 DOI: 10.1111/biom.12701

Hidden Markov Models for Extended Batch Data

Laura L. E. Cowen,[1*] Panagiotis Besbeas,[2,3,] Byron J. T. Morgan,[3,***] and Carl J. Schwarz[4,****]**

[1]Mathematics and Statistics, University of Victoria, PO Box 1700 STN CSC, Victoria BC, Canada, V8W 2Y2
[2]Department of Statistics, Athens University of Business and Economics, 10434 Athens, Greece
[3]National Centre for Statistical Ecology, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7FS, England
[4]Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada
email: lcowen@uvic.ca
**email:* P.T.Besbeas@kent.ac.uk
***email:* B.J.T.Morgan@kent.ac.uk
****email:* cschwarz@stat.sfu.ca

Summary. Batch marking provides an important and efficient way to estimate the survival probabilities and population sizes of wild animals. It is particularly useful when dealing with animals that are difficult to mark individually. For the first time, we provide the likelihood for extended batch-marking experiments. It is often the case that samples contain individuals that remain unmarked, due to time and other constraints, and this information has not previously been analyzed. We provide ways of modeling such information, including an open N-mixture approach. We demonstrate that models for both marked and unmarked individuals are hidden Markov models; this provides a unified approach, and is the key to developing methods for fast likelihood computation and maximization. Likelihoods for marked and unmarked individuals can easily be combined using integrated population modeling. This allows the simultaneous estimation of population size and immigration, in addition to survival, as well as efficient estimation of standard errors and methods of model selection and evaluation, using standard likelihood techniques. Alternative methods for estimating population size are presented and compared. An illustration is provided by a weather-loach data set, previously analyzed by means of a complex procedure of constructing a pseudo likelihood, the formation of estimating equations, the use of sandwich estimates of variance, and piecemeal estimation of population size. Simulation provides general validation of the hidden Markov model methods developed and demonstrates their excellent performance and efficiency. This is especially notable due to the large numbers of hidden states that may be typically required

Key words: Batch marking; Integrated population modeling; Mark-recapture; Open N-mixture models; Viterbi algorithm; Weather-loach.

1. Introduction

The standard protocol for capture–recapture studies of animals is to use individually numbered tags so that the capture history can be constructed, determining whether an individual was captured at each sampling occasion. Many sophisticated models have been built for this type of data (McCrea and Morgan, 2014). However, it may not be feasible to use individually numbered tags for some species because the small size of the animal makes it difficult to use a unique tag, or for reasons of cost and convenience. In these cases, batch marks can be used instead.

In extended batch marking, individuals are captured at several sampling occasions. Some of the unmarked individuals are given a common batch-mark, with different marks applied at different sampling occasions; without loss of generality, we shall talk in terms of different colored tags. At each subsequent capture occasion, marked individuals are counted and their tag colors are noted. A random sample of unmarked individuals are given a new colored mark and then all marked animals are released. Some unmarked individuals do not receive tags, and these animals may or may not be released; in the case of the weather-loach data that we analyze in this article, the unmarked individuals that are not tagged were not released.

There are many examples of batch marking in the literature, involving, for example, species of fish, insects, and amphibians. A wide variety of batch marks are available that are often species dependent, including tattoo, brand, fin-clip, o-rings, dyes, polymer, antibiotic, and radioisotope marks as possible fisheries batch marks, and dust, paint, dye, painted labels, self-marking baits, wire tags, genetic markers, and mutilation for marking insects.

Closed population models have been developed for extended batch-mark studies, aiming at estimating population size. For two-sample studies the Petersen estimator can be used to estimate population size and the Schnabel

estimator (under model M_t in Otis et al., 1978) is used with more than two sample occasions (Pine et al., 2003). Individual fish can be marked and released upriver, then caught downriver to enumerate salmon runs; these studies can be analyzed using a ratio estimator developed by Laplace (Rawson, 2009). For open populations, Skalski et al. (2009) review marking methods for small fish and analysis methods to estimate survival in both batch and uniquely tagged individuals. However, models for open populations and the use of standard likelihood methods are difficult because of the need to account for potential, non-identifiable, multiple detections of the same individuals over the different sampling occasions.

