



Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: [www.elsevier.com/locate/stamet](http://www.elsevier.com/locate/stamet)



# A generalised likelihood framework for partially observed capture–recapture–recovery models



R. King<sup>a,\*</sup>, R.S. McCrea<sup>b</sup>

<sup>a</sup> Centre for Research into Ecological and Environmental Modelling and School of Mathematics and Statistics, University of St Andrews, St Andrews, Fife, KY16 9LZ, UK

<sup>b</sup> National Centre for Statistical Ecology and School of Mathematics, Statistics and Actuarial Science, Cornwallis Building, University of Kent, Canterbury, Kent, CT2 7NF, UK

## ARTICLE INFO

### Article history:

Received 13 December 2012

Received in revised form

16 May 2013

Accepted 9 July 2013

### Keywords:

Capture–recapture–recovery data

Closed form likelihood

Multi-state

Partially observed states

Sufficient statistics

## ABSTRACT

We provide a closed form likelihood expression for multi-state capture–recapture–recovery data when the state of an individual may be only partially observed. The corresponding sufficient statistics are presented in addition to a matrix formulation which facilitates an efficient calculation of the likelihood. This likelihood framework provides a consistent and unified framework with many standard models applied to capture–recapture–recovery data as special cases.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Models for data collected on populations of wild animals are becoming increasingly complex in order to more realistically model the underlying biological traits exhibited by the animals. We focus on capture–recapture–recovery data. This essentially involves uniquely identifying individuals (either by natural markings or via applying a form of mark) and subsequently observing live individuals at future capture events and/or recovering dead individuals. This approach typically assumes that individuals are uniquely identifiable via the markings. The data can then be recorded in the form of an individual encounter history for each marked animal. The standard capture–recapture presence/absence data of live resightings can be modelled by the Cormack–Jolly–Seber (CJS) model [8,10,27], conditioning on the initial capture of each individual. Under the standard model assumptions, including, for

\* Corresponding author. Tel.: +44 1334461820.

E-mail addresses: [ruth@mcs.st-and.ac.uk](mailto:ruth@mcs.st-and.ac.uk) (R. King), [r.s.mccrea@kent.ac.uk](mailto:r.s.mccrea@kent.ac.uk) (R.S. McCrea).

example, no misidentification of individuals and individuals behaving independently of each other, a likelihood expression can be constructed, permitting the estimation of time-dependent survival and recapture probabilities (see for example [12,19,26] for further discussion). An explicit expression for the likelihood of integrated capture–recapture–recovery data, using sufficient statistics was derived by [4].

In addition to simply recording the presence/absence of an individual at each capture event (either alive or dead) additional information such as their location or state in which an animal is in may also be recorded upon capture/recapture. This has led to the important development of the formulation of the Arnason–Schwarz (AS) model [1–3,25], which is the multi-state generalisation of the CJS model, once more conditioning on the initial capture and the corresponding observed state of each individual (see [17] for a review of the AS model). The AS model was initially developed for multi-site capture–recapture data, where individuals are observed alive at different (discrete) sites, but can be generally applied to individuals being observed in different (discrete) states [16]. This multi-state approach allows the dependence of the survival and/or recapture probabilities to be state-dependent and the estimation of transition probabilities between the different states (typically assuming that the model parameters are first-order Markovian). [14] presented an efficient construction of the closed-form likelihood for the AS model, via sufficient statistics, for multi-state capture–recapture–recovery data, extending the likelihood expression of [4]. The AS multi-state model however does not account for uncertainty in state assignment (i.e. all states are known without error). The more general multievent model extends the AS model and incorporates the possibility of state uncertainty [22]. These models belong to the more general class of hidden Markov models which are becoming increasingly popular within the analysis of capture–recapture–recovery data (see for example [12]).

In this paper we provide a concise closed form likelihood, via the use of sufficient statistics, for multi-state capture–recapture–recovery data where at any given time the state of an individual may be completely known (without error), observed with error, partly observed (so that it is known to be in a subset of possible states), or completely unobserved. We partition the observed encounter histories into component sections to optimise efficiency and permit the specification of non-overlapping sufficient statistics necessary, for example, for goodness-of-fit tests. As for the AS model, we assume a first-order Markovian structure for transitions between states, conditioning on the initial capture and known or partially observed state of each individual (although note that in multievent models, the observed state is not conditioned on and additional parameters are included corresponding to the probabilities of being in each possible state at initial capture). Finally, we assume that when an individual is observed, it is known whether the corresponding state of the individual is completely known (i.e. observed without error) or only partially observed (i.e. observed with error, in a set of possible states or simply the state is unobserved).

To demonstrate the different types of data that can result in partially observed states we briefly describe three motivating examples.

#### *Example 1—multiple mark types.*

In many mark–recapture–recovery studies multiple mark types are used, each uniquely identifying an individual. Different mark types may include, for example, brands, tags, rings and photo-identification techniques. This can lead to multiple marks applied to a single individual [23,28]. The state of an individual corresponds to the set of identifying marks. Typically the recapture probability will be dependent on the state of the individual. However when an individual is observed, the set of marks present may not be fully recorded (for example, not all markings may be visible). This can result in partially observed states. Further, within double-tagging studies the transition probabilities between states are often of particular interest as this estimates the probability of tag-loss.

#### *Example 2—breeding.*

Upon capture, individuals are recorded as breeding or non-breeding. The state of the individual is simply its breeding status. Both the recapture and survival probabilities may be state-dependent. However, it may not always be possible to assign a breeding status to the individual when it is observed, so the state of “unassigned” may be reported. In addition, a probable state may be recorded which could be regarded as a state observed with potential error and again can be regarded as a partial

state. For example, an individual bird seen close to a nest may be assigned as “probably breeding”, allowing for potential error within the observed state.

*Example 3—the number of offspring.*

An individual may have up to  $M$  offspring. When an individual is observed, the number of offspring is recorded as the state of the individual. The survival probability of the individual may be dependent on the number of offspring produced in a given year (higher breeding effort may result in decreased survival probability). However, the number of offspring produced may not be completely observed, but instead the partial state may be expressed in the form of “at least  $m$  offspring”. Note that “at least 0 offspring” could be interpreted as an unassigned (i.e. completely unknown) state.

The fitting of general partial observation models is possible using multievent software E-SURGE [5]. Within this package, the corresponding likelihood is calculated by evaluating the probabilities associated with each individual observed encounter history. Partial observations can be incorporated by including additional “states” so that the set of possible states is the union of the true and partially observed states. However, this model fitting approach can lead to computational problems, due to the expansion in the number of “states” defined. An alternative model fitting approach is to use a Bayesian data augmentation approach [29]. For the AS model, the unobserved (or missing) states corresponding to when an individual is unobserved are imputed using an auxiliary variable approach [9,11,15]. The joint posterior distribution of the model parameters and additional auxiliary variables is formed and the auxiliary variables integrated out within a Markov chain Monte Carlo algorithm to obtain an estimate of the posterior distribution of only the model parameters. However, this data augmentation can be computationally expensive (due to the number of imputed states in the Markov chain algorithm) and makes model discrimination difficult, although [13] present a reversible jump algorithm for AS models for obtaining posterior model probabilities. This approach can be extended to the case of partially observed states, imputing the true states of partial observations of individuals [28].

The explicit modular likelihood we propose within this paper allows complex models to be fitted within a classical framework in an efficient manner. In turn this permits the investigation of model selection using standard model discrimination techniques (such as information criteria) and, due to the non-overlapping structure of the sufficient statistics, the application of absolute goodness-of-fit tests. In Section 2 we present an expression for the generalised likelihood before extending the likelihood expression to that of matrix form. In Section 3 we present a number of standard models as special cases of the generalised likelihood before presenting a real example in Section 4. We conclude with a discussion in Section 5.

## 2. Likelihood and sufficient statistics

### 2.1. Capture histories

Suppose that captures occur at times  $t \in \{1, \dots, T\}$ . We let  $\mathcal{R} = \{r_1, \dots, r_{|\mathcal{R}|}\}$  denote the set of possible states and  $\mathcal{P} = \{m_1, \dots, m_{|\mathcal{P}|}\}$  the set of partial states an individual can be additionally observed within. We note that when there are several discrete covariates a single state  $r$  corresponds to a specific combination of each covariate value, and the set of all states,  $\mathcal{R}$ , corresponding to the set of all possible combinations of covariate values. The term partial state includes states that may be observed with error, partially observed (so the state is known to be within a set of possible states) or completely unobserved. For the case of multiple discrete covariates, a partial state includes the case where some of the covariate values are known, but at least one covariate is unknown (i.e. this is an example where the true state is known to be in a subset of possible states,  $\mathcal{S} \subseteq \mathcal{R}$ , corresponding to the set of those states with the given observed covariate values). We assume it is known whether an individual is fully observed or only partially observed so that  $\mathcal{P} \cap \mathcal{R} = \emptyset$  and that there is no error with individuals assigned to known states. In other words, if an individual is assigned state  $r \in \mathcal{R}$  the true state of the individual is  $r$  (i.e. there is no error). Conversely, if an individual is assigned to a partial state  $m \in \mathcal{P}$ , there is uncertainty with regard to their true state. A typical encounter history

for an individual may be expressed as  $\mathbf{x} = \{x_{(1)}, \dots, x_{(T)}\}$  where  $x_{(t)}$  denotes the encounter of the individual at time  $t = 1, \dots, T$ , such that,

$$x_{(t)} = \begin{cases} 0 & \text{if individual is unobserved at time } t; \\ r & \text{if individual is observed at time } t \text{ and assigned state } r \in \mathcal{R} \cup \mathcal{P}; \\ \dagger & \text{if individual is recovered dead in the interval } (t, t + 1]. \end{cases}$$

