + n - 1)!}{x!(n - 1)!} q^n (1 - q)^x.$$

This is the functional form of the distribution used with BUGS/JAGS.

426 Alternative prior distributions include  $p(N) \propto N^{-c}$  for some positive constant  $c$ , proposed by  
 427 [Fienberg \*et al.\* \(1999\)](#). For  $c > 1$  the tail of the distribution for  $N$  decays faster than for Jeffreys'  
 428 prior; while  $c < 1$  leads to a heavier tailed distribution. Alternatively, for an informative prior dis-  
 429 tribution for  $N$ , a Poisson or Poisson-Gamma (equivalently a Negative-Binomial) prior distribution  
 430 is often specified on  $N$  ([King and Brooks, 2001](#)). It can also be noted that specifying  $N \sim Po(\lambda)$   
 431 and  $\lambda \sim \Gamma(\delta, \delta)$  for small  $\delta$  provides another approximate Jeffreys' prior for  $N$ . These alternative  
 432 prior distributions are able to be implemented within BUGS/JAGS (typically using the zeros or  
 433 ones trick, [Lunn \*et al.\* \(2013\)](#) - see [King \*et al.\* \(2015\)](#) for associated sample JAGS code).

434 **3.3. Special case.** We note that the approach presented by [Bonner and Schofield \(2014\)](#) is  
 435 a special case of the semi-complete data likelihood approach applied to a covariate model. In  
 436 particular, [Bonner and Schofield \(2014\)](#) consider a time invariant individual covariate model given  
 437 in Example 1 of Section 2. Using the terminology presented above, so that the notation differs  
 438 to that given in [Bonner and Schofield \(2014\)](#), they describe the particular case where  $\epsilon^{Obs} = \epsilon_{1:n}$   
 439 and  $\epsilon^{Mis} = \epsilon_{n+1:N}$ . In other words, the individual heterogeneity terms are known for individuals  
 440 observed within the study (though it is implied in their discussion that the approach is more  
 441 generally applicable). The posterior distribution is then formed analogous to Equation (3.4). The  
 442 probability of not being observed within the study, given in Equation (3.2) is estimated using Monte  
 443 Carlo integration.

444 **4. Examples.** We consider two real examples, relating to model  $M_h$  and SECR, described  
 445 in Section 2. We note that as with all performance metrics for comparing the efficiency of differ-  
 446 ent model-fitting approaches these are dependent on numerous factors, such as the programming  
 447 language, specific application, data, model specification (including the pseudo-priors specified for  
 448 the super-population approach), initial starting values and machine used. In order to be able to  
 449 draw sensible comparisons for each example we present results obtained from same machine and  
 450 language using the JAGS codes provided in [King \*et al.\* \(2015\)](#).

451 **4.1. Model  $M_h$  - snowshoe hares.** To demonstrate our proposed semi-complete data likelihood  
 452 approaches for model  $M_h$ , we revisit the snowshoe hare data originally examined in the seminal  
 453 paper of [Otis \*et al.\* \(1978\)](#) and subsequently analyzed by many others (for example [Coull and](#)  
 454 [Agresti, 1999](#); [Dorazio and Royle, 2003](#); [Royle \*et al.\*, 2007](#); [Link, 2013](#)). Over  $T = 6$  days of  
 455 trapping,  $n = 68$  hares were captured with observed frequencies  $\mathbf{n} = (25, 22, 13, 5, 1, 2)'$  where  
 456  $n_t = \sum_{i=1}^n I(y_i = t)$  and  $y_i = \sum_{j=1}^T x_{ij}$  for  $t = 1, \dots, T$ . We assume  $\text{logit}(p_{it}) = \alpha + \epsilon_i$  and  
 457  $\epsilon_i \sim N(0, \sigma^2)$  for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ , with  $\boldsymbol{\theta} = \{\alpha\}$  and  $\boldsymbol{\eta} = \{\sigma^2\}$ .