An important advance was made by Huggins et al. (2010), hereafter denoted HWK, who provide methods for an extended batch mark study. Conditional on the number of individuals released at each sample time, they developed a pseudo likelihood for the recaptured individuals. They then derived estimating equations to estimate survival and capture probabilities, with error estimation obtained from a sandwich estimator. They show how population size can be derived using a Horvitz–Thompson-like estimator also accompanied by a large-sample sandwich variance estimator. HWK do not model individuals that were captured but not marked; however, they did include these when estimating population size. During a week-long industrial modeling camp, Dang et al. (2009) derived a likelihood for unmarked individuals, but ignored the dependencies between the numbers of individuals captured at successive times. Cowen et al. (2014) developed a likelihood approach for the marked data, and compared the efficiency of the extended batch-mark study, analyzed using both likelihood and pseudo-likelihood, with a traditional capture–recapture study (using the Jolly–Seber model). Although the analysis only involved marked individuals, direct computation of the likelihood was a formidable computational exercise.

In this article, we present hidden Markov models (HMMs) for the extended batch-mark survey incorporating both marked and unmarked individuals captured and released at each sampling occasion. We illustrate our methods using data on the oriental weather-loach, *Misgurnus anguillicaudatu*, studied by HWK.

The article is structured as follows: the data motivating the work are described and presented in Section 2; Section 3 lists the notation used, including model specification, and explains the relevance of HMMs; Section 4 presents the two likelihood components, corresponding to marked and unmarked individuals. It is explained how the component likelihoods are efficiently computed using the methodology of HMMs. A framework for model selection and evaluation is provided in Section 5 and alternative procedures for estimating population size are presented in Section 6. The results are given in Section 7 and the article ends with discussion and avenues for future research in Section 8.

2. Sampling Design and Data

Data from marked individuals can be summarized in a similar way to the m-arrays used in band-recovery experiments (Brownie et al., 1985), but individuals can be captured on multiple occasions, and significantly the use of multinomial distributions is no longer appropriate.

The weather-loach study is described in detail by HWK. Here, different colored batch tags were given to a random sample of unmarked individuals at each occasion. The data are provided in Table 1 for this 11-sample occasion study. We do not use the information on the numbers lost (ℓ_t), but it is presented to illuminate the data as a whole. The analysis of

Table 1

Weather-loach batch mark data array taken from HWK. The number unmarked is the number of individuals caught without a mark, before marking takes place. Thus, trivially this number is 306 at the first sampling occasion. Of this number, 280 individuals are given a batch mark, and 32 of them are recaptured at sample occasion 2, 22 at occasion 3, etc. In order to understand the table structure, note, for example, that at sample occasion 4, $207 = 23 + 17 + 20 + 147$*, etc. To illustrate the notation,* $T = 11, G = 10, S_2 = 219, R_2 = 139, u_2 = 187, r_{23} = 28, r_{24} = 17$*, etc. At each sample occasion a number of the fish sampled are not marked, and they are not returned to the water, becoming lost to the study. Thus, for example, at recapture occasion 2,* $48 = 187 - 139$*, etc.*

Sample occasion (g)	Number sampled(S_g)	Number marked (R_g)	Recapture occasion									
			2	3	4	5	6	7	8	9	10	11
1	306	280	32	22	23	8	3	1	2	1	1	0
2	219	139		28	17	3	2	0	2	1	2	0
3	189	115			20	6	3	1	3	2	1	0
4	207	126				5	0	3	2	2	0	1
5	111	80					12	3	1	1	2	1
6	96	65						2	4	8	5	2
7	30	14							2	1	1	0
8	68	50								4	5	1
9	83	54									9	1
10	81	55										6
11	50	0										
	Number unmarked (u_t)		187	139	147	89	76	20	52	63	55	38
	Number lost (ℓ_t)		48	24	21	9	11	6	2	9	0	38

marked individuals is conditional upon the numbers of marked individuals released, and the analysis of unmarked individuals is based upon the numbers of unmarked individuals sampled, and so the lost individuals do not bias the analyses. Note that the number of sampling occasions ($T = 11$) and the number of individuals marked at the first occasion ($R_1 = 280$) are sufficiently large for direct computation of the likelihood, as in Cowen et al. (2014), to be computationally prohibitive, as discussed below.