Any individual encounter history,  $\mathbf{x}$ , can be subdivided into two distinct capture histories dependent on whether the state is known at a given time or only partially observed. In particular, we denote the encounter history of *known* states by  $\mathbf{y} = \{y_{(1)}, \dots, y_{(T)}\}$ , and the encounter history of *partial* states by  $\mathbf{z} = \{z_{(1)}, \dots, z_{(T)}\}$ , where

$$y_{(t)} = \begin{cases} x_{(t)} & \text{for } x_{(t)} \in \mathcal{R} \cup \dagger; \\ 0 & \text{otherwise;} \end{cases} \quad \text{and} \quad z_{(t)} = \begin{cases} x_{(t)} & \text{for } x_{(t)} \in \mathcal{P}; \\ 0 & \text{otherwise.} \end{cases}$$

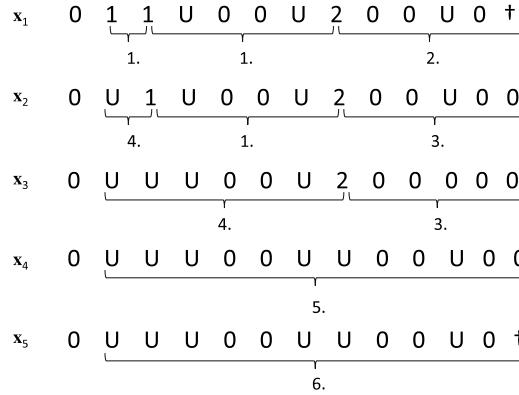
For notational convenience, we set  $\mathbf{z}_{s:t} = \{z_{(s)}, \dots, z_{(t)}\}$ , corresponding to the partially observed states (including unobserved) between times  $s$  and  $t$  inclusive. We let  $\mathbf{z}_{s:t}^* = \{z_{(i)} : z_{(i)} \in \mathcal{P} \cup \{0\}, i = s, \dots, t\}$  denote the set of all partial encounter histories and  $|\mathbf{z}_{s:t}^*| = (|\mathcal{P}| + 1)^{t-s+1}$  the total number of partial histories from times  $s$  to  $t$  (for  $t \geq s$ ). For notational convenience, for  $t < s$ , we set  $\mathbf{z}_{s:t} \equiv \emptyset$ . The set of possible partial encounter histories is ordered such that in the set  $\mathbf{z}_{s:t}^*$  each  $z_{(i)}$  term increases in cyclic order from 0 to  $m_{|\mathcal{P}|}$ , such that  $z_{(t)}$  changes between each element of the set; the  $z_{(t-1)}$  term changes every  $|\mathcal{P}| + 1$  element and so on. For example, suppose that there is only one partial state, denoted by  $m$ , then  $\mathbf{z}_{1:2}^* = \{\{0, 0\}, \{0, m\}, \{m, 0\}, \{m, m\}\}$ . This is the ordering that is used within the definition of the sufficient statistics (which is important when the corresponding matrix formulation of the likelihood is given). We note that in many applications (such as the multiple mark type example in Section 1 and the tag-loss model described in Section 3.4) transitions between some states (and hence the possible partial states they may be observed in) are not possible, reducing the set of *possible* partial encounter histories. However, we retain the most general notation here for simplicity (but note that the corresponding sufficient statistics for impossible part histories will simply be equal to 0, as well as the corresponding probability of any such history).

## 2.2. Model parameters

The model parameters are defined to be:

- $\phi_t$ : row vector of length  $|\mathcal{R}|$  with  $i$ th element,  $\phi_t(r_i)$ , the probability an individual survives until time  $t + 1$  given the individual is alive and in state  $r_i$  at time  $t$ , for  $t = 1, \dots, T - 1$ ;
- $\mathbf{p}_t$ : row vector of length  $|\mathcal{R}|$  with  $i$ th element,  $p_t(r_i)$ , the probability an individual is recaptured at time  $t$  given the individual is alive and in state  $r_i$  at time  $t$ , for  $t = 2, \dots, T$ ;
- $\lambda_t$ : row vector of length  $|\mathcal{R}|$  with  $i$ th element,  $\lambda_t(r_i)$ , the probability an individual is recovered in  $(t, t + 1]$  given the individual is in state  $r_i$  at time  $t$  and dies in  $(t, t + 1]$  for  $t = 1, \dots, T - 1$ ;
- $\psi_t$ : matrix of dimension  $|\mathcal{R}| \times |\mathcal{R}|$  with  $(i, j)$ th element,  $\psi_t(r_i, r_j)$ , the probability an individual is in state  $r_j$  at time  $t + 1$  given the individual is in state  $r_i$  at time  $t$  and is alive at time  $t + 1$ , for  $t = 1, \dots, T - 1$ ;
- $\alpha_t$ : row vector of length  $|\mathcal{R}|$  with  $i$ th element,  $\alpha_t(r_i)$ , the probability an individual is assigned to known state  $r_i$  at time  $t$  given the individual is observed at time  $t$ , for  $t = 2, \dots, T$ ;
- $\beta_t$ : matrix of dimension  $|\mathcal{P}| \times |\mathcal{R}|$  with  $(i, j)$ th element,  $\beta_t(m_i, r_j)$ , the probability an individual is assigned partial state  $m_i$  at time  $t$  given the individual is observed and in state  $r_j$  at this time, for  $t = 2, \dots, T$ ;
- $\pi_t$ : matrix of dimension  $|\mathcal{R}| \times |\mathcal{P}|$  with  $(i, j)$ th element,  $\pi_t(r_i, m_j)$ , the probability an individual is in state  $r_i$  at time  $t$ , given the individual is assigned partial state  $m_j$  at time  $t$ , for  $t = 1, \dots, T - 1$ .

Let  $\phi = \{\phi_1, \dots, \phi_{T-1}\}$  with similar notation for the other parameters. The survival, recapture, recovery and transition parameters ( $\phi$ ,  $\mathbf{p}$ ,  $\lambda$  and  $\psi$ ) are those required for the AS model (allowing for dead recoveries). However there are additional terms: known state assignment probabilities,  $\alpha$ ; partial state assignment probabilities,  $\beta$ ; and conditional state probabilities,  $\pi$ . We note that by definition,  $\sum_{m \in \mathcal{P}} \beta_t(m, r) = 1 - \alpha_t(r)$ , for all  $t = 2, \dots, T$  and  $r \in \mathcal{R}$ , and  $\sum_{r \in \mathcal{R}} \pi_t(r, m) = 1$ , for all  $t = 1, \dots, T - 1$  and  $m \in \mathcal{P}$ .



**Fig. 1.** A diagrammatic representation of the decomposition of five encounter histories  $x_1, \dots, x_5$  into the six composite parts (denoted by 1., 2., ..., 6.) provided in Section 2.3. Known observed states are denoted by 1 and 2 (i.e.  $\mathcal{R} = \{1, 2\}$ ) with an unknown state denoted by U (i.e.  $\mathcal{P} = \{U\}$ ).

### 2.3. Decomposing the likelihood

We decompose the likelihood into distinct parts considering two different scenarios for an individual encounter history: (i) an individual is observed in a known live state at least once during the study (i.e.  $y_{(t)} \in \mathcal{R}$  for some  $t \in \{1, \dots, T\}$ ); (ii) an individual is never observed in a known live state (i.e.  $y_{(t)} = 0$  for all  $t \in \{1, \dots, T\}$ ). For case (i) an individual encounter history can be decomposed into four distinct parts:

1. consecutive observations with known live states:  $y_{(t)}, 0, \dots, 0, y_{(t+j)}$ , for  $y_{(t)}, y_{(t+j)} \in \mathcal{R}$  and  $y_{(t+1)} = \dots = y_{(t+j-1)} = 0$  (when  $j > 1$ );
2. final sighting in a known live state to being recovered dead:  $y_{(t)}, 0, \dots, 0, \dagger$ , for  $y_{(t)} \in \mathcal{R}, y_{(t+j)} = \dagger$  and  $y_{(t+1)} = \dots = y_{(t+j-1)} = 0$  (when  $j > 1$ );
3. final sighting in a known live state to the end of the study, given not recovered dead:  $y_{(t)}, 0, \dots, 0$  for  $y_{(t)} \in \mathcal{R}$  and  $y_{(t+1)} = \dots = y_{(T)} = 0$ ; and
4. first sighting in a partial state to first sighting in a known live state:  $z_{(t)}, \dots, y_{(t+j)}$ , for  $z_{(t)} \in \mathcal{P}, y_{(t+j)} \in \mathcal{R}$  and  $y_{(1)} = \dots = y_{(t+j-1)} = 0$  (for  $j \geq 1$ ).