458 We fit the semi-complete data likelihood and complete data likelihood Bayesian super-population  
 459 (CD:R and CD:DE) approaches in R ([R Core Team, 2014](#)) using the `rjags` package ([Plummer,](#)  
 460 [2013](#)) - see Appendix A of [King \*et al.\* \(2015\)](#) for the associated JAGS code. For each analysis we  
 461 specify the priors,  $\alpha \sim N(0, 100)$  and  $\sigma^2 \sim \Gamma^{-1}(0.01, 0.01)$ . We specify Jeffreys' prior for  $N$ , for  
 462 the semi-complete data likelihood and CD:DE. For ease of comparison with CD:R we set  $\psi \sim$   
 463  $Beta(0.001, 1)$ , which induces an approximate (truncated) Jeffreys' prior for  $N$  on  $1, \dots, M$  ([Link,](#)  
 464 [2013](#)). We note that we consider two JAGS specifications for the semi-complete data likelihood. The  
 465 first approach (SCD1) uses the Jeffreys' prior specification for  $N$  explicitly in the model component  
 466 of the code. However, since Jeffreys' prior is improper we need to specify an upper bound for  $N$ ,  
 467 which we again denote by  $M$  (essentially this is a truncated Jeffreys' prior at  $M$ ). The second  
 468 approach (SCD2) specifies the (predictive) posterior conditional distribution for  $N - n$ , which is of  
 469 Negative-Binomial form (see Section 3.2).

470 Following [Link \(2013\)](#), we specify an upper bound of  $M = 1000$  for the maximum total population  
 471 size for the complete data likelihood super-population approaches and the first semi-complete data  
 472 likelihood approach (SCD1) in JAGS. For the semi-complete data likelihood approach, the integral

TABLE 1

Posterior summaries for snowshoe hare abundance ( $N$ ) under model  $M_h$  using the semi-complete data likelihood (SCD) approach, CD:R and CD:DE. The semi-complete data likelihood approaches correspond to specifying the prior for  $N$  (SCD1) and the posterior conditional distribution for  $N - n$  (SCD2) in the model component of the JAGS code. For SCD1, CD:R and CD:DE, we specify an upper limit of  $M = 1000$ . Effective sample size (ESS) and effective sample size per second (ESS/s) are included for each approach. A total of 30 million iterations are used in each case with the realisations thinned by 10.

method	mean	median	SD	95% CI	ESS	ESS/s
SCD1	100.3	93	32.8	(74, 171)	168347	7.67
SCD2	101.1	93	74.9	(74, 173)	167680	7.74
CD:R	100.6	93	32.7	(74, 171)	13080	0.10
CD:DE	101.3	93	36.2	(74, 178)	9626	0.03

473 in Equation (3.2) is evaluated using Gauss-Hermite quadrature:

$$474 \quad (4.1) \quad 1 - p^* \approx \sum_{j=1}^q \frac{w_j}{\sqrt{\pi} [1 + \exp(\sqrt{2}\sigma v_j + \alpha)]^T},$$

475 where  $w_j$  and  $v_j$  are the weights and nodes corresponding to  $q$  quadrature points (sensu McClintock  
476 *et al.*, 2009). The degree of accuracy of this approximation increases with  $q$ , and larger  $q$  is required  
477 for larger  $\sigma$ . For our analyses, we specify  $q = 100$ .

478 For each approach, we ran three chains of 10 million iterations (after initial pilot tuning and  
479 burn-in) from overdispersed starting values, thinning the realisations by 10 for memory storage  
480 purposes. Chain convergence was assessed based on visual inspection and Brooks-Gelman-Rubin  
481 diagnostics (no lack of convergence was identified). On a computer running 64-bit Windows 7  
482 (3.4GHz Intel Core i7 processor, 16Gb RAM), the analyses required about 6.1 hrs for the first  
483 semi-complete data likelihood (prior distribution for  $N$  specified) approach, 6.0 hrs for the second  
484 semi-complete data likelihood (posterior conditional distribution for  $N - n$  specified), 35.1 hrs for  
485 CD:R and 83.3 hours for CD:DE. We note that the run times should be interpreted comparatively,  
486 as they will in general differ across different computers as a result of different processors, operating  
487 systems etc. The marginal posterior summaries are provided in Table 1, coupled with the effective  
488 sample sizes (per second) for each approach.