3. Notation and Model Development

3.1. *Models and Assumptions*

We allow for immigration and emigration/death between sample times, but all emigration is assumed permanent. Other assumptions are similar to those of typical capture–recapture experiments namely: all individuals have the same probability of survival between sample times, all individuals have the same probability of capture at a sample time, tags are not lost, the color of the tags is identifiable, sampling is instantaneous, and individuals are independent with respect to capture and survival.

As ages of individuals are unknown in the case study, we only consider possible time-dependence in model parameters. We specify models using a standard notation so that the four models for marked individuals are (ϕ, p), (ϕ_t, p), (ϕ, p_t), and (ϕ_t, p_t), where ϕ and p denote, respectively, survival and recapture probabilities, which may be constant or time dependent. The (ϕ_t, p_t) model is the batch-marking version of the Cormack–Jolly–Seber model, see Buckland and Morgan (2016) and McCrea and Morgan (2014, p. 70). There are also no covariates in the case study, but if relevant time-varying covariates are available then these might be accommodated by appropriate logistic regressions.

Hidden Markov models (HMMs) are a particular type of state-space model where the state space is discrete. Mark-recapture clearly fits under the HMM framework; see King (2012). Here, a capture–recapture observation is generated by a distribution that is dependent on the state of an unobserved Markov process (Zucchini et al., 2016). For typical capture–recapture models there are two hidden states for each animal, and the true sequences of states are often only partially observed; when an animal is captured then it is known to be alive, whereas when it is not then its true state may be unknown. This is explained in detail in Laake (2013). A more complex example is provided by Chapter 24 of Zucchini et al. (2016), in which there are three hidden states at each time, according to whether individuals are alive, have died since the previous sampling time, or have died prior to that. In that application observed capture histories are then described in terms of survival, recapture, and recovery probabilities. For the extended batch-marking experiment, the hidden states are at the batch, rather than the individual level, and the true sequence of states is entirely unobserved. This gives rise to a large number of states.

3.2. *Primary Notation*

The following notation is used in specifying the likelihoods:

Constants and Statistics

- T the number of capture–recapture occasions.
- G the number of batch-marked release groups; $G \leq T$.
- S_g the number of individuals sampled at sampling occasion g; $g = 1, 2, \ldots, G$.
- R_g the number of individuals marked and released at sampling occasion g from batch group g; $g = 1, 2, \ldots, G$. We condition on the $\{R_g\}$ when we form the likelihood for marked individuals.
- r_{gt} the number of individuals from batch group g recaptured at recapture occasion t; $g = 1, 2, \ldots, G$, $t = g + 1, \ldots, T$.
- u_t the number of individuals captured at sampling occasion t that were not marked; $t = 1, \ldots, T$.
- ℓ_t the number of individuals lost at sampling occasion t; $t = 2, \ldots, T$. $\ell_t = u_t - R_t$.

Latent Variables

- X_{gt} the number of individuals present at occasion t from marked group g; $g = 1, 2, \ldots, G$, $t = g + 1, \ldots, T$.
- d_{gt} the number of individuals from marked group g that die at sampling occasion t.