For case (ii) the entire encounter history needs to be modelled but individual encounter histories can be subdivided into two distinct categories:

5. individuals only observed in partial states on each capture and not recovered dead:  $z_{(1)}, \dots, z_{(t)}, \dots, z_{(T)}$ , for  $z_{(t)} \in \mathcal{P} \cup \{0\}$  for  $t = 1, \dots, T$  such that  $\exists j$  where  $z_{(j)} \in \mathcal{P}$  (and  $y_{(1)} = \dots = y_{(T)} = 0$ );
6. individuals only observed in partial states on each capture and recovered dead:  $z_{(1)}, \dots, y_{(t)} = \dagger$ , for  $z_{(1)}, \dots, z_{(t-1)} \in \mathcal{P} \cup \{0\}$  such that  $\exists j$  where  $z_{(j)} \in \mathcal{P}$  and  $y_{(1)} = \dots = y_{(t-1)} = 0$ .

All capture histories can be decomposed into one or more of these non-overlapping parts. To illustrate this graphically Fig. 1 provides a diagrammatical representation of the decomposition of a series of encounter histories into the six composite parts identified above, conditional on the first time an individual is observed. For these histories (as for the example considered in Section 4) there are two observed states denoted by 1 and 2 (so that  $\mathcal{R} = \{1, 2\}$ ) and a single unknown state, denoted by U (so that  $\mathcal{P} = \{U\}$ ) corresponding to the state being unknown when an individual is observed.

### 2.4. Sufficient statistics

We consider the number of individuals with each part history provided in the previous section, allowing for partially observed states. The corresponding sufficient statistics are:

1.  $\mathbf{n}_{s,t}$  for  $s = 1, \dots, T-1$  and  $t = s, \dots, T-1$ : matrix of dimension  $|\mathcal{R}| \times |\mathcal{R}|(|\mathcal{P}|+1)^{t-s}$  composed of  $|\mathcal{R}|^2$  sub-matrices with the  $(i, j)$ th sub-matrix consisting of the row vector  $\mathbf{n}_{s,t}(r_i, r_j)$  of length  $(|\mathcal{P}|+1)^{t-s}$ , with  $(\mathbf{z}_{s+1:t})$ th element, denoted by  $n_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t})$ , corresponding to the number of individuals observed at time  $s$  in state  $r_i$ , next observed in a known state at time  $t+1$  in state  $r_j$  with partial history  $\mathbf{z}_{s+1:t}$ ;
2.  $\mathbf{d}_{s,t}$  for  $s = 1, \dots, T-1$  and  $t = s, \dots, T-1$ : matrix of dimension  $|\mathcal{R}| \times (|\mathcal{P}|+1)^{t-s}$  with  $(i, \mathbf{z}_{s+1:t})$ th element, denoted by  $d_{s,t}(r_i, \mathbf{z}_{s+1:t})$ , corresponding to the number of individuals observed at time  $s$  in state  $r_i$  next observed in a known state at time  $t+1$  when recovered dead with partial history  $\mathbf{z}_{s+1:t}$ ;
3.  $\mathbf{v}_{s,t}$  for  $s = 1, \dots, T-1$  and  $t = s, \dots, T$ : matrix of dimension  $|\mathcal{R}| \times \{|\mathcal{P}|(|\mathcal{P}|+1)^{t-s-1}\}^{\mathbb{I}_{\{t>s\}}}$  (where  $\mathbb{I}_{\{\cdot\}}$  denotes the indicator function) with  $(i, \mathbf{z}_{s+1:t})$ th element, denoted by  $v_{s,t}(r_i, \mathbf{z}_{s+1:t})$ , corresponding to the number of individuals observed at time  $s$  in state  $r_i$  that are not observed again in a known state (including death) but observed for the final time at time  $t$  with partial history  $\mathbf{z}_{s+1:t}$ ;
4.  $\mathbf{w}_{s,t}$  for  $s = 1, \dots, T-1$  and  $t = s, \dots, T-1$ : matrix of dimension  $|\mathcal{P}| \times |\mathcal{R}|(|\mathcal{P}|+1)^{t-s}$  composed of  $|\mathcal{P}| \times |\mathcal{R}|$  sub-matrices with the  $(i, j)$ th sub-matrix consisting of a row vector  $\mathbf{w}_{s,t}(m_i, r_j)$  of length  $(|\mathcal{P}|+1)^{t-s}$  with  $(\mathbf{z}_{s+1:t})$ th element, denoted by  $w_{s,t}(m_i, r_j, \mathbf{z}_{s+1:t})$ , corresponding to the number of individuals observed for the first time at time  $s$  in partial state  $m_i$  and observed in a known state for the first time at  $t+1$  in state  $r_j$  with partial history  $\mathbf{z}_{s+1:t}$ ;
5.  $\mathbf{b}_{s,t}$  for  $s = 1, \dots, T-1$  and  $t = s, \dots, T$ : matrix of dimension  $|\mathcal{P}| \times \{|\mathcal{P}|(|\mathcal{P}|+1)^{t-s-1}\}^{\mathbb{I}_{\{t>s\}}}$  with  $(i, \mathbf{z}_{s+1:t})$ th element, denoted by  $b_{s,t}(m_i, \mathbf{z}_{s+1:t})$ , corresponding to the number of individuals not observed in a known state (including death) at any time but first observed at time  $s$  in partial state  $m_i$  and last observed at time  $t$  with partial history  $\mathbf{z}_{s+1:t}$ ;
6.  $\mathbf{h}_{s,t}$  for  $s = 1, \dots, T-1$  and  $t = s, \dots, T-1$ : matrix of dimension  $|\mathcal{P}| \times (|\mathcal{P}|+1)^{t-s}$  with  $(i, \mathbf{z}_{s+1:t})$ th element, denoted by  $h_{s,t}(m_i, \mathbf{z}_{s+1:t})$ , corresponding to the number of individuals not observed in a live known state at any time but first observed at time  $s$  in partial state  $m_i$  and recovered dead in the interval  $(t, t+1]$  with partial history  $\mathbf{z}_{s+1:t}$ .

Notationally, let  $\mathbf{n}$  be the matrix with  $(s, t)$ th sub-matrix  $\mathbf{n}_{s,t}$  for  $s = 1, \dots, T-1$  and  $t = s, \dots, T-1$ ; and similarly for the other terms. The sufficient statistics are  $\mathbf{n}, \mathbf{d}, \mathbf{v}, \mathbf{w}, \mathbf{b}, \mathbf{h}$ . We initially construct the likelihood using recursive formulae, before expressing the likelihood in matrix form.

## 2.5. Likelihood

We consider the probabilities associated with each distinct decomposed part capture history identified previously. These are obtained by considering the one-step ahead forwards (or backwards) probabilistic arguments, following a similar argument to that proposed by [14]. The technical details deriving the probabilistic arguments are provided in the [Appendix](#). We begin by defining the following two probability statements and providing the corresponding recursive formulae for each that will be repeatedly used within the likelihood function.

Let  $Z_t(r_i, z_{(t)})$  denote the probability that an individual not observed in a known state at time  $t$  has partial observation encounter  $z_{(t)} \in \mathcal{P} \cup \{0\}$ , given that they are in state  $r_i \in \mathcal{R}$  at time  $t$ . Immediately, we have that,

$$Z_t(r_i, z_{(t)}) = \{1 - p_t(r_i)\}^{\mathbb{I}_{\{z_{(t)}=0\}}} \{p_t(r_i)\beta_t(z_{(t)}, r_i)\}^{\mathbb{I}_{\{z_{(t)} \neq 0\}}},$$

where  $\mathbb{I}_{\{\cdot\}}$  denotes the indicator function.

Let  $Q_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t})$  denote the probability that an individual in state  $r_i \in \mathcal{R}$  at time  $s$  is only partially observed or unobserved between times  $s+1$  and  $t$  with part history  $\mathbf{z}_{s+1:t}$  and is in state  $r_j \in \mathcal{R}$  at time  $t+1$ . Then,

$$Q_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t}) = \begin{cases} \phi_t(r_i)\psi_t(r_i, r_j) & s = t; \\ \phi_s(r_i) \sum_{u \in \mathcal{R}} \psi_s(r_i, u) Z_{s+1,t}(u, z_{(s+1)}) Q_{s+1,t}(u, r_j, \mathbf{z}_{s+2:t}) & s < t. \end{cases} \quad (1)$$

We now consider the probability associated with each decomposed part capture history.