489 Although setting  $M = 1000$  may appear conservative, this did appear to influence the skewness  
490 of the right tail of the marginal posterior distribution for  $N$  relative to the (unbounded) posterior  
491 distribution for  $N$  when using the second semi-complete data likelihood approach (SCD2). We  
492 therefore reran the first semi-complete data likelihood (SCD1) analysis with  $M = 10000$  leading  
493 to posterior summary results more similar to the second complete data likelihood approach ( $N$   
494 posterior mean = 100.9, median = 93, SD = 56.1, 95% credible interval (CI) = (74, 172)), but with  
495 noticeably reduced effective sample size (ESS = 74928) and increased computation time (ESS/s =  
496 2.81). Nevertheless, specifying larger  $M$  for the first semi-complete data likelihood approach comes  
497 at considerably less computational cost compared to the super-population complete data likelihood  
498 approaches (CD:R and CD:DE). Avoidance of the need to specify  $M$  when using BUGS/JAGS  
499 remains an advantage of the general semi-complete data likelihood approach (this is true even  
500 when using Jeffreys' prior on  $N$  by specifying the posterior conditional distribution for  $N - n$  in  
501 the model component of the code).

502 For approach SCD1, using an explicit Negative-Binomial or Beta-Binomial approximation to  
503 Jeffreys' prior (code is provided in Appendix A of King *et al.* (2015)) unsurprisingly lead to similar  
504 results in terms of ESS and ESS/s as for the use of the explicit (truncated) Jeffreys' prior. However,  
505 within the model specification code, using the distributions' hierarchical form where an auxiliary  
506 variable is introduced for the Poisson mean or Binomial probability and imputed within the MCMC  
507 algorithm lead to lower ESS and ESS/s as a result of poorer mixing due to posterior correlation

508 between parameters. We do not consider these prior specifications further.

509 Finally, we note that  $q = 100$  appeared to be sufficient in the Gauss-Hermite quadrature approach  
 510 for these analyses, but in general proper specification of  $q$  will be case dependent. For example,  
 511 using our estimated posterior median  $\alpha = -1.2$  and the 99.9% quantile  $\sigma = 3.3$ , Equation (4.1)  
 512 with  $q = 100$  is accurate to a precision of five decimal places. However, for  $\sigma = 10$ ,  $q = 100$  it is only  
 513 accurate to two decimal places. Care must therefore be taken when specifying  $q$  using the semi-  
 514 complete data likelihood approach in JAGS. If computation speed is of little concern Equation (3.2)  
 515 could alternatively be approximated in OpenBUGS using the inbuilt `integral` function, which also  
 516 has an inbuilt default value for  $q$ .

517 *4.2. Model SECR - gibbons.* To illustrate the proposed semi-complete data likelihood approach  
 518 in the context of SECR models we use acoustic survey data from a population of northern yellow-  
 519 cheeked gibbon from northeastern Cambodia. These data were collected from 13 replicate survey  
 520 locations, each consisting of a 3 by 1 linear array of listening posts spaced 0.5km apart. Each  
 521 listening post was manned by a single human observer who recorded the timing of calls at each and  
 522 an estimated compass bearing to each detected gibbon group. Recaptured groups were determined  
 523 using the estimated bearings and detection times. Over  $T = 1$  survey days a total of  $n = 77$  gibbon  
 524 groups were detected across the 13 arrays. We specify the half-normal function for  $g$  of the form,

$$525 \quad p_{ijt} = \exp\left(-\frac{\|\mathbf{u}_j - \boldsymbol{\epsilon}_i\|^2}{2\sigma^2}\right).$$

526 For each analysis we specify the prior  $\sigma \sim U[0, 10]$  and assume that the home range centres are  
 527 uniformly distributed over the given area, i.e.  $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_i | \boldsymbol{\eta}) = \frac{1}{A}$  where  $A$  is the area of  $\mathbb{S}$  for each  
 528  $i = 1, \dots, N$  (in this case  $A = 546\text{km}^2$ ). Thus we set  $\psi \sim \text{Beta}(0.001, 1)$  for the super-population  
 529 approach CD:R and Jeffreys' prior for  $N$  for the complete data likelihood (CD:DE) semi-complete  
 530 likelihood approaches.

