CAPTURE–RECAPTURE ABUNDANCE ESTIMATION USING A SEMI-COMPLETE DATA LIKELIHOOD APPROACH[1]

BY RUTH KING[*], BRETT T. MCCLINTOCK[†],
DARREN KIDNEY[‡] AND DAVID BORCHERS[‡]

University of Edinburgh[*], NOAA National Marine Mammal Laboratory[†] and University of St Andrews[‡]

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 superpopulation 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 data sets: 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, Johnson and Junker (1999)], disease prevalence [Manrique-Vallier and Fienberg (2008)] and animal populations [Borchers, Buckland and Zucchini (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, Buckland and Zucchini (2002), Williams, Nichols and Conroy (2002) and McCrea and Morgan (2015).

In general, the likelihood of capture–recapture data can be expressed in multi-nomial 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), behavioral (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 maximized 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), Pledger (2005)]. 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) numeral integration to estimate the *marginal* (or observed) data likelihood [Borchers and Efford (2008), Coull and Agresti (1999), 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), King and Brooks (2008), King et al. (2009), Royle and Dorazio (2012), Royle,

Dorazio and Link (2007), Royle et al. (2009)]. 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, Johnson and Junker (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 data set 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 $i = 1, \ldots, n$ and the unobserved individuals $i = n + 1, \ldots, N$. Let p_{it} denote the capture probability of individual $i = 1, \ldots, N$ at time $t = 1, \ldots, T$. Further, for standard capture–recapture data, $\mathbf{x}_i = \{x_{it} : t = 1, \ldots, T\}$ denotes the capture history of individual $i = 1, \ldots, N$, such that

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

We consider individual heterogeneity specified such that

$$p_{it} = g(\boldsymbol{\theta}, \boldsymbol{\varepsilon}_i),$$

for some function g, where $\boldsymbol{\theta}$ denotes the model parameters associated with the capture probabilities (which may include, for example, temporal and/or behavioral effect terms, regression coefficients for covariate values, etc.) and $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_i : i =$

$1, \ldots, N\}$ such that $\boldsymbol{\varepsilon}_i \in \mathbb{S} \subset \mathbb{R}^k$, corresponding to the individual heterogeneity term for individual $i = 1, \ldots, N$. Further, we assume an underlying model for the individual heterogeneity, such that $\boldsymbol{\varepsilon}$ is a function of the parameters $\boldsymbol{\eta}$, and that the individual heterogeneity terms, $\boldsymbol{\varepsilon}_i$, are independent of each other conditional on $\boldsymbol{\eta}$. The associated joint probability density function of the heterogeneity terms is given by $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|N, \boldsymbol{\eta}) = \prod_{i=1}^{N} f_{\varepsilon}(\boldsymbol{\varepsilon}_i|\boldsymbol{\eta})$, using the conditional independence assumption (and dropping the dependence on N for the conditional density function of the individual heterogeneity terms for individual i). Further, to provide a general framework for both observed and unobserved individual heterogeneity, we additionally write $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}^{\text{Obs}}, \boldsymbol{\varepsilon}^{\text{Mis}}\}$, where $\boldsymbol{\varepsilon}^{\text{Obs}}$ denotes the set of observed individual heterogeneity components and $\boldsymbol{\varepsilon}^{\text{Mis}}$ the set of unobserved individual heterogeneity components. Similarly, we write $\boldsymbol{\varepsilon}_i = \{\boldsymbol{\varepsilon}_i^{\text{Obs}}, \boldsymbol{\varepsilon}_i^{\text{Mis}}\}$, for $i = 1, \ldots, N$ with obvious interpretation. Finally, we assume that the capture histories of the individuals are independent of each other given the capture probability model parameters, $\boldsymbol{\theta}$, and individual heterogeneity terms, $\boldsymbol{\varepsilon}$.

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

$$f_m(\mathbf{x}, \boldsymbol{\varepsilon}^{\text{Obs}}|N, \boldsymbol{\theta}, \boldsymbol{\eta}) = \int_{\boldsymbol{\varepsilon}_1^{\text{Mis}}} \cdots \int_{\boldsymbol{\varepsilon}_N^{\text{Mis}}} f_c(\mathbf{x}, \boldsymbol{\varepsilon}|N, \boldsymbol{\theta}, \boldsymbol{\eta}) \, d\boldsymbol{\varepsilon}_N^{\text{Mis}} \cdots d\boldsymbol{\varepsilon}_1^{\text{Mis}}$$

$$(2.1) \qquad = \int_{\boldsymbol{\varepsilon}_1^{\text{Mis}}} \cdots \int_{\boldsymbol{\varepsilon}_N^{\text{Mis}}} f_{\mathbf{x}}(\mathbf{x}|N, \boldsymbol{\theta}, \boldsymbol{\varepsilon}) f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|N, \boldsymbol{\eta}) \, d\boldsymbol{\varepsilon}_N^{\text{Mis}} \cdots d\boldsymbol{\varepsilon}_1^{\text{Mis}}$$