Parameters

- $\boldsymbol{\Gamma}_{gt}$ the $(R_g + 1) \times (R_g + 1)$ state transition probability matrix of the Markov chain for the marked individuals of group g, describing transitions between sample occasions t and $t + 1$.
- $\boldsymbol{\Gamma}^u_t$ the $(U_{max} + 1) \times (U_{max} + 1)$ state transition probability matrix of the Markov chain for the unmarked individuals describing transitions between sample occasions t and $t + 1$.
- $\boldsymbol{P}_{gt}(m)$ the $(R_g + 1) \times (R_g + 1)$ diagonal matrix containing the state-dependent probabilities of observing m recaptures at occasion t for group g. Thus, $\boldsymbol{P}_{gt}(m) = \text{diag}(q_0(m), \ldots, q_{R_g}(m))$, where $q_i(m) = P(r_{gt} = m | X_{gt} = i)$.
- $\boldsymbol{P}^u_t(m)$ the $(U_{max} + 1) \times (U_{max} + 1)$ diagonal matrix containing the state-dependent probabilities of observing m unmarked individuals at occasion t. Thus, $\boldsymbol{P}^u_t(m) = \text{diag}(q^u_0(m), \ldots, q^u_{U_{max}}(m))$, where $q^u_i(m) = P(u_t = m | U_t = i)$.
- $\boldsymbol{\delta}_g$ the initial distribution of the Markov chain for marked group g.
- $\boldsymbol{\delta}^u$ the initial distribution of the Markov chain for the unmarked population.
- ϕ_t the probability of surviving and remaining in the population between occasions t and $t + 1$, given an individual was alive and in the population at occasion t. We use $\boldsymbol{\phi}$ to indicate the set of survival parameters in a model.
- p_t the probability of capture at occasion t. We use $\boldsymbol{p}$ to indicate the full set of recapture probabilities in a model, and $\boldsymbol{p}_{2:T}$ when p_1 is omitted.
- λ the initial mean abundance (at occasion 1) for the unmarked population.

η the recruitment rate into the unmarked population. Note that we shall consider constant and density-dependent versions of recruitment.

U_t the total number of unmarked individuals in the population available for capture at occasion t. We may use $\boldsymbol{U}$ to denote the full set of the numbers of unmarked individuals in the population at times $t = 1, \ldots, T$. In the Student model of Section 4.2.1, these numbers are parameters which are estimated directly from the likelihood, whereas in the open N-mixture model the $\boldsymbol{U}$ are obtained as derived variables, which are functions of the other model parameters.

U_{max} the maximum number of unmarked individuals available for capture on any occasion.

N_t the total population size at time t.

4. Likelihood Constructions

As we can see from Table 1, in examples of extended batch-mark studies, not all individuals that are captured are marked. This may occur for reasons such as handling time constraints where animals must be released after being contained for a set period of time, or practical constraints, such as having a limited number of marks available. We shall provide the likelihood for marked individuals for extended batch-marking experiments. We shall also incorporate information on unmarked individuals into the likelihood by developing separate models for such individuals. This was not done by HWK or Cowen et al. (2014); it will allow us to include additional information on the capture probabilities in the analysis and also to estimate simultaneously the numbers of unmarked individuals in the population.

4.1. *Likelihood for Marked Individuals*

HWK regarded the likelihood for marked individuals as "intractable," and suggested that a possible EM approach would be "complicated and computationally intensive." Instead, they present the pseudo likelihood shown below,

$$\tilde{\mathrm{L}}_m(\boldsymbol{\phi}, \boldsymbol{p}_{2:T}; \{r_{gt}\}\{R_g\}) \propto \prod_{g=1}^{G} \prod_{t=g+1}^{T} Q_{gt}(\boldsymbol{\phi}, \boldsymbol{p})^{r_{gt}} \{1 - Q_{gt}(\boldsymbol{\phi}, \boldsymbol{p})\}^{R_g - r_{gt}},$$

where $Q_{gt}(\boldsymbol{\phi}, \boldsymbol{p}) = p_t \left(\prod_{i=g}^{t-1} \phi_i\right)$ denotes the probability that an individual released from group g is recaptured at occasion t. In fact, we can obtain an expression for the likelihood for the data from marked individuals by conditioning upon the unknown numbers of dead individuals, $\{d_{gt}\}$, to obtain