531 As in Section 4.1 we fit both forms of the semi-complete data likelihood (Equations (3.1) and  
 532 (3.3)) and the super-population complete data likelihood Bayesian approaches CD:R and CD:DE  
 533 using the `rjags` package (see Appendix B of King *et al.* (2015) for sample JAGS code). For the  
 534 complete data likelihood approaches and first semi-complete data likelihood (specifying Jeffreys'  
 535 prior on  $N$  within the model component of the JAGS code) we specify an upper bound of  $M = 1000$   
 536 for the discrete support of  $N$ . For both semi-complete likelihoods the integral in Equation (3.2) was  
 537 approximated by a summation over a rectangular grid of 4200 points. Note that a suitable choice  
 538 of grid will be case dependent, with increases in accuracy resulting from greater spatial extents and  
 539 decreased distances between neighbouring grid points, but at the expense of computational time.  
 540 An exploratory analysis suggested that the grid used was relatively conservative, achieving good  
 541 numerical accuracy.

542 To compare the performance of the different approaches, each MCMC algorithm is run for 500,000  
 543 iterations, following a burn-in period of 10,000 iterations (no lack of convergence was identified for  
 544 simulations of this length). On a computer running Windows Server 2008 R2 Enterprise (3.1GHz  
 545 Intel Xeon CPU E5-2687, 256Gb RAM), the analyses required about 46.6 minutes for the first  
 546 semi-complete data likelihood (SCD1; specifying (truncated) Jeffreys' prior on  $N$  in the model  
 547 component) approach, 42.3 minutes for the second semi-complete data likelihood (SCD2; specifying  
 548 the posterior conditional distribution for  $N - n$ ), 2.5 hours for CD:R and 6.8 hours for CD:DE. As  
 549 for the snowshoe hare example, marginal posterior summaries were similar for all parameters using  
 550 all approaches, but the semi-complete data likelihood approaches required far less computation  
 551 time and yielded greater effective sample sizes than the data-augmented complete data likelihood  
 552 approaches (Table 2).

TABLE 2

Posterior summaries for gibbon group abundance ( $N$ ) under the SECR models using the semi-complete data likelihood (SCD) approach, CD:R and CD:DE. The semi-complete data likelihood approaches correspond to specifying the prior for  $N$  (SCD1) and the posterior conditional distribution for  $N - n$  (SCD2) in the model component of the JAGS code. For SCD1, CD:R and CD:DE, we specify an upper limit of  $M = 1000$ . Effective sample size (ESS) and effective sample size per second (ESS/s) are included for each approach. A total of 500,000 iterations are used in each case.

model	mean	median	SD	95% CI	ESS	ESS/s
SCD1	357.1	328	176.2	(119, 766)	2763	1.01
SCD2	357.7	327	178.4	(120, 775)	3872	1.56
CD:R	355.3	326	176.9	(118, 768)	865	0.09
CD:DE	362.7	338	173.2	(122, 765)	622	0.03

553 **5. Discussion.** For closed population models, the semi-complete data likelihood specifies the  
554 joint probability density function of the model parameters and associated unobserved individual  
555 heterogeneity terms for only those individuals observed, conditional on the observed capture histo-  
556 ries and observed individual heterogeneity components. This likelihood is specified as an integral of  
557 the individual heterogeneity component for the unobserved individuals. The integral is analytically  
558 intractable but of dimension equal to the dimension of the individual heterogeneity component of  
559 the model, and hence typically small. This permits the the use of standard (efficient) numerical ap-  
560 proximation techniques to estimate the integral (for example, in OpenBUGS, the inbuilt `integral`  
561 function can be used to conduct one dimensional integration; with similar inbuilt functions in R for  
562 one or multi-dimensional integrals). The semi-complete data likelihood approach can be applied to  
563 a range of different individual heterogeneity models.