$$\propto \frac{N!}{(N-n)!} \prod_{i=1}^{N} \int_{\boldsymbol{\varepsilon}_i^{\text{Mis}}} f_x(\mathbf{x}_i|\boldsymbol{\theta}, \boldsymbol{\varepsilon}_i) f_{\varepsilon}(\boldsymbol{\varepsilon}_i|\boldsymbol{\eta}) \, d\boldsymbol{\varepsilon}_i^{\text{Mis}},$$

using the multinomial distributional form of the capture–recapture data (omitting the constant multinomial coefficients for simplicity) and conditional independence of the random effect terms. The term $f_c(\mathbf{x}, \boldsymbol{\varepsilon}|N, \boldsymbol{\theta}, \boldsymbol{\eta})$ corresponds to the *complete* data likelihood (i.e., the joint probability density function of the capture histories and individual effects), $f_{\mathbf{x}}(\mathbf{x}|N, \boldsymbol{\theta}, \boldsymbol{\varepsilon})$ the conditional likelihood of the capture histories (where the conditioning includes the individual heterogeneity terms), and $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|N, \boldsymbol{\eta})$ the joint probability density function of the individual heterogeneity terms. The term $f_x(\mathbf{x}_i|\boldsymbol{\theta}, \boldsymbol{\varepsilon}_i)$ corresponds to the conditional likelihood of the capture history for individual $i = 1, \ldots, N$, and $f_{\varepsilon}(\boldsymbol{\varepsilon}_i|\boldsymbol{\eta})$ the conditional probability density function of the individual heterogeneity component for individual $i = 1, \ldots, N$ (where in each case we drop the dependence on N).

EXAMPLE 1 (Continuous individual covariates). We consider the case with q time-invariant continuous individual covariates $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N\}$, where $\boldsymbol{\varepsilon}_i \in \mathbb{S} \subseteq \mathbb{R}^q$ denotes the covariate values associated with individual $i = 1, \ldots, N$. Since the covariate values are time-invariant, the associated capture probabilities for each individual are also time-invariant, so that $p_{it} = p_i$ for $t = 1, \ldots, T$. Assuming that

the capture probabilities are linearly related to the covariate values via some link function, we may specify

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

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

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

The complete data likelihood is of the form

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

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

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 observed). The first term of the complete data likelihood corresponds to the conditional likelihood (conditional on the individual covariate terms) and the second term to the individual covariate component.

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

$$f_m(\mathbf{x}, \boldsymbol{\varepsilon}^{\text{Obs}}|N, \boldsymbol{\theta}, \boldsymbol{\eta}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{T-y_i} f_{\varepsilon}(\boldsymbol{\varepsilon}_i|\boldsymbol{\eta})$$

$$\times \prod_{i=n+1}^{N} \int_{\boldsymbol{\varepsilon}_i} p_i^{y_i} (1 - p_i)^{T-y_i} f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_i|\boldsymbol{\eta}) \, d\boldsymbol{\varepsilon}_i$$

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

$$\times \left[\int_{\boldsymbol{\varepsilon}_0} (1 - p_0)^T f_{\varepsilon}(\boldsymbol{\varepsilon}_0|\boldsymbol{\eta}) \, d\boldsymbol{\varepsilon}_0 \right]^{N-n},$$

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

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

EXAMPLE 2 (M_h-type models).   For M_h-type models the individual heterogeneity corresponds to an unobserved individual random effect component (so that $\boldsymbol{\varepsilon}^{\mathrm{Obs}} = \varnothing$ and $\boldsymbol{\varepsilon}^{\mathrm{Mis}} = \boldsymbol{\varepsilon}$). For example, for model M_h we may set $\boldsymbol{\theta} = \{\alpha\}$ and $\boldsymbol{\eta} = \{\sigma^2\}$ such that

$$\varepsilon_i \sim N(0, \sigma^2),$$

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

$$g^{-1}(p_i) = \alpha + \varepsilon_i,$$

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

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

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

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 likelihood corresponds to the conditional likelihood (conditional on the individual random effect terms) and the second term to the individual effect component.