1. Define  $O_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t})$  to be the probability an individual in state  $r_i \in \mathcal{R}$  at time  $s$  is only partially observed or unobserved between capture times  $s + 1$  and  $t$  with partial encounter history  $\mathbf{z}_{s+1:t}$  and is observed at time  $t + 1$ , in known state  $r_j \in \mathcal{R}$ . Then, for  $s \leq t$ ,

$$O_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t}) = Q_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t})p_{t+1}(r_j)\alpha_{t+1}(r_j). \quad (2)$$

2. Define  $D_{s,t}(r_i, \mathbf{z}_{s+1:t})$  to be the probability an individual in state  $r_i \in \mathcal{R}$  at time  $s$  is only partially observed or unobserved between times  $s + 1$  and  $t$  with partial encounter history given by  $\mathbf{z}_{s+1:t}$  and recovered dead in the interval  $(t, t + 1]$ . Then, for  $s \leq t$ ,

$$D_{s,t}(r_i, \mathbf{z}_{s+1:t}) = \begin{cases} \{1 - \phi_t(r_i)\}\lambda_t(r) & s = t \\ \sum_{u \in \mathcal{R}} Q_{s,t-1}(r_i, u, \mathbf{z}_{s+1:t-1})Z_t(u, z_{(t)})\{1 - \phi_t(u)\}\lambda_t(u) & s < t. \end{cases} \quad (3)$$

3. Define  $\gamma_{s,t}(r_i, \mathbf{z}_{s+1:t})$  to be the probability an individual in state  $r_i \in \mathcal{R}$  at time  $s$  is only partially observed or unobserved from time  $s + 1$  to the end of the study, such that they are observed for the final time at time  $t \geq s$  (in known or partial state) with partial encounter history  $\mathbf{z}_{s+1:t}$  (so that  $z_{(t+1)} = \dots = z_{(T)} = 0$  and when  $t > s$ ,  $z_{(t)} \in \mathcal{P}$ ). We let  $\chi_t(r_i)$  denote the probability that an individual in state  $r_i \in \mathcal{R}$  at time  $t$  is not observed after time  $t$  (either in a known or partial state). Then for  $r_i \in \mathcal{R}$ ,

$$\chi_t(r_i) = \begin{cases} 1 & t = T \\ \{1 - \phi_t(r_i)\}\{1 - \lambda_t(r_i)\} + \phi_t(r_i) \sum_{u \in \mathcal{R}} \psi_t(r_i, u)\{1 - p_{t+1}(u)\}\chi_{t+1}(u) & t < T; \end{cases} \quad (4)$$

and,

$$\gamma_{s,t}(r_i, \mathbf{z}_{s+1:t}) = \begin{cases} \chi_t(r_i) & s = t; \\ \sum_{u \in \mathcal{R}} Q_{s,t-1}(r_i, u, \mathbf{z}_{s+1:t-1})p_t(u)\beta_t(z_{(t)}, u)\chi_t(u) & s < t. \end{cases} \quad (5)$$

4. Define  $\zeta_{s,t}(m_i, r_j, \mathbf{z}_{s+1:t})$  to be the probability an individual that is initially observed at time  $s$  in partial state  $m_i \in \mathcal{P}$ , is only partially observed or unobserved from time  $s + 1$  to  $t$  with partial encounter history  $\mathbf{z}_{s+1:t}$  and is observed at time  $t + 1$ , when they are in known state  $r_j \in \mathcal{R}$ . Then, for  $s \leq t$ ,

$$\zeta_{s,t}(m_i, r_j, \mathbf{z}_{s+1:t}) = \sum_{u \in \mathcal{R}} \pi_s(u, m_i)O_{s,t}(u, r_j, \mathbf{z}_{s+1:t}). \quad (6)$$

5. Define  $\rho_{s,t}(m_i, \mathbf{z}_{s+1:t})$  to be the probability an individual not observed in a known state within the study, nor recovered dead in the study, has partial encounter history  $\mathbf{z}_{s+1:t}$ , such that they are observed for the final time at time  $t$ , (so that  $z_{(t+1)} = \dots = z_{(T)} = 0$  and when  $s < t$ ,  $z_{(t)} \in \mathcal{P}$ ), given that they are first observed within the study at time  $s$  in partial state  $m_i \in \mathcal{P}$ . Then, for  $s \leq t$ ,

$$\rho_{s,t}(m_i, \mathbf{z}_{s+1:t}) = \sum_{u \in \mathcal{R}} \pi_s(u, m_i)\gamma_{s,t}(u, \mathbf{z}_{s+1:t}). \quad (7)$$

Note that  $s = t$  corresponds to an individual only observed once in the study when they are observed in a partial state.

6. Define  $\omega_{s,t}(m_i, \mathbf{z}_{s+1:t})$  to be the probability an individual not observed in a known state within the study, is recovered dead in the time interval  $(t, t + 1]$  and has encounter history  $\mathbf{z}_{s+1:t}$ , given that they are first observed at time  $s$  in partial state  $m_i \in \mathcal{P}$ . Then, for  $s \leq t$ ,

$$\omega_{s,t}(m_i, \mathbf{z}_{s+1:t}) = \sum_{u \in \mathcal{R}} \pi_s(u, m_i)D_{s,t}(u, \mathbf{z}_{s+1:t}). \quad (8)$$

The corresponding likelihood of the data is simply the product of a series of multinomial distributions, where the “cells” of each multinomial distribution correspond to each distinct possible encounter history, conditional on the initial capture time and assigned state. Thus, the likelihood is in the form of the multinomial coefficient terms (which is simply a function of the observed data

and hence independent of the model parameters) multiplied by the product over each individual of the probability of their corresponding encounter histories. The probability of any encounter history is simply the product of the probabilities associated with the part histories of which it is composed. Thus, producting over all individuals observed within the study, the likelihood can be factored into each distinct part history to the power of the number of individuals with the given part history (i.e. the corresponding sufficient statistic). Letting the set of sufficient statistics be denoted by  $\eta = \{\mathbf{n}, \mathbf{d}, \mathbf{v}, \mathbf{w}, \mathbf{b}, \mathbf{h}\}$ , with model parameters  $\theta = \{\phi, \mathbf{p}, \lambda, \psi, \alpha, \beta, \pi\}$  the likelihood is expressible in the form,

$$\begin{aligned} L(\theta; \eta) \propto & \prod_{s=1}^{T-1} \prod_{t=s}^{T-1} \prod_{r_i \in \mathcal{R}} \prod_{\mathbf{z}_{s:t} \in \mathcal{Z}_{s:t}^*} \left[ \{D_{s,t}(r_i, \mathbf{z}_{s:t})\}^{d_{s,t}(r_i, \mathbf{z}_{s:t})} \prod_{r_j \in \mathcal{R}} \{O_{s,t}(r_i, r_j, \mathbf{z}_{s:t})\}^{n_{s,t}(r_i, r_j, \mathbf{z}_{s:t})} \right] \\ & \times \prod_{s=1}^{T-1} \prod_{t=s}^T \left[ \prod_{\substack{\mathbf{z}_{s+1:t-1} \in \mathcal{Z}_{s+1:t-1}^* \\ \mathbf{z}_{(t)} \in \mathcal{P}: t \neq s}} \left[ \prod_{r_i \in \mathcal{R}} [\{\gamma_{s,t}(r_i, \mathbf{z}_{s+1:t})\}^{v_{s,t}(r_i, \mathbf{z}_{s+1:t})}] \right. \right. \\ & \times \left. \left. \prod_{m_i \in \mathcal{P}} [\{\rho_{s,t}(m_i, \mathbf{z}_{s+1:t})\}^{b_{s,t}(m_i, \mathbf{z}_{s+1:t})}] \right] \right] \\ & \times \prod_{s=1}^{T-1} \prod_{t=s}^{T-1} \prod_{m_i \in \mathcal{P}} \prod_{\mathbf{z}_{s+1:t} \in \mathcal{Z}_{s+1:t}^*} \left[ \{\omega_{s,t}(m_i, \mathbf{z}_{s+1:t})\}^{h_{s,t}(m_i, \mathbf{z}_{s+1:t})} \right] \\ & \times \prod_{r_j \in \mathcal{R}} [\{\zeta_{s,t}(m_i, r_j, \mathbf{z}_{s+1:t})\}^{w_{s,t}(m_i, r_j, \mathbf{z}_{s+1:t})}]. \end{aligned}$$

Note that we use the convention that  $\prod_{\emptyset} \equiv 1$ .

The likelihood, although perhaps somewhat intuitive in nature, is complex to express. An alternative presentation of the likelihood (and data itself) is in matrix form, extending the likelihood specification of [3,25] for multi-state capture–recapture and capture–recovery data where all states are observed as known. We provide this alternative matrix specification next, before considering some special cases of these general likelihoods.