564 Using this semi-complete data likelihood within a Bayesian analysis of closed capture-recapture  
565 data in the presence of individual heterogeneity, removes the need for trans-dimensional algo-  
566 rithms to explore the posterior distribution of the parameters due to the “unknown number of  
567 parameters” problem. Consequently, the models can be fitted efficiently in standard software, such  
568 as BUGS/JAGS without using a super-population approach. The semi-complete data likelihood  
569 approach is significantly more efficient than the previous super-population approaches, as demon-  
570 strated in Section 4, where the improvement for the examples that we considered using the codes  
571 provided in King *et al.* (2015) was up to two orders of magnitude. The improvement is in terms  
572 of both computational time and effective sample sizes (as a result of improved mixing within the  
573 MCMC algorithm). The efficiency of the super-population approaches is heavily dependent on the  
574 upper limit specified for the super-population,  $M$ . This makes the Bayesian approach feasible for  
575 fitting to a significantly wider range of data, particularly for spatially explicit capture-recapture,  
576 where the use of a Bayesian data augmentation technique can be particularly inefficient. In general,  
577 the ESS and ESS/s for the different approaches is dependent on numerous factors including the  
578 exact form of the model specification, the pseudo-priors specified in the super-population approach,  
579 initial starting values and computer on which the simulations are being run.

580 This semi-complete data approach has been developed for closed population models in the pres-  
581 ence of individual heterogeneity. As discussed in Example 1 of Section 2 the inclusion of additional  
582 observable individual level covariates is immediate and can be seen to be a generalisation of the  
583 Monte Carlo in MCMC approach proposed by Bonner and Schofield (2014) (see Section 3.3). The  
584 individual heterogeneity terms correspond to the covariate values and are typically known when  
585 individuals are observed, though this need not be the case (missing covariate values for individuals  
586 observed within the study can again be treated as auxiliary variables within the complete data like-  
587 lihood component). In the presence of time-varying continuous individual covariates the increase  
588 in dimension of the necessary integral in the associated marginal data likelihood can be reduced by  
589 efficiently approximating the underlying state process as a hidden Markov model (Langrock and  
590 King, 2013). The approach can also be immediately applied to other forms of data. For example,

591 these include stopover models permitting arrivals to, and departures from, the study population  
 592 (Pledger *et al.*, 2009) and conventional distance sampling (Buckland *et al.*, 2001). For the latter  
 593 case the capture history is a univariate binary term (1 if an individual is observed and 0 if un-  
 594 observed), the individual heterogeneity component is the perpendicular distance of the individual  
 595 from the line/point transect (known for observed individuals), assumed to have a uniform distribu-  
 596 tion (for line transects) or triangular distribution (for point transects), see for example, Equation  
 597 (7.10) on page 141 of Borchers *et al.* (2002). Further work lies in identifying and developing similar  
 598 approaches for different forms of data. In addition, for more general Bayesian analyses, highly cor-  
 599 related parameters often leads to inefficient MCMC algorithm, due to poor mixing. To address this  
 600 issue, a reparameterisation may often be used and/or block-updates implemented. An alternative  
 601 approach, motivated by this semi-complete data approach, would be to identify and integrate out  
 602 (using a numerical approximation) the highly correlated parameters. This is an area of current  
 603 research.

604 **Acknowledgments.** The authors would like to thank two referees and an AE for helpful  
 605 comments on an earlier version of the manuscript. We would also like to thank Dr Benjamin  
 606 Rawson (Conservation International, Cambodia) for permission to use the gibbon data analysed in  
 607 Section 4.2 and Dr John Durban and Prof David Elston for providing sample BUGS code for the  
 608 CD:DE approach. This work was part-funded by EPSRC grant EP/I000917/1. The findings and  
 609 conclusions in the paper are those of the author(s) and do not necessarily represent the views of  
 610 the National Marine Fisheries Service, NOAA. Any use of trade, product or firm names does not  
 611 imply an endorsement by the US Government.