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

$$f_m(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta})$$

$$\propto \frac{N!}{(N-n)!} \prod_{k=0}^{T} \left[\int_{\varepsilon_k \in \mathbb{S}} (p_k)^k (1 - p_k)^{T-k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_k^2}{2\sigma^2}\right) d\varepsilon_k \right]^{n_k},$$

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

EXAMPLE 3 (SECR models). For traditional spatially explicit capture–recapture models, $\mathbb{S} \subset \mathbb{R}^2$ and the individual heterogeneity corresponds to the unobserved activity center of the individual (so that $\boldsymbol{\varepsilon}^{\text{Obs}} = \varnothing$ and $\boldsymbol{\varepsilon}^{\text{Mis}} = \boldsymbol{\varepsilon}$). The range of possible models is greater for SECR than nonspatial capture–recapture, as SECR models involve multiple traps or detectors at different locations on each occasion and take account of the location(s) of observations within occasions. To this end, we define $\mathbf{u}_j = (u_{j1}, u_{j2}) \in \mathbb{R}^2$ to be the Cartesian coordinates of trap j, for $j = 1, \ldots, J$. We consider the likelihood for a study with binary detection data within occasion, such that

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

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

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

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

The complete data likelihood can be written as

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

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

The marginal data likelihood integrates out the $\boldsymbol{\varepsilon}_i$ terms and can be expressed as

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

once more omitting the term $\boldsymbol{\varepsilon}^{\text{Mis}} = \varnothing$.

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

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

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

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

2.1.2. *Complete data likelihood.* The Bayesian complete data likelihood approach specifies the unobserved individual heterogeneity terms, $\boldsymbol{\varepsilon}^{\text{Mis}}$, as auxiliary

variables (or additional parameters). The joint posterior distribution of the parameters and auxiliary variables is then formed and given by

$$\pi\left(N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}^{\text{Mis}} | \mathbf{x}, \boldsymbol{\varepsilon}^{\text{Obs}}\right) \propto f_c(\mathbf{x}, \boldsymbol{\varepsilon} | N, \boldsymbol{\theta}, \boldsymbol{\eta}) p(N, \boldsymbol{\theta}, \boldsymbol{\eta})$$
$$= f_{\mathbf{x}}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\varepsilon}) f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon} | N, \boldsymbol{\eta}) p(N, \boldsymbol{\theta}, \boldsymbol{\eta}),$$

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

An additional computational model-fitting difficulty arises since $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N\}$ and, hence, $\boldsymbol{\varepsilon}$ is itself a function of the unknown parameter, N. To address this issue, King and Brooks (2008) describe a reversible jump (RJ) MCMC algorithm for M_h-type models that is able to explore the joint posterior distribution, where the number of parameters is able to vary within the constructed Markov chain. This involved writing bespoke computer code. Alternatively, Durban and Elston (2005), Royle and Dorazio (2012), Royle, Dorazio and Link (2007), Royle et al. (2009) use data augmentation techniques that can be fitted in BUGS/JAGS. The underlying idea is to specify a superpopulation of size M, with associated individual random effect terms $\boldsymbol{\varepsilon}_i$ for $i = 1, \ldots, M$. The encounter histories for individuals $n + 1, \ldots, M$ correspond to not being observed within the study. Within the MCMC algorithm, the random effect term for each individual in this superpopulation is imputed in addition to a binary indicator variable, z_i for $i = 1, \ldots, M$, identifying which members of the superpopulation are members of the target population of interest (by definition, $z_i = 1$ for $i = 1, \ldots, n$, i.e., for all individuals observed at least once within the study). This binary indicator variable has been implemented using two different techniques, each with different consequences. Durban and Elston (2005) specify the binary variables, such that $z_1, \ldots, z_N = 1$ and $z_{N+1}, \ldots, z_M = 0$ (i.e., the indicator variables are ordered), whereas Royle, Dorazio and Link (2007), Royle et al. (2009) do not induce any such structure on the indicator variables relating to unobserved individuals, setting $z_i = 1$ for $i = 1, \ldots, n$ and modeling each indicator variable z_i for $i = n + 1, \ldots, M$. The estimate of N is obtained as the sum of nonzero indicator variables, that is, $N = \sum_{i=1}^{M} z_i$. In other words, Durban and Elston (2005) define the indicator variables, conditional on N, whereas Royle, Dorazio and Link (2007), Royle et al. (2009) define N, conditional on the indicator variables. For ease of reference we refer to the complete data likelihood data approach of Durban and Elston (2005)

as CD:DE (complete data: Durban and Elston) and of Royle and Dorazio (2012), Royle, Dorazio and Link (2007), Royle et al. (2009) as CD:R (complete data: Royle).

Several issues arise with regard to these superpopulation data augmentation approaches. For both approaches M needs to be specified and corresponds to an upper bound for the total population size. This necessarily leads to a trade-off between the size specified for M and the computational speed of the code. The larger the value of M, the greater the computational time due to the imputation of the random effect term (and binary indicator variable for CD:R) for each individual in the superpopulation. Too small a value for M will lead to a truncation of the posterior distribution and biased inference. In addition, for CD:R, since N is derived as a deterministic function of the indicator variables, it has a more limited prior specification (see Section 3.2 for further discussion regarding prior specification). Alternatively, for the approach of CD:DE, due to the more restricted nature of the indicator variable specification, mixing issues can arise. To aid in the efficiency of the computational algorithm, Durban and Elston (2005) advocate the use of a pseudo-prior for the corresponding random effect terms for individuals not in the population (i.e., for ε_i for all $i = N + 1, \ldots, M$). The pseudo-prior is obtained from an initial MCMC run, using the estimated posterior distribution for the random effect of an unobserved individual. For further discussion of data augmentation techniques (particularly focusing on CD:R), see for example, Link (2013) and Schofield and Barker (2014).