## 2.6. Matrix formulation

In this section we derive the corresponding matrix formulation of the above likelihood expression. The previous likelihood probability terms and results can be derived by expanding the corresponding matrix terms. For notational convenience we let  $\text{Diag}(\cdot)$  denote the diagonal matrix with ordered elements  $(\cdot)$ . We define the  $|\mathcal{R}| \times |\mathcal{R}|(|\mathcal{P}| + 1)$  matrix,

$$\mathbf{Z}_t = [\text{Diag}(\mathbf{1} - \mathbf{p}_t) \quad \text{Diag}(\mathbf{p}_t)\text{Diag}(\boldsymbol{\beta}_t(m_1)) \quad \cdots \quad \text{Diag}(\mathbf{p}_t)\text{Diag}(\boldsymbol{\beta}_t(m_{|\mathcal{P}|}))],$$

for  $t = 2, \dots, T$ , where  $\mathbf{1}$  denotes the row vector (of length  $|\mathcal{R}|$ ) with  $i$ th element equal to unity and  $\boldsymbol{\beta}_t(m_i)$  the  $i$ th row of the matrix  $\boldsymbol{\beta}_t$  (i.e. row vector of length  $|\mathcal{R}|$  with  $j$ th element  $\beta_t(m_i, r_j)$ ). This  $\mathbf{Z}_t$  matrix corresponds to the probabilities associated with not being observed, or observed and assigned a partial state at time  $t$ . For  $s \leq t$  we define,

$$\mathbf{Q}_{s,t} = \begin{cases} \text{Diag}(\boldsymbol{\phi}_t)\boldsymbol{\psi}_t & s = t; \\ \text{Diag}(\boldsymbol{\phi}_s)\boldsymbol{\psi}_s \mathbf{Z}_{s+1} [\mathbf{I}_{(|\mathcal{P}|+1)} \otimes \mathbf{Q}_{s+1,t}] & s < t, \end{cases}$$

where  $\mathbf{I}_{(|\mathcal{P}|+1)}$  is the  $(|\mathcal{P}| + 1) \times (|\mathcal{P}| + 1)$  identity matrix and  $\otimes$  denotes the Kronecker product. Thus,  $\mathbf{Q}_{s,t}$  is a matrix of dimension  $|\mathcal{R}| \times |\mathcal{R}|(|\mathcal{P}| + 1)^{t-s}$  corresponding to the probabilities that an



individual is partially observed between times  $s + 1$  and  $t$  (i.e. has partial encounter history  $\mathbf{z}_{s+1:t}$ ), and survives until time  $t + 1$ , given they are last observed in a known state at time  $s$ . We again consider each of the decomposed part histories in turn and provide their corresponding probabilities in matrix form.

1. For consecutive sightings in known states, at times  $s$  and  $t + 1$ , the associated probability matrix is,

$$\mathbf{O}_{s,t} = \mathbf{Q}_{s,t} \left[ \mathbf{I}_{\{(|\mathcal{P}|+1)^{t-s}\}} \otimes \text{Diag}(\mathbf{p}_{t+1}) \text{Diag}(\boldsymbol{\alpha}_{t+1}) \right],$$

which is of dimension  $|\mathcal{R}| \times |\mathcal{R}|(|\mathcal{P}| + 1)^{t-s}$ . Note that we use the notational convention that  $\mathbf{I}_0 \equiv \mathbf{1}$ .

2. From final sighting in known state at time  $s$  to being recovered dead in  $(t, t + 1]$ , the corresponding probability matrix is,

$$\mathbf{D}_{s,t} = \begin{cases} \text{Diag}(\mathbf{1} - \boldsymbol{\phi}_t) \boldsymbol{\lambda}_t^T & s = t; \\ \mathbf{Q}_{s,t-1} \left[ \mathbf{I}_{\{(|\mathcal{P}|+1)^{t-s-1}\}} \otimes \mathbf{Z}_t \right] \left[ \mathbf{I}_{\{(|\mathcal{P}|+1)^{t-s}\}} \otimes \text{Diag}(\mathbf{1} - \boldsymbol{\phi}_t) \boldsymbol{\lambda}_t^T \right] & s < t, \end{cases}$$

and is of dimension  $|\mathcal{R}| \times (|\mathcal{P}| + 1)^{t-s}$ .

3. From final sighting in known state at time  $s$  to the end of the study, when observed for the final time (in known or partial state) at time  $t$ . We initially define the column vector of length  $|\mathcal{R}|$  for the probabilities of not being observed again after time  $t$ ,

$$\boldsymbol{\chi}_t = \begin{cases} \mathbf{1}^T & t = T; \\ \text{Diag}(\mathbf{1} - \boldsymbol{\phi}_t)(\mathbf{1} - \boldsymbol{\lambda}_t)^T + \text{Diag}(\boldsymbol{\phi}_t) \boldsymbol{\psi}_t \text{Diag}(\mathbf{1} - \mathbf{p}_{t+1}) \boldsymbol{\chi}_{t+1} & t < T. \end{cases}$$

The probability matrix for the associated decomposed part history is,

$$\boldsymbol{\gamma}_{s,t} = \begin{cases} \boldsymbol{\chi}_t & s = t; \\ \mathbf{Q}_{s,t-1} \left[ \mathbf{I}_{\{(|\mathcal{P}|+1)^{t-s-1}\}} \otimes \text{Diag}(\mathbf{p}_t) \text{Diag}(\boldsymbol{\chi}_t) \boldsymbol{\beta}_t^T \right] & s < t, \end{cases}$$

and is of dimension  $|\mathcal{R}| \times \{|\mathcal{P}|(|\mathcal{P}| + 1)^{t-s-1}\}^{\mathbb{I}_{\{s < t\}}}$ .

4. For initial capture at time  $s$  in a partially observed state to the first observed time in a known state at time  $t + 1$ , the corresponding probability matrix is given by,

$$\boldsymbol{\zeta}_{s,t} = \boldsymbol{\pi}_s^T \mathbf{O}_{s,t},$$

so that  $\boldsymbol{\zeta}_{s,t}$  is of dimension  $|\mathcal{P}| \times |\mathcal{R}|(|\mathcal{P}| + 1)^{t-s}$ .

5. For capture histories with no known observed states (including death), initial capture at time  $s$  and final observed sighting at  $t$ , the associated probability matrix for the capture history is,

$$\boldsymbol{\rho}_{s,t} = \boldsymbol{\pi}_s^T \boldsymbol{\gamma}_{s,t},$$

and is of dimension  $|\mathcal{P}| \times \{|\mathcal{P}|(|\mathcal{P}| + 1)^{t-s-1}\}^{\mathbb{I}_{\{s < t\}}}$ .

6. Finally, the probability matrix for capture histories with no live known observed states, initial capture at time  $s$  and subsequently recovered dead in  $(t, t + 1]$  is,

$$\boldsymbol{\omega}_{s,t} = \boldsymbol{\pi}_s^T \mathbf{D}_{s,t},$$

and is of dimension  $|\mathcal{P}| \times (|\mathcal{P}| + 1)^{t-s}$ .

Conditional on the initial capture time and assigned state, we have a multinomial distribution for the number of individuals with each possible encounter history, which can be written as a function of the decomposed part histories,

$$L(\mathbf{x}; \mathbf{p}, \boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) \propto \prod_{s=1}^{T-1} \left( \prod_{t=s}^{T-1} \left[ \mathbf{O}_{s,t}^{n_{s,t}} \mathbf{D}_{s,t}^{d_{s,t}} \boldsymbol{\zeta}_{s,t}^{w_{s,t}} \boldsymbol{\omega}_{s,t}^{h_{s,t}} \right] \prod_{t=s}^T \left[ \boldsymbol{\gamma}_{s,t}^{v_{s,t}} \boldsymbol{\rho}_{s,t}^{b_{s,t}} \right] \right),$$

using the convention that for matrices  $\mathbf{A}$  and  $\mathbf{a}$  each of dimension  $(M \times N)$ , the term  $\mathbf{A}^{\mathbf{a}} = \prod_{i=1}^M \prod_{j=1}^N A(i, j)^{a(i, j)}$ , where  $A(i, j)$  and  $a(i, j)$  denote the  $(i, j)$ th elements of  $\mathbf{A}$  and  $\mathbf{a}$ .

### 3. Special cases

We present a number of “standard” models as special cases of partially observed data (though note that this list is not exhaustive) and provide the corresponding likelihood expressions, before considering how the likelihood can be used for absolute goodness-of-fit.

#### 3.1. Arnason–Schwarz model

The standard AS model considers only mark–recapture–recovery data where all states are known without error when an individual is observed (i.e.  $\mathbf{z} = \mathbf{0}$ ), so that,

$$\mathbf{Q}_{s,t} = \begin{cases} \text{Diag}(\boldsymbol{\phi}_t)\boldsymbol{\psi}_t & s = t; \\ \text{Diag}(\boldsymbol{\phi}_s)\boldsymbol{\psi}_s\text{Diag}(\mathbf{1} - \mathbf{p}_{s+1})\mathbf{Q}_{s+1,t} & s < t. \end{cases}$$