#### SUPPLEMENTARY MATERIAL

612 **Supplement: Supplement to “Capture-recapture abundance estimation using a semi-  
 613 complete data likelihood approach”**  
 614 (doi: [COMPLETED BY THE TYPESETTER](#)). The supplement consists of Appendices A and B  
 615 that provide sample JAGS codes for the examples provided in the text using the different model-  
 616 fitting algorithms (referenced in Sections 3.2, 4.1 and 4.2).

#### References.

- 618 Borchers, D. L., Buckland, S. T. and Zucchini W. (2002) *Estimating Animal Abundance, Closed Populations*. London:  
 619 Springer. 314 pp.
- 620 Borchers, D. L. and Efford, M. G. (2008) Spatially explicit maximum likelihood methods for capture-recapture studies.  
 621 *Biometrics*, **64**, 377-85
- 622 Bonner, S. and Schofield, M. R. (2014) MC(MC)MC: exploring Monte Carlo integration within MCMC for mark-  
 623 recapture models with individual covariates. *Methods in Ecology and Evolution*, in press
- 624 Buckland, S. T., Anderson D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. and Thomas, L. (2001) Introduction  
 625 to distance sampling. Oxford: Oxford University Press
- 626 Coull, B. and Agresti, A. (1999) The use of mixed logit models to reflect heterogeneity in capture-recapture studies.  
 627 *Biometrics*, **55**, 294-301
- 628 Dorazio, R. M. and Royle, J. A. (2003) Mixture models for estimating the size of a closed population when capture  
 629 rates vary among individuals. *Biometrics* **59**, 351-64
- 630 Durban, J. W. and Elston, D. A. (2005) Mark-recapture with occasion and individual effects: abundance estimation  
 631 through Bayesian model selection in a fixed dimensional parameter space. *Journal of Agricultural, Biological, and*  
 632 *Environmental Statistics* **10**, 291-305
- 633 Fienberg, S. E., Johnson, M. S. and Junker, B. W. (1999) Classical multilevel and Bayesian approaches to population  
 634 size estimation using multiple lists. *Journal of the Royal Statistical Society: Series A* **162**, 383-405
- 635 Gimenez, O. and Choquet, R. (2010) Individual heterogeneity in studies on marked animals using numerical integra-  
 636 tion: capturerecapture mixed models. *Ecology* **91**, 951-957
- 637 King, R. and Brooks, S. P. (2001) On the Bayesian analysis of population size. *Biometrika*, **88**, 317-336
- 638 King, R. and Brooks, S. P. (2008) On the Bayesian estimation of a closed population size in the presence of hetero-  
 639 geneity and model uncertainty. *Biometrics*, **64**, 816-824