In general, without any prior information, the choice of analysis (classical marginal data likelihood or Bayesian complete data likelihood) may be data dependent. In general, for a given data set, there is a computational trade-off between these different approaches. The marginal data likelihood requires the numerical approximation of the integrals over the individual random effects; the complete data likelihood is fast to evaluate but the individual random effects need to be updated within the MCMC algorithm (using either RJMCMC or a superpopulation approach). To avoid the use of explicitly approximating multiple integrals or the need to use a superpopulation or transdimensional algorithm, we propose a hybrid semi-complete data likelihood approach. This involves numerical integration for the part of the likelihood corresponding to unobserved individuals (as in the marginal likelihood approach), while for the observed individuals any unobserved individual heterogeneity terms are treated as auxiliary variables within a data augmentation approach (as in the complete data likelihood approach). In this case, the number of auxiliary variables is known so that the dimension of the parameter space is known and fixed. Standard BUGS/JAGS software readily accommodates this approach, which involves approximation of only a single integral of dimension $\dim(\mathbb{S})$. We describe this approach in more detail next.

3. Semi-complete data likelihood. We propose a *semi-complete* data likelihood approach, combining the complete data likelihood for the individuals that are

observed within the study (i.e., individuals $i = 1, \ldots, n$) with a marginal data likelihood for the individuals that are not observed within the study (i.e., individuals $i = n + 1, \ldots, N$). The semi-complete likelihood is expressed in the form

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

where $\boldsymbol{\varepsilon}_{1:n} = \{\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n\}$, $f_{\mathbf{x}^*}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}_{1:n})$ denotes the likelihood of the capture histories conditional on the model parameters (N, $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$) and individual heterogeneity terms for the observed individuals only ($\boldsymbol{\varepsilon}_{1:n}$), and $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_{1:n} | N, \boldsymbol{\eta})$ the joint probability density function of the individual heterogeneity component for the observed individuals. Further, we let $\mathbf{x}_{a:b} = \{\mathbf{x}_a, \ldots, \mathbf{x}_b\}$ and define the following conditional likelihood functions: $f_{\mathbf{x}^*}(\mathbf{x}_{1:n} | N, \boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:n})$ for the capture histories of the observed individuals only, conditional on the model parameters and individual heterogeneity terms for the observed individuals (dropping the dependence on $\boldsymbol{\eta}$ since these are conditionally independent given $\boldsymbol{\varepsilon}_{1:n}$); $f_{\mathbf{x}^*}(\mathbf{x}_{n+1:N} | N, \boldsymbol{\theta}, \boldsymbol{\eta})$ for the capture histories of the unobserved individuals, conditional on the model parameters; and $f_{x^*}(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\eta})$ for the capture history for unobserved individual $i = n + 1, \ldots, N$, given the capture probability and individual heterogeneity model parameters (in the latter two cases dropping the conditioning on $\boldsymbol{\varepsilon}_{1:n}$). We express the conditional likelihood in the form

$$f_{\mathbf{x}^*}(\mathbf{x} | N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}_{1:n}) = f_{\mathbf{x}^*}(\mathbf{x}_{1:n} | N, \boldsymbol{\theta}, \boldsymbol{\varepsilon}_{1:n}) f_{\mathbf{x}^*}(\mathbf{x}_{n+1:N} | N, \boldsymbol{\theta}, \boldsymbol{\eta})$$

$$\propto \frac{N!}{(N-n)!} \prod_{i=1}^{n} f_x(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\varepsilon}_i) \times \prod_{i=n+1}^{N} f_{x^*}(\mathbf{x}_i | \boldsymbol{\theta})$$

$$(3.1) \qquad = \prod_{i=1}^{n} f_x(\mathbf{x}_i | \boldsymbol{\theta}, \boldsymbol{\varepsilon}_i) \times \frac{N!}{(N-n)!} (1 - p^*)^{N-n},$$

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

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

and $\boldsymbol{\omega} = \mathbf{0}$ denotes the encounter history of an individual who is unobserved within the study, $f_x(\boldsymbol{\omega} = \mathbf{0} | \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}_{\boldsymbol{\omega}})$ the conditional likelihood function associated with an individual not observed within the study and $f_{\varepsilon}(\boldsymbol{\varepsilon}_{\boldsymbol{\omega}} | \boldsymbol{\theta}, \boldsymbol{\eta})$ 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)
$$f_{\mathbf{x}^*}(\mathbf{x}|N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}_{1:n})$$
$$\propto \frac{1}{(p^*)^n} \prod_{i=1}^{n} f_x(\mathbf{x}_i|\boldsymbol{\theta}, \boldsymbol{\varepsilon}_i) \times \frac{N!}{(N-n)!}(p^*)^n(1-p^*)^{N-n},$$