Then  $\mathbf{O}_{s,t} = \mathbf{Q}_{s,t}\text{Diag}(\mathbf{p}_{t+1})$  and is of dimension  $|\mathcal{R}| \times |\mathcal{R}|$ . An individual is always observed for the final time in a known state (as all states are known without error), so that

$$\boldsymbol{\chi}_t = \text{Diag}(\boldsymbol{\phi}_t)\boldsymbol{\psi}_t\text{Diag}(\mathbf{1} - \mathbf{p}_{t+1})\boldsymbol{\chi}_{t+1},$$

for  $t < T$ . The likelihood reduces to,

$$L(\boldsymbol{\chi}; \boldsymbol{\phi}, \mathbf{p}, \boldsymbol{\psi}) \propto \prod_{s=1}^{T-1} \left( \boldsymbol{\chi}_s^{\mathbf{v}_{s,s}} \prod_{t=s}^{T-1} [\mathbf{O}_{s,t}^{\mathbf{n}_{s,t}}] \right),$$

where  $\mathbf{v}_{s,s}$  denotes the column vector of length  $|\mathcal{R}|$  corresponding to the number of individuals observed for the final time at time  $s$ . This expression is equivalent to the matrix form likelihood presented in [3]. Including dead recoveries simply involves including the additional part histories that end with the observed deaths. In particular, setting,

$$\mathbf{D}_{s,t} = \mathbf{Q}_{s,t-1}\text{Diag}(\mathbf{1} - \boldsymbol{\phi}_t)\boldsymbol{\lambda}_t^T,$$

and using the form of  $\boldsymbol{\chi}_t$  provided in case 3 of Section 2.6. The likelihood is,

$$L(\boldsymbol{\chi}; \boldsymbol{\phi}, \mathbf{p}, \boldsymbol{\psi}) \propto \prod_{s=1}^{T-1} \left( \boldsymbol{\chi}_s^{\mathbf{v}_{s,s}} \prod_{t=s}^{T-1} [\mathbf{O}_{s,t}^{\mathbf{n}_{s,t}} \mathbf{D}_{s,t}^{\mathbf{d}_{s,t}}] \right).$$

#### 3.2. Observation error model

In this case the state of an individual is always observed with error so that all assigned states are partially observed (i.e.  $\mathbf{y} = \mathbf{0}$ ). The likelihood contains only two terms with

$$L(\boldsymbol{\chi}; \boldsymbol{\phi}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\pi}) \propto \prod_{s=1}^{T-1} \left( \prod_{t=s}^T [\boldsymbol{\rho}_{s,t}^{\mathbf{b}_{s,t}}] \prod_{t=s}^{T-1} [\boldsymbol{\omega}_{s,t}^{\mathbf{h}_{s,t}}] \right),$$

for  $\boldsymbol{\rho}_{s,t}$  and  $\boldsymbol{\omega}_{s,t}$  given in Section 2.6.

#### 3.3. Pledger’s mixture model

The mixture models proposed by [21] assume there are  $N$  unobservable sub-populations with each individual belonging to one sub-population. Thus  $\mathcal{R} = \{1, 2, \dots, N\}$ ,  $\mathcal{P} = \{U\}$  ( $U \equiv$  “unknown”) and we have the following simplifications:

1.  $\boldsymbol{\psi}_t = \mathbf{I}_{|\mathcal{R}|}$  for  $t = 1, \dots, T-1$ : (an individual does not change state i.e. sub-population);
2.  $\boldsymbol{\beta}_t = \mathbf{1}$  and  $\boldsymbol{\alpha} = \mathbf{0}$  for  $t = 2, \dots, T$ : (all individuals are assigned state  $U$  so  $|\mathcal{P}| = 1$ );
3.  $\mathbf{Z}_t = [\text{Diag}(\mathbf{1} - \mathbf{p}_t)\text{Diag}(\mathbf{p}_t)]$ : (since  $|\mathcal{P}| = 1$  and  $\boldsymbol{\beta}_t = \mathbf{1}$ );

4.  $\pi_t = (\pi(1), \dots, \pi(N))^T$  for all  $t = 1, \dots, T - 1$ , where  $\pi(r)$  denotes the (time-invariant) probability of belonging to sub-population  $r \in \mathcal{R}$ .

As for the observation error model, no individuals are observed in a known state, (i.e.  $\mathbf{y} = \mathbf{0}$ ), so that the likelihood (allowing for observable deaths) is

$$L(\mathbf{x}; \boldsymbol{\phi}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\pi}) \propto \prod_{s=1}^{T-1} \left( \prod_{t=s}^T [\boldsymbol{\rho}_{s,t}^{b_{s,t}}] \prod_{t=s}^{T-1} [\boldsymbol{\omega}_{s,t}^{h_{s,t}}] \right),$$

for  $\boldsymbol{\rho}_{s,t}$  and  $\boldsymbol{\omega}_{s,t}$  provided in Section 2.6, with the above simplifications substituted into the relevant probability matrices.

### 3.4. Tag-loss model

We consider a double-tagging capture–recapture study, where the state of an individual corresponds to the number of tags (an individual with zero tags cannot be identified and hence cannot be recaptured), so that  $\mathcal{R} = \{0, 1, 2\}$ . Partial observations may occur when an individual is observed, but it is unknown whether they have 1 or 2 tags so that  $\mathcal{P} = \{1+\}$  (i.e. they are observed and identified by one tag, but the presence/absence of the other tag is unknown). Assuming that each tag is lost independently of each other,

$$\boldsymbol{\psi}_t = \begin{pmatrix} 1 & 0 & 0 \\ \tau & 1 - \tau & 0 \\ \tau^2 & 2\tau(1 - \tau) & (1 - \tau)^2 \end{pmatrix},$$

for all  $t = 1, \dots, T - 1$ , where  $\tau$  denotes the probability of losing a single tag (and for simplicity is assumed to be constant over time). Individuals with 0 tags (i.e. individuals who have lost all the tags applied) cannot be observed, so that  $\mathbf{p}_t = (0, p_t(1), p_t(2))$ . Additionally, assuming that survival probabilities are independent of tags applied,  $\phi_t(r) = \phi_t$  for all  $r \in \mathcal{R}$ .

### 3.5. Goodness-of-fit

The likelihood construction, written as a function of non-overlapping sufficient statistics, permits the application of absolute goodness-of-fit tests. In theory, the approach of [20] can be directly applied to partially observed mark–recapture–recovery data. In particular the observed matrices of sufficient statistics can be compared with the expected values through the evaluation of the corresponding probability matrices, given the specified model. However, we note that in any real application (such as the example in Section 4), the pooling of cells will be necessary, due to small expected values (see [20] and Section 5 for further discussion).

## 4. Example

We consider partially observed multi-state capture–recapture data (i.e. there are no dead recoveries) relating to house finches (*Carpodacus mexicanus*), where the state corresponds to the presence/absence of the disease *Mycoplasma gallisepticum* conjunctivitis. We let the set of observable states be denoted by  $\mathcal{R} = \{1, 2\}$ , where 1  $\equiv$  absence and 2  $\equiv$  presence of the disease, respectively. However, when an individual is observed, their state may also be recorded as “unknown”, and we set  $\mathcal{P} = \{U\}$ . The data we consider consists of  $N = 813$  individuals with  $T = 16$  capture events. There are a total of 1619 captures (both initial captures and recaptures) of which 1374 record the absence of conjunctivitis, 178 record the presence of conjunctivitis and 67 record an unknown disease state (corresponding to approximately 4% of the captures). A total of 59 individuals (approximately 7%) have an unknown disease status recorded within their encounter history. For further discussion of the data and previous analyses see [7] (for a classical analysis using E-SURGE) and [24] (for a Bayesian analysis).

**Table 1**

The total number and the corresponding mean of the non-zero sufficient statistics for each of the sets of sufficient statistics corresponding to the different decomposed part encounter histories. Note that we omit the terms **d** and **h** as these correspond to dead recoveries which are not observed for these data.

Sufficient statistic	Total number of non-zero terms	Mean of non-zero terms
<b>n</b>	167	4.5
<b>v</b>	48	15.1
<b>w</b>	6	1
<b>b</b>	10	1.4

We initially consider the sufficient statistics for these data. Table 1 provides the number of non-zero sufficient statistics and the corresponding mean of the non-zero cell entries for each set of sufficient statistics. The majority of the sufficient statistics are zero, thus reducing the number of necessary calculations to evaluate the likelihood function. In addition due to the recursive formulae for the calculation of the different decomposed capture histories (provided in Section 2.5, or the equivalent matrix specification in Section 2.6) the number of necessary calculations in order to evaluate the likelihood is again reduced (this is essentially due to the Markovian structure of the data).

We note that, as expected, the largest number of non-zero terms for the sufficient statistics corresponds to the statistic **n**, relating to consecutive sightings in known states. For these data, this summary statistic can be decomposed further into the number of consecutive sightings in known states with 0, 1 or 2 captures in between these times where the state is recorded as unknown, with 153, 13 and 1 non-zero elements, respectively (with the corresponding mean of the non-zero terms given by 4.8, 1.5 and 1, respectively). Similarly, the number of non-zero elements of **v** can be decomposed further with regard to the number of times (0 or 1) an individual is observed in an unknown state following final sighting in a known state. All possible part histories (a total of 30) are observed corresponding to being observed for the final time (at time  $s = 1, \dots, T - 1$ ) in a known state and not being observed again (with a mean of these non-zero terms of 23.5). There are a further 18 distinct part histories of being observed in a known state followed by being observed one further time but in an unknown state (the mean of these non-zero terms is 1.2).