$$\begin{aligned}\mathrm{L}_m(\boldsymbol{\phi}, \boldsymbol{p}_{2:T}; \{r_{gt}\}\{R_g\}) = \prod_{g=1}^{G} \sum_{d_{gg}} \cdots \sum_{d_{g,T-1}} \prod_{t=g+1}^{T} \\ P(r_{gt}|d_{gg}, \ldots, d_{g,T-1}, R_g) \\ \times P(d_{gg}, \ldots, d_{g,T-1}|R_g). \end{aligned} \quad (1)$$

Evaluating the conditional probabilities in equation (1) then results in the explicit expression for the likelihood,

$$\begin{aligned}\mathrm{L}_m(\boldsymbol{\phi}, \boldsymbol{p}_{2:T}; \{r_{gt}\}\{R_g\}) = \prod_{g=1}^{G} \sum_{d_{gg}} \cdots \sum_{d_{g,T-1}} \left\{ \prod_{t=g+1}^{T} \binom{R_g - \sum_{m=g}^{t-1} d_{gm}}{r_{gt}} \right. \\ \left. \times \left(p_t \prod_{m=g}^{t} \phi_m\right)^{r_{gt}} \left(1 - p_t \prod_{m=g}^{t} \phi_m\right)^{R_g - \sum_{m=g}^{t-1} d_{gm} - r_{gt}} \right\} \\ \times \frac{R_g!}{d_{gg}! \ldots d_{g,T-1}! (R_g - \sum_{m=g}^{T-1} d_{gm})!} \pi_{gg}^{d_{gg}} \\ \ldots \pi_{g,T-1}^{d_{g,T-1}} (1 - \pi_{gg} - \ldots - \pi_{g,T-1})^{(R_g - \Sigma_{m=g}^{T-1} d_{gm})}, \end{aligned} \quad (2)$$

where $\pi_{gk} = (1 - \phi_k) \prod_{j=1}^{k-1} \phi_j$, and by convention, $\prod_{j=1}^{0} = 1$.

Direct computation of the likelihood $\mathrm{L_m}$ can be slow, due to the evaluation of the multiple summations involved, which is $O((R_1 + 1)^T)$, equating to $O(281^{11})$ for the case study. Consequently, we have only been able to maximize the likelihood in this form for the weather-loach data from the first 7 samples only. However, the results provide a useful check of the HMM computations which follow.

The model for each release group is a HMM, with the $\{d_{gt}\}$ being the hidden information, described by the conditional multinomial distributions in equation (1). It is the fact that the $\{d_{gt}\}$ are unknown/hidden that results in the expensive summations in equation (2). We can therefore make use of the efficiency of the standard forward probability approach for HMMs to compute the likelihood, in addition to other benefits; see Zucchini et al. (2016, p. 37). Therefore, the HMM likelihood component for release group g of the marked individuals, $\mathrm{L}_{m,g}$, can be written in the usual way as a product of the initial distribution vector $\boldsymbol{\delta}_g$, the appropriate transition probability matrices $\{\boldsymbol{\Gamma}_{gt}\}$, and the state-dependent probability matrices $\{\boldsymbol{P}_{gt}\}$ for each sample occasion t. Thus, from Zucchini et al. (2016, p. 37) we can write,

$$\mathrm{L}_{m,g} = \boldsymbol{\delta}_g \boldsymbol{P}_{g,g+1}(r_{g,g+1}) \boldsymbol{\Gamma}_{g,g+1} \boldsymbol{P}_{g,g+2}(r_{g,g+2}) \boldsymbol{\Gamma}_{g,g+2} \ldots \boldsymbol{\Gamma}_{g,T-1} \boldsymbol{P}_{gT}(r_{gT}) \mathbf{1}',$$