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 $\boldsymbol{\varepsilon}_{1:n}^{\text{Obs}}$ and $\boldsymbol{\varepsilon}_{1:n}^{\text{Mis}}$ 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

$$\pi(N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}_{1:n}^{\text{Mis}}|\mathbf{x}, \boldsymbol{\varepsilon}_{1:n}^{\text{Obs}}) \propto f_s(\mathbf{x}, \boldsymbol{\varepsilon}_{1:n}|N, \boldsymbol{\theta}, \boldsymbol{\eta})p(N, \boldsymbol{\theta}, \boldsymbol{\eta})$$

(3.4)
$$= f_{\mathbf{x}^*}(\mathbf{x}|N, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\varepsilon}_{1:n})f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_{1:n}|N, \boldsymbol{\eta})p(N, \boldsymbol{\theta}, \boldsymbol{\eta}),$$

where $f_s(\mathbf{x}, \boldsymbol{\varepsilon}_{1:n}|N, \boldsymbol{\theta}, \boldsymbol{\eta})$ 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, $\pi(N, \boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{x}, \boldsymbol{\varepsilon}_{1:n}^{\text{Obs}})$. The number of auxiliary variables needed within this Bayesian data augmentation approach, using the semi-complete likelihood, is fixed and simply equal to $|\boldsymbol{\varepsilon}_{1:n}^{\text{Mis}}|$ (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, $\boldsymbol{\varepsilon}^{\text{Mis}}$, given in equation (2.2), since $\boldsymbol{\varepsilon}^{\text{Mis}} = \{\boldsymbol{\varepsilon}_{1:n}^{\text{Mis}}, \boldsymbol{\varepsilon}_{n+1:N}\}$ 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, Johnson and Junker (1999)] for Rasch-type (M_{th}) models and superpopulation data augmentation techniques. The first two approaches require bespoke code, while the superpopulation data augmentation

approaches can be implemented within BUGS/JAGS [Durban and Elston (2005), Royle and Dorazio (2012), Royle, Dorazio and Link (2007)] but require the specification of an upper bound M and imputation of the $(M - n)$ individual random effect terms $\boldsymbol{\varepsilon}_{n+1:M}$ (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 $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and N. However, the semi-complete data likelihood removes the necessity to impute the terms $\boldsymbol{\varepsilon}_{n+1:M}$ 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. (2016), for sample 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 (e.g., 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, Dorazio and Link (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 superpopulation is a member of the population of interest, denoted ψ. 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 easy 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 $N \sim \text{Beta-Binomial}(M, a, b)$, where M is the superpopulation 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, e.g., King and Brooks (2008), Madigan and York (1997)]. We note that specifying Jeffreys' prior and using the semi-complete data likelihood expression given in equation (3.3) lead to a standard posterior conditional distribution for N, that is,

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

for p^* given in equation (3.1).[2] 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. (2016) for sample JAGS code for the above Negative-Binomial posterior conditional distribution case]. See also Fienberg, Johnson and Junker (1999) for further discussion.

Alternative prior distributions include $p(N) \propto N^{-c}$ for some positive constant c, proposed by Fienberg, Johnson and Junker (1999). For $c > 1$ the tail of the distribution for N decays faster than for Jeffreys' prior, while $c < 1$ leads to a heavier tailed distribution. Alternatively, for an informative prior distribution for N, a Poisson or Poisson-Gamma (equivalently, a Negative-Binomial) prior distribution is often specified on N [King and Brooks (2001)]. It can also be noted that specifying $N \sim \text{Po}(\lambda)$ and $\lambda \sim \Gamma(\delta, \delta)$ for small δ provides another approximate Jeffreys' prior for N. These alternative prior distributions are able to be implemented within BUGS/JAGS [typically using the zeros or ones trick, Lunn et al. (2013)—see King et al. (2016) for associated sample JAGS code].

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

---

[2]We use the form of the Negative-Binomial distribution such that for $X \sim \text{Neg-Bin}(n, q)$ 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 in BUGS/JAGS.

4. Examples. We consider two real examples, relating to model M_h and SECR, described in Section 2. We note that as with all performance metrics for comparing the efficiency of different model-fitting approaches these are dependent on numerous factors, such as the programming language, specific application, data, model specification (including the pseudo-priors specified for the superpopulation approach), initial starting values and machine used. In order to be able to draw sensible comparisons for each example, we present results obtained from the same machine and language using the JAGS codes provided in King et al. (2016).