- 640 King, R., Bird, S. M., Overstall, A. M., Hay, G. and Hutchinson, S. J. (2014) Estimating prevalence of injecting drug  
 641 users and associated death rates in England using regional data and incorporating prior information. *Journal of*  
 642 *the Royal Statistical Society: Series A*, **77**, 209-236
- 643 King, R., McClintock, B., Kidney, D. and Borchers, D. L. (2015). Supplement to "Capture-recapture abundance  
 644 estimation using a semi-complete data likelihood approach".
- 645 King, R., Morgan, B. J. T., Gimenez, O. and Brooks, S. P. (2009) *Bayesian Analysis for Population Ecology* Boca  
 646 Raton: CRC Press
- 647 Langrock, R. and King, R. (2013) Maximum likelihood estimation of mark-recapture-recovery models in the presence  
 648 of continuous covariates. *Annals of Applied Statistics*, **7**, 1709-1732
- 649 Link, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial  $N$ . *Ecology*, **94**, 2173-2179
- 650 Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2013) *The BUGS Book: A Practical Introduction*  
 651 *to Bayesian Analysis*. Boca Raton: CRC Press
- 652 Madigan, D. and York, J. C. (1997) Bayesian methods for estimation of the size of a closed population. *Biometrika*,  
 653 **84**, 19-31
- 654 Manrique-Vallier, D. and Fienberg, S. E. (2008) Population size estimation using individual level mixture models.  
 655 *Biometrical Journal*, **50**, 1051-63
- 656 McClintock, B.T., White, G.C., Burnham, K.P. and Pryde, M.A. (2009) A generalized mixed effects model of abun-  
 657 dance for mark-resight data when sampling is without replacement. *Modeling Demographic Processes in Marked*  
 658 *Populations* (eds D.L. Thomson, E.G. Cooch and M.J. Conroy), pp. 271-289. Springer, New York, USA.
- 659 McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of Capture-Recapture Data*. Boca Raton: CRC Press.
- 660 Morgan, B. J. T. and Ridout, M. S. (2008) A new mixture model for capture heterogeneity. *Journal of the Royal*  
 661 *Statistical Society: Series C* **57**, 433-46
- 662 Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978), Statistical inference from capture data on  
 663 closed animal populations. *Wildlife Monographs* **62**, 1-135
- 664 Overstall, A. M., King, R., Bird, S. M., Hutchinson, S. J. and Hay, G. (2014) Estimating the number of people who  
 665 inject drugs in Scotland using multi-list data with left censoring. *Statistics in Medicine* **33**, 1564-79
- 666 Pledger S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Bio-*  
 667 *metrics*, **56**, 434-42
- 668 Pledger S. (2005) The performance of mixture models in heterogeneous closed population capture-recapture. *Bio-*  
 669 *metrics*, **61**, 868-73
- 670 Pledger, S., Efford, M., Pollock, K. H., Collazo, J. A. and Lyons, J. E. (2009), Stopover duration analysis with depar-  
 671 ture probability dependent on unknown time since arrival. *Modeling Demographic Processes in Marked Populations*  
 672 (eds D.L. Thomson, E.G. Cooch and M.J. Conroy), pp. 349-363. Springer, New York, USA.
- 673 Plummer, M. (2013) rjags: Bayesian graphical models using MCMC. R package version 3-11
- 674 R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Com-  
 675 puting, Vienna, Austria. URL <http://www.R-project.org/>.
- 676 Royle, J. A. and Dorazio, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture-  
 677 recapture models. *Journal of Ornithology*, **152**, S521-37.
- 678 Royle J. A., Dorazio R. M. and Link W. A. (2007) Analysis of multinomial models with unknown index using data  
 679 augmentaton. *Journal of Computational and Graphical Statistics*, **16**, 67-85
- 680 Royle J. A., Karanth K. U., Gopalaswamy A. M. and Kumar N. S. (2009) Bayesian inference in camera trapping  
 681 studies for a class of spatial capture-recapture models. *Ecology* **90**, 3233-44
- 682 Schofield, M. R. and Barker, R. J. (2014) Hierarchical modeling of abundance in closed population capture-recapture  
 683 models under heterogeneity. *Environmental and Ecological Statistics* **21**, 435-51
- 684 Williams, B. K., Nichols, J. D. and Conroy, M. J. (2002) *Analysis and Management of Animals Populations*. *Academic*  
 685 *Press*

686 ADDRESS OF KING  
 SCHOOL OF MATHEMATICS, UNIVERSITY OF EDINBURGH,  
 JAMES CLERK MAXWELL BUILDING,  
 THE KING'S BUILDINGS, PETER GUTHRIE TAIT ROAD,  
 EDINBURGH, UK. EH9 3FD.

ADDRESS OF MCCLINTOCK  
 NATIONAL MARINE MAMMAL LABORATORY,  
 ALASKA FISHERIES SCIENCE CENTER,  
 NOAA NATIONAL MARINE FISHERIES SERVICE,  
 7600 SAND POINT WAY NE, SEATTLE,  
 WASHINGTON 98115 USA

687 ADDRESS OF KIDNEY AND BORCHERS  
 CENTRE FOR RESEARCH INTO ECOLOGICAL AND ENVIRONMENTAL MODELLING  
 AND SCHOOL OF MATHEMATICS AND STATISTICS,  
 THE OBSERVATORY, BUCHANAN GARDENS,  
 UNIVERSITY OF ST ANDREWS, ST ANDREWS, UK. KY16 9LZ.