Following a model search algorithm using the AIC statistic to discriminate between models, the model  $p(t + r)/\phi(t + r)/\psi/\alpha(r)$  is identified as optimal, corresponding to time ( $t$ ) and state ( $r$ ) dependent capture and survival probabilities (additive on the logit scale), time-independent transition probabilities and state dependent state assignment probabilities (where the probability of positively identifying an individual with conjunctivitis is greater than the probability of identifying the individual as not having conjunctivitis, i.e.  $\alpha(2) > \alpha(1)$ , see [7] for further details). Note that due to the sum to unity constraint of the  $\alpha$  and  $\beta$  terms (i.e. known and partial state assignments), this necessarily implies a state dependent partial state assignment probability. This model also identifies a reduced survival probability for an individual with conjunctivitis coupled with an increased capture probability. The same model is identified by [7] who also provide further discussion of the model and associated parameter estimates.

Of particular interest may be the (neighbouring) model  $p(t + r)/\phi(t + r)/\psi/\alpha$ , with the corresponding  $\Delta$  AIC of 5 when compared to the optimal model. This model assumes that the state assignment probabilities are no longer state dependent, i.e. the probability of observing an individual house finch in a known state (with or without conjunctivitis) is independent of their underlying state (see for example [24] for further discussion). In other words, the unknown state values (conditional on an individual being observed) are not missing at random (we note further that due to the dependence of the capture probabilities on the state, the missing at random assumption is again violated).

## 5. Discussion

We have presented a closed form expression for the likelihood of partially observed multi-state mark-recapture-recovery models defined via non-overlapping sufficient statistics. This likelihood unifies many commonly used capture-recapture-recovery models into a single framework. The different models can be seen to be of modular structure, corresponding to the union of the possible

decomposed part encounter histories, with the likelihood simply the product over the probabilities for the part histories observed within any given dataset. The recursive nature of the terms within the likelihood provides an efficient formulation, using the underlying Markovian structure of the model. The further matrix construction of the likelihood is an immediate and concise specification and allows the use of efficient matrix calculations within computer programs for model fitting.

The likelihood construction of this paper in terms of non-overlapping sufficient statistics provides a mechanism for the assessment of absolute goodness-of-fit due to its relationship with the King and Brooks [14] likelihood [20]. This approach compares the observed sufficient statistics with their corresponding expected values, for the given model. However, this will typically involve significant pooling of matrix (or cell) entries to ensure that the goodness-of-fit asymptotic approximation is valid. For example, for the dataset in Section 4 the majority of the sufficient statistics of the observed data are equal to 0; and of the non-zero cell entries 74% of these have an observed value of less than 5. Clearly, there are numerous ways in which any form of pooling may be implemented and investigating different possible approaches is an area of current research.

The general likelihood assumes full time dependence on the parameters. However, sub-models will often be fitted with restrictions placed on the parameters. An explicit likelihood function permits model selection to be performed in a standard framework, using likelihood-ratio tests and/or information criteria. For example, there may be interest in whether the states are missing at random (as in Section 4) and can be investigated within this modelling framework. We note that for multi-state data, model discrimination can be non-trivial due to the large number of possible models contained in a candidate model set. Some structured solutions have been presented for the related multi-state models with known state [13,18], and such approaches can be extended to partially observed mark–recapture–recovery data. In addition, the issue of parameter redundancy for complex models may appear. An explicit likelihood expression structured in terms of multinomial probabilities allows formal parameter redundancy tests to be developed for given models [6]. Applying these methods to the generalised likelihood form is a focus of current research.

## Acknowledgements

We wish to thank Byron Morgan for interesting discussions and Evan Cooch and Paul Conn for providing the data used in Section 4. McCrea was funded by the NERC fellowship grant NE/J018473/1.

## Appendix. Derivation of probability results in Section 2.5

We note that the results for the probabilistic statement given in Section 2.5 are all applications of the Chapman–Kolmogorov equations.

**Proof of Eq. (2).** We initially consider the form given for  $Q_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t})$  in Eq. (1). Initially consider the case  $s = t$  (i.e. when individuals are observed at consecutive capture events), so that,

$$\begin{aligned} Q_{t,t}(r_i, r_j, \mathbf{z}_{t+1:t} = \emptyset) &= \mathbb{P}(\text{individual is in state } r_j \in \mathcal{R} \text{ at time } t + 1 \mid \text{in state } r_i \text{ at time } t) \\ &= \phi_t(r_i)\psi_t(r_i, r_j). \end{aligned}$$

We now consider the more general case for  $s < t$ ,

$$\begin{aligned} Q_{s,t}(r_i, r_j, \mathbf{z}_{s+1:t}) &= \mathbb{P}(\text{individual has partial history } \mathbf{z}_{s+1:t} \text{ and in state } r_j \in \mathcal{R} \text{ at time } t + 1 \\ &\quad \mid \text{in state } r_i \text{ at time } s) \\ &= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{individual has partial history } \mathbf{z}_{s+1:t} \text{ and in state } r_j \in \mathcal{R} \text{ at} \\ &\quad \text{time } t + 1 \mid \text{in state } r_i \text{ at time } s \text{ and state } u \text{ at time } s + 1) \\ &\quad \times \mathbb{P}(\text{individual in state } u \text{ at time } s + 1 \mid \text{in state } r_i \text{ at time } s) \\ &= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{individual has partial history } \mathbf{z}_{s+2:t} \text{ and in state } r_j \in \mathcal{R} \text{ at} \end{aligned}$$

$$\begin{aligned}
& \text{time } t + 1 \mid \text{in state } u \text{ at time } s + 1 \text{ and assigned partial encounter} \\
& z_{(s+1)} \text{ at time } s + 1) \\
& \times \mathbb{P}(\text{partial encounter observation } z_{(s+1)} \text{ at time } s + 1 \mid \text{in state } u \text{ at} \\
& \text{time } s + 1) \phi_s(r_i) \psi_s(r_i, u) \\
& = \phi_s(r_i) \sum_{u \in \mathcal{R}} \psi_s(r_i, u) Q_{s+1,t}(u, r_j, \mathbf{z}_{s+2:t}) Z_{s+1}(u, z_{(s+1)}).
\end{aligned}$$

**Proof of Eq. (3).** We again consider two different cases. We begin by considering the probability that an individual is recovered dead at time  $t + 1$ , given they are observed in known state  $r_i \in \mathcal{R}$  at time  $t$  (i.e.  $s = t$ ). Then,

$$\begin{aligned}
D_{t,t}(r_i, \mathbf{z}_{t+1:t} = \emptyset) &= \mathbb{P}(\text{individual dies and is recovered in } (t, t + 1] \mid \text{in state } r_i \text{ at time } t) \\
&= \{1 - \phi_t(r_i)\} \lambda_t(r_i).
\end{aligned}$$

Now consider the more general case, where an individual is last observed in known state at time  $s$  and recovered dead in the interval  $(t, t + 1)$  (for  $s < t$ ). Then, we have that,

$$\begin{aligned}
D_{s,t}(r_i, \mathbf{z}_{s+1:t}) &= \mathbb{P}(\text{individual dies and is recovered in } (t, t + 1] \text{ and has partial encounter} \\
& \text{history } \mathbf{z}_{s+1:t} \mid \text{individual is in state } r_i \text{ at time } s) \\
&= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{individual dies and is recovered in } (t, t + 1], \text{ has partial encounter} \\
& \text{history } \mathbf{z}_{s+1:t} \text{ and is in state } u \text{ at time } t \mid \text{individual in state } r_i \text{ at time } s) \\
&= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{individual dies and is recovered in } (t, t + 1] \mid \text{in state } u \text{ at time } t) \\
& \quad \times \mathbb{P}(\text{partial encounter history } \mathbf{z}_{s+1:t} \text{ and in state } u \text{ at time } t \mid \text{in} \\
& \quad \text{state } r_i \text{ at time } s) \\
&= \sum_{u \in \mathcal{R}} \{1 - \phi_t(u)\} \lambda_t(u) \\
& \quad \times \mathbb{P}(\text{individual has assigned partial encounter } z_{(t)} \text{ at time } t \mid \text{in} \\
& \quad \text{state } u \text{ at time } t) \\
& \quad \times \mathbb{P}(\text{partial encounter history } \mathbf{z}_{s+1:t-1} \text{ and in state } u \text{ at time } t \mid \text{in} \\
& \quad \text{state } r_i \text{ at time } s) \\
&= \sum_{u \in \mathcal{R}} \{1 - \phi_t(u)\} \lambda_t(u) Z_t(u, z_{(t)}) Q_{s,t-1}(r_i, u, \mathbf{z}_{s+1:t-1}).
\end{aligned}$$