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

We fit the semi-complete data likelihood and complete data likelihood Bayesian superpopulation (CD:R and CD:DE) approaches in R [R Core Team (2014)] using the rjags package [Plummer (2013)]—see Appendix A of King et al. (2016) for the associated JAGS code. For each analysis we 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 the semi-complete data likelihood and CD:DE. For ease of comparison with CD:R we set $\psi \sim \mathrm{Beta}(0.001, 1)$, which induces an approximate (truncated) Jeffreys' prior for N on $1, \ldots, M$ [Link (2013)]. We note that we consider two JAGS specifications for the semi-complete data likelihood. The first approach (SCD1) uses the Jeffreys' prior specification for N explicitly in the model component of the code. However, since Jeffreys' prior is improper, we need to specify an upper bound for N, which we again denote by M (essentially this is a truncated Jeffreys' prior at M). The second approach (SCD2) specifies the (predictive) posterior conditional distribution for $N - n$, which is of Negative-Binomial form (see Section 3.2).

Following Link (2013), we specify an upper bound of $M = 1000$ for the maximum total population size for the complete data likelihood superpopulation approaches and the first semi-complete data likelihood approach (SCD1) in JAGS. For the semi-complete data likelihood approach, the integral in equation (3.2) is evaluated using Gauss–Hermite quadrature:

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

where w_j and v_j are the weights and nodes corresponding to q quadrature points [see McClintock et al. (2009)]. The degree of accuracy of this approximation in-

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 obtained using the R *package,* coda. *A total of* 30 *million iterations are used in each case with the realizations thinned by* 10

Method	Mean	Median	SD	95% CI	ESS	ESS/s
SCD1	100.3	93	32.8	(74, 171)	168,347	7.67
SCD2	101.1	93	74.9	(74, 173)	167,680	7.74
CD:R	100.6	93	32.7	(74, 171)	13,080	0.10
CD:DE	101.3	93	36.2	(74, 178)	9626	0.03

creases with q, and larger q is required for larger σ. For our analyses, we specify $q = 100$.

For each approach, we ran three chains of 10 million iterations (after initial pilot tuning and burn-in) from overdispersed starting values, thinning the realizations by 10 for memory storage purposes. Chain convergence was assessed based on visual inspection and Brooks–Gelman–Rubin diagnostics (no lack of convergence was identified). On a computer running 64-bit Windows 7 (3.4 GHz Intel Core i7 processor, 16 Gb RAM), the analyses required about 6.1 hrs for the first semi-complete data likelihood (prior distribution for N specified) approach, 6.0 hrs for the second semi-complete data likelihood (posterior conditional distribution for $N - n$ specified), 35.1 hrs for CD:R and 83.3 hours for CD:DE. We note that the run times should be interpreted comparatively, as they will in general differ across different computers as a result of different processors, operating systems, etc. The marginal posterior summaries are provided in Table 1, coupled with the effective sample sizes (per second) using the R package coda [Plummer et al. (2006)] for each approach.

Although setting $M = 1000$ may appear conservative, this did appear to influence the skewness of the right tail of the marginal posterior distribution for N relative to the (unbounded) posterior distribution for N when using the second semi-complete data likelihood approach (SCD2). We therefore reran the first semi-complete data likelihood (SCD1) analysis with $M = 10,000$, leading to posterior summary results more similar to the second complete data likelihood approach [N posterior mean $= 100.9$, median $= 93$, SD $= 56.1$, 95% credible interval (CI) $= (74, 172)$], but with noticeably reduced effective sample size (ESS $= 74,928$) and increased computation time (ESS/s $= 2.81$). Nevertheless, specifying larger M for the first semi-complete data likelihood approach comes at considerably less computational cost compared to the superpopulation complete data likelihood approaches (CD:R and CD:DE). Avoidance of the need to specify M when using BUGS/JAGS remains an advantage of the general semi-complete

data likelihood approach (this is true even when using Jeffreys' prior on N by specifying the posterior conditional distribution for $N - n$ in the model component of the code).

For approach SCD1, using an explicit Negative-Binomial or Beta-Binomial approximation to Jeffreys' prior [code is provided in Appendix A of King et al. (2016)] unsurprisingly led to similar results in terms of ESS and ESS/s as for the use of the explicit (truncated) Jeffreys' prior. However, within the model specification code, using the distributions' hierarchical form where an auxiliary variable is introduced for the Poisson mean or Binomial probability and imputed within the MCMC algorithm led to lower ESS and ESS/s as a result of poorer mixing due to posterior correlation between parameters. We do not consider these prior specifications further.

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

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

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

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

As in Section 4.1, we fit both forms of the semi-complete data likelihood [equations (3.1) and (3.3)] and the superpopulation complete data likelihood Bayesian

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 using the R package coda. *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

approaches CD:R and CD:DE using the rjags package [see Appendix B of King et al. (2016) for sample JAGS code]. For the complete data likelihood approaches and first semi-complete data likelihood (specifying Jeffreys' prior on N within the model component of the JAGS code), we specify an upper bound of $M = 1000$ for the discrete support of N. For both semi-complete likelihoods the integral in equation (3.2) was approximated by a summation over a rectangular grid of 4200 points. Note that a suitable choice of grid will be case dependent, with increases in accuracy resulting from greater spatial extents and decreased distances between neighboring grid points, but at the expense of computational time. An exploratory analysis suggested that the grid used was relatively conservative, achieving good numerical accuracy.

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

5. Discussion. For closed population models, the semi-complete data likelihood specifies the joint probability density function of the model parameters and associated unobserved individual heterogeneity terms for only those individuals

observed, conditional on the observed capture histories and observed individual heterogeneity components. This likelihood is specified as an integral of the individual heterogeneity component for the unobserved individuals. The integral is analytically intractable but of dimension equal to the dimension of the individual heterogeneity component of the model, and hence typically small. This permits the use of standard (efficient) numerical approximation techniques to estimate the integral (for example, in OpenBUGS, the inbuilt `integral` function can be used to conduct one-dimensional integration; with similar inbuilt functions in R for one or multi-dimensional integrals). The semi-complete data likelihood approach can be applied to a range of different individual heterogeneity models.

Using this semi-complete data likelihood within a Bayesian analysis of closed capture–recapture data in the presence of individual heterogeneity removes the need for trans-dimensional algorithms to explore the posterior distribution of the parameters due to the "unknown number of parameters" problem. Consequently, the models can be fitted efficiently in standard software, such as BUGS/JAGS without using a superpopulation approach. The semi-complete data likelihood approach is significantly more efficient than the previous superpopulation approaches, as demonstrated in Section 4, where the improvement for the examples that we considered using the codes provided in King et al. (2016) was up to two orders of magnitude. The improvement is in terms of both computational time and effective sample sizes (as a result of improved mixing within the MCMC algorithm). The efficiency of the superpopulation approaches is heavily dependent on the upper limit specified for the superpopulation, M. This makes the Bayesian approach feasible for fitting to a significantly wider range of data, particularly for spatially explicit capture–recapture, where the use of a Bayesian data augmentation technique can be particularly inefficient. In general, the ESS and ESS/s for the different approaches is dependent on numerous factors, including the exact form of the model specification, the pseudo-priors specified in the superpopulation approach, initial starting values and the computer on which the simulations are being run.

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

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

SUPPLEMENTARY MATERIAL

Supplement to "Capture–recapture abundance estimation using a semi-complete data likelihood approach" (DOI: 10.1214/15-AOAS890SUPP; .pdf). The supplement consists of Appendices A and B that provide sample JAGS codes for the examples provided in the text using the different model-fitting algorithms (referenced in Sections 3.2, 4.1 and 4.2).

REFERENCES

BONNER, S. and SCHOFIELD, M. R. (2014). MC(MC)MC: Exploring Monte Carlo integration within MCMC for mark-recapture models with individual covariates. *Methods in Ecology and Evolution*. **5**. 1305–1315.

BORCHERS, D. L., BUCKLAND, S. T. and ZUCCHINI, W. (2002). *Estimating Animal Abundance, Closed Populations*. Springer, London.

BORCHERS, D. L. and EFFORD, M. G. (2008). Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics* **64** 377–385, 664. MR2432407

BUCKLAND, S. T., ANDERSON, D. R., BURNHAM, K. P., LAAKE, J. L., BORCHERS, D. L. and THOMAS, L. (2001). *Introduction to Distance Sampling*. Oxford Univ. Press, Oxford.

COULL, B. A. and AGRESTI, A. (1999). The use of mixed logit models to reflect heterogeneity in capture–recapture studies. *Biometrics* **55** 294–301.

DORAZIO, R. M. and ROYLE, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59** 351–364. MR1987402

DURBAN, J. W. and ELSTON, D. A. (2005). Mark-recapture with occasion and individual effects: Abundance estimation through Bayesian model selection in a fixed dimensional parameter space. *J. Agric. Biol. Environ. Stat.* **10** 291–305.

FIENBERG, S. E., JOHNSON, M. S. and JUNKER, B. W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *J. Roy. Statist. Soc. Ser. A* **162** 383–405.

GIMENEZ, O. and CHOQUET, R. (2010). Individual heterogeneity in studies on marked animals using numerical integration: Capture–recapture mixed models. *Ecology* **91** 951–957.

KING, R. and BROOKS, S. P. (2001). On the Bayesian analysis of population size. *Biometrika* **88** 317–336. MR1844834

KING, R. and BROOKS, S. P. (2008). On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics* **64** 816–824. MR2526632

KING, R., MORGAN, B. J. T., GIMENEZ, O. and BROOKS, S. P. (2009). *Bayesian Analysis for Population Ecology Coba*. CRC Press, Boca Raton, FL.