**Proof of Eq. (5).** Consider the probability that an individual observed in known state  $r_i \in \mathcal{R}$  at time  $t$  is not observed again, denoted by  $\chi_t(r_i)$ . [14] derive the result in Eq. (4) and so this proof is omitted. Now consider an individual in known state  $r_i \in \mathcal{R}$  at time  $s$  observed for the final time at time  $t$  (in known or partial state). Initially consider  $s = t$ . By definition, it is immediate that for all  $r_i \in \mathcal{R}$ ,

$$\gamma_{t,t}(r_i, \mathbf{z}_{t+1:t} = \emptyset) = \chi_t(r_i).$$

Now consider the case where  $t > s$  so that the individual is observed for the final time in a known state at time  $s$  and is observed in a partial state for the final time at time  $t$  (so that  $z_{(t)} \in \mathcal{P}$ ) and has partial encounter history  $\mathbf{z}_{s+1:t}$ . Then,

$$\begin{aligned}
\gamma_{s,t}(r_i, \mathbf{z}_{s+1:t}) &= \mathbb{P}(\text{individual has encounter history } \mathbf{z}_{s+1:t} \\
& \mid \text{in state } r_i \text{ at time } s \text{ and not observed in a known state after time } s) \\
&= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{individual has encounter history } \mathbf{z}_{s+1:t} \text{ and in state } u \text{ at time } t \\
& \mid \text{in state } r_i \text{ at time } s \text{ and not observed in a known state after time } s)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{encounter history } \mathbf{z}_{s+1:t-1} \text{ and in state } u \text{ at time } t \mid \text{in state } r_i \text{ at time } s) \\
&\quad \times \mathbb{P}(\text{individual observed at time } t \text{ and assigned state } z_{(t)} \mid \text{in state } u \text{ at time } t) \\
&\quad \times \mathbb{P}(\text{individual not observed after time } t \mid \text{in state } u \text{ at time } t) \\
&= \sum_{u \in \mathcal{R}} Q_{s,t-1}(r_i, u, \mathbf{z}_{s+1:t-1}) p_t(u) \beta_t(z_{(t)}, u) \chi_t(u).
\end{aligned}$$

**Proof of Eq. (6).** Consider the probability of observing an individual in known state  $r_j \in \mathcal{R}$  for the first time at time  $t + 1$  and have partial encounter history  $\mathbf{z}_{s+1:t}$ , given they are initially observed at time  $s$  in partial state  $m_i \in \mathcal{P}$ . For  $s \leq t$ , the corresponding probability is given by,

$$\begin{aligned}
\zeta_{s,t}(m_i, r_j, \mathbf{z}_{s+1:t}) &= \mathbb{P}(\text{partial encounter history } \mathbf{z}_{s+1:t} \text{ and observed at time } t + 1 \text{ in known} \\
&\quad \text{state } m_i \in \mathcal{R} \mid \text{initially observed and assigned partial state } m_i \text{ at time } s) \\
&= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{partial encounter history } \mathbf{z}_{s+1:t} \text{ and observed at time } t + 1 \text{ in} \\
&\quad \text{known state } m \in \mathcal{R} \mid \text{in state } u \text{ at time } s) \\
&\quad \times \mathbb{P}(\text{individual is in state } u \text{ at time } s \mid \text{assigned partial state } m_i \text{ at time } s) \\
&= \sum_{u \in \mathcal{R}} \pi_s(u, m_i) O_{s,t}(u, r_j, \mathbf{z}_{s+1:t}).
\end{aligned}$$

**Proof of Eq. (7).** Consider the capture history of an individual with encounter history  $\mathbf{z}_{s+1:t}$  such that they are not recovered dead within the study and are observed for the final time at time  $t$  in partial state  $z_{(t)} \in \mathcal{P}$  given they are initially observed at time  $s$  in partial state  $m_i \in \mathcal{P}$ . The corresponding probability is given by,

$$\begin{aligned}
\rho_{s,t}(m_i, \mathbf{z}_{s+1:t}) &= \mathbb{P}(\text{individual has partial encounter history } \mathbf{z}_{s+1:t} \text{ and observed for the last} \\
&\quad \text{time at time } t \mid \text{initially observed at time } s \text{ and assigned partial state } m_i) \\
&= \sum_{u \in \mathcal{R}} \mathbb{P}(\text{partial encounter history } \mathbf{z}_{s+1:t} \text{ and observed for the last time at} \\
&\quad \text{time } t \mid \text{in state } u \text{ at time } s) \\
&\quad \times \mathbb{P}(\text{individual in state } u \text{ at time } s \mid \text{assigned partial state } m_i \text{ at time } s) \\
&= \sum_{u \in \mathcal{R}} \pi_s(u, m_i) \chi_{s,t}(u, \mathbf{z}_{s+1:t}).
\end{aligned}$$

**Proof of Eq. (8).** This proof follows analogously to that for Eq. (7) and so is omitted for brevity.

## References

- [1] A.N. Arnason, Parameter estimates from mark–recapture–recovery experiments on two populations subject to migration and death, *Researches on Population Ecology* 13 (1972) 97–113.
- [2] A.N. Arnason, The estimation of population size, migration rates, and survival in a stratified population, *Researches on Population Ecology* 15 (1973) 1–8.
- [3] C. Brownie, J.E. Hines, J.D. Nichols, K.H. Pollock, J.B. Hestbeck, Capture–recapture studies for multiple strata including non-Markovian transitions, *Biometrics* 49 (1993) 1173–1187.
- [4] E.A. Catchpole, S.N. Freeman, B.J.T. Morgan, M.P. Harris, Integrated recovery/recapture data analysis, *Biometrics* 54 (1998) 33–46.
- [5] R. Choquet, L. Rouan, R. Pradel, Program E-SURGE: a software application for fitting multievent models, in: David L. Thomson, Evan G. Cooch, Michael J. Conroy (Eds.), *Series: Environmental and Ecological Statistics*, 2009, pp. 845–865.
- [6] D.J. Cole, B.J.T. Morgan, D.M. Titterton, Determining the parametric structure of non-linear models, *Mathematical Biosciences* 228 (2010) 16–30.
- [7] P.B. Conn, E.G. Cooch, Multistate capture–recapture analysis under imperfect state observation: an application to disease models, *Journal of Applied Ecology* 46 (2009) 486–492.
- [8] R.M. Cormack, Estimates of survival from the sighting of marked animals, *Biometrika* 51 (1964) 429–438.
- [9] J.A. Dupuis, Bayesian estimation of movement and survival probabilities from capture–recapture data, *Biometrika* 82 (1995) 761–772.

- [10] G.M. Jolly, Explicit estimates from capture–recapture data with both death and immigration–stochastic model, *Biometrika* 52 (1965) 225–247.
- [11] R. King, A review of Bayesian state–space modelling of capture–recapture–recovery data, *Interface Focus* 2 (2012) 190–204.
- [12] R. King, *Statistical ecology*, *Annual Review of Statistics and its Application* (2014) in press.
- [13] R. King, S.P. Brooks, Bayesian model discrimination for multiple strata capture–recapture data, *Biometrika* 89 (2002) 785–806.
- [14] R. King, S.P. Brooks, Closed-form likelihoods for Arnason–Schwarz models, *Biometrika* 90 (2003) 435–444.
- [15] R. King, B.J.T. Morgan, O. Gimenez, S.P. Brooks, *Bayesian Analysis for Population Ecology*, CRC Press, Boca Raton, 2009.
- [16] J.-D. Lebreton, T. Almeras, R. Pradel, Competing events, mixtures of information and multistratum recapture models, *Bird Study* 46 (1999) S39–S46.
- [17] J.D. Lebreton, J.D. Nichols, R. Barker, R. Pradel, J. Spindelw, Modeling individual animal histories with multistate capture–recapture models, in: H. Caswell (Ed.), *Advances in Ecological Research*, Vol. 41, 2009, pp. 81–173.
- [18] R.S. McCrea, B.J.T. Morgan, Multi-state mark–recapture model selection using score tests, *Biometrics* 67 (2011) 234–241.
- [19] R.S. McCrea, B.J.T. Morgan, *Analysis of Capture–Recapture Data*, CNC Press, Boca Raton, 2014.
- [20] R.S. McCrea, B.J.T. Morgan, T. Bregnballe, Model comparison and assessment for multi-state capture–recapture–recovery models, *Journal of Ornithology* 152 (2011) 293–391.
- [21] S. Pledger, K.H. Pollock, J.L. Norris, Open capture–recapture models with heterogeneity: I. Cormack–Jolly–Seber model, *Biometrics* 59 (2003) 786–794.
- [22] R. Pradel, Multievent: an extension of multistate capture–recapture models to uncertain states, *Biometrics* 61 (2005) 442–447.
- [23] T. Reynolds, R. King, J. Harwood, M. Frederiksen, S. Wanless, M. Harris, Integrated data analyses in the presence of emigration and tag-loss, *Journal of Agricultural, Biological, and Environmental Statistics* 14 (2008) 411–431.
- [24] M.R. Schofield, R.J. Barker, Full open population capture–recapture models with individual covariates, *Journal of Agricultural, Biological, and Environmental Statistics* 16 (2011) 253–268.
- [25] C.G. Schwarz, J.F. Schweigert, A.N. Arnason, Estimating migration rates using tag recovery data, *Biometrics* 59 (1993) 291–318.
- [26] C.J. Schwarz, G.A.F. Seber, A review of estimating animal abundance III, *Statistical Science* 14 (1999) 427–456.
- [27] G.A.F. Seber, A note on the multiple-recapture census, *Biometrics* 52 (1965) 249–259.
- [28] S.C. Smout, R. King, P.P. Pomeroy, Integrating heterogeneity of detection and mark loss to estimate survival and transience in UK colonies of great seal, *Journal of Applied Ecology* 48 (2011) 364–372.
- [29] M.A. Tanner, W.H. Wong, The calculation of posterior distribution by data augmentation, *Journal of the American Statistical Association* 82 (1987) 528–540.