KING, R., BIRD, S. M., OVERSTALL, A. M., HAY, G. and HUTCHINSON, S. J. (2014). Estimating prevalence of injecting drug users and associated heroin-related death rates in England by using regional data and incorporating prior information. *J. Roy. Statist. Soc. Ser. A* **177** 209–236. MR3158675

KING, R., MCCLINTOCK, B. T., KIDNEY, D. and BORCHERS, D. (2016). Supplement to "Capture–recapture abundance estimation using a semi-complete data likelihood approach." DOI:10.1214/15-AOAS890SUPP.

LANGROCK, R. and KING, R. (2013). Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Ann. Appl. Stat.* **7** 1709–1732. MR3127965

LINK, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial N. *Ecology* **94** 2173–2179.

LUNN, D., JACKSON, C., BEST, N., THOMAS, A. and SPIEGELHALTER, D. (2013). *The BUGS Book*: *A Practical Introduction to Bayesian Analysis*. CRC Press, Boca Raton, FL.

MADIGAN, D. and YORK, J. C. (1997). Bayesian methods for estimation of the size of a closed population. *Biometrika* **84** 19–31. MR1450189

MANRIQUE-VALLIER, D. and FIENBERG, S. E. (2008). Population size estimation using individual level mixture models. *Biom. J.* **50** 1051–1063. MR2649394

MCCLINTOCK, B. T., WHITE, G. C., BURNHAM, K. P. and PRYDE, M. A. (2009). A generalized mixed effects model of abundance for mark-resight data when sampling is without replacement. In *Modeling Demographic Processes in Marked Populations* (D. L. Thomson, E. G. Cooch and M. J. Conroy, eds.) 271–289. Springer, New York.

MCCREA, R. S. and MORGAN, B. J. T. (2015). *Analysis of Capture–Recapture Data*. Chapman & Hall/CRC, Boca Raton, FL. MR3330977

MORGAN, B. J. T. and RIDOUT, M. S. (2008). A new mixture model for capture heterogeneity. *J. Roy. Statist. Soc. Ser. C* **57** 433–446. MR2526127

OTIS, D. L., BURNHAM, K. P., WHITE, G. C. and ANDERSON, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildl. Monogr.* **62** 1–135.

OVERSTALL, A. M., KING, R., BIRD, S. M., HUTCHINSON, S. J. and HAY, G. (2014). Incomplete contingency tables with censored cells with application to estimating the number of people who inject drugs in Scotland. *Stat. Med.* **33** 1564–1579. MR3240769

PLEDGER, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* **56** 434–442.

PLEDGER, S. (2005). The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics* **61** 868–876. MR2196177

PLEDGER, S., EFFORD, M., POLLOCK, K. H., COLLAZO, J. A. and LYONS, J. E. (2009). Stopover duration analysis with departure probability dependent on unknown time since arrival. In *Modeling Demographic Processes in Marked Populations* (D. L. Thomson, E. G. Cooch and M. J. Conroy, eds.) 349–363. Springer, New York.

PLUMMER, M. (2013). rjags: Bayesian graphical models using MCMC. R package version 3-11.

PLUMMER, M., BEST, N., COWLES, K. and VINES,K. (2015). CODA: Convergence diagnosis and output analysis for MCMC. *R News* **6** 7–11.

R CORE TEAM (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.

ROYLE, J. A. and DORAZIO, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. *J. Ornithol*. **152** S521–S537.

ROYLE, J. A., DORAZIO, R. M. and LINK, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *J. Comput. Graph. Statist*. **16** 67–85. MR2345748

ROYLE, J. A., KARANTH, K. U., GOPALASWAMY, A. M. and KUMAR, N. S. (2009). Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. *Ecology* **90** 3233–3244.

SCHOFIELD, M. R. and BARKER, R. J. (2014). Hierarchical modeling of abundance in closed population capture–recapture models under heterogeneity. *Environ. Ecol. Stat*. **21** 435–451. MR3248533

WILLIAMS, B. K., NICHOLS, J. D. and CONROY, M. J. (2002). *Analysis and Management of Animals Populations*. Academic Press, San Diego, CA.

R. KING
SCHOOL OF MATHEMATICS
UNIVERSITY OF EDINBURGH
JAMES CLERK MAXWELL BUILDING
THE KING'S BUILDINGS
PETER GUTHRIE TAIT ROAD
EDINBURGH EH9 3FD
UNITED KINGDON
E-MAIL: Ruth.King@ed.ac.uk

B. T. MCCLINTOCK
NATIONAL MARINE MAMMAL LABORATORY
ALASKA FISHERIES SCIENCE CENTER
NOAA NATIONAL MARINE FISHERIES SERVICE
7600 SAND POINT WAY NE
SEATTLE, WASHINGTON 98115
USA
E-MAIL: Brett.McClintock@noaa.gov

D. KIDNEY
D. 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 KY16 9LZ
UNITED KINGDON
E-MAIL: darrenkidney@googlemail.com
         dlb@st-andrews.ac.uk