where $\mathbf{1}$ denotes the unit row vector,

$$\boldsymbol{P}_{gt}(r_{gt}) = \begin{bmatrix} P(r_{gt}|X_{gt} = 0) & & & \\ & P(r_{gt}|X_{gt} = 1) & & 0 \\ & & \ddots & \\ 0 & & & P(r_{gt}|X_{gt} = R_g) \end{bmatrix},$$

with $P(r_{gt}|X_{gt} = i) = 0$ for $i < r_{gt}$, and otherwise we have the binomial forms: $P(r_{gt}|X_{gt} = i) = \binom{i}{r_{gt}} p_t^{r_{gt}} (1 - p_t)^{i - r_{gt}}$, for $i \geq r_{gt}$.

For each sample time t, the state transition probability matrix $\mathbf{\Gamma}_{gt}$ has elements $P(X_{g,t+1} = j|X_{gt} = i)$, and has the form,

$$\mathbf{\Gamma}_{gt} = \begin{array}{c} \\ i=0 \\ 1 \\ 2 \\ \vdots \\ R_g \end{array} \begin{array}{c} \begin{array}{cccc} j=0 & 1 & 2 & R_g \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ (1-\phi_t) & \phi_t & 0 & \dots & 0 \\ (1-\phi_t)^2 & 2\phi_t(1-\phi_t) & \phi_t^2 & \dots & 0 \\ \vdots & \vdots & & & \\ (1-\phi_t)^{R_g} & R_g\phi_t(1-\phi_t)^{R_g-1} & \binom{R_g}{2}\phi_t^2(1-\phi_t)^{R_g-2} & \dots & \phi_t^{R_g} \end{bmatrix}_{(R_g+1)\times(R_g+1)} \end{array}.$$

Each row of $\mathbf{\Gamma}_{gt}$ corresponds to a binomial distribution, and the rows sum to unity, as required. As the number of animals alive at $t = 0$ for group g is known to be R_g for each g, the initial state distribution for group g is given by

$$\begin{aligned}\boldsymbol{\delta}_g &= [P(X_{g0} = 0),\ P(X_{g0} = 1), P(X_{g0} = 2), \dots, P(X_{g0} = R_g)] \\ &= (0, 0, \dots, 0, 1).\end{aligned}$$

If we assume independence between release groups, the likelihood for the marked individuals is then given as

$$\mathrm{L}_m(\boldsymbol{\phi}, \boldsymbol{p}_{2:T}; \{r_{gt}\}, \{R_g\}) = \prod_{g=1}^{G} \mathrm{L}_{m,g}. \tag{3}$$

This is exactly the same expression as equation (2), but in a different formulation for efficient computation.

4.2. *Likelihood for Unmarked Individuals*

4.2.1. Student approach. Dang et al. (2009) employed a product-binomial likelihood,

$$\mathrm{L}_u(\boldsymbol{p}, \boldsymbol{U}; \{u_t\}) = \prod_{t=1}^{T} \binom{U_t}{u_t} (p_t)^{u_t}(1-p_t)^{U_t-u_t}. \tag{4}$$

This likelihood has the same structural form as the first part of the Jolly–Seber likelihood, (McCrea and Morgan, 2014, p. 150), and we shall refer to the underlying model as the Student model. The model includes $\boldsymbol{U}$, the elements of which denote the numbers of unmarked individuals in the population at the different times, as a set of parameters to be estimated. By itself of course this likelihood is over parameterized, as there are two unknowns for each degree of freedom. There are various ways of dealing with this, and a referee has suggested using Bayesian modeling with a joint regularization prior on the $\{U_t\}$ with a positive correlation structure to constrain the $\{U_t\}$, for example, a random walk of order 2 or a Gaussian Markov random field; see Schmidt et al. (2015). Here, we combine the likelihood with the likelihood for the marked individuals. The likelihoods for the marked and unmarked individuals can be multiplied to get the overall joint likelihood, as there are no individuals in common:

$$\begin{aligned}\mathrm{L}_{\mathrm{j}}(\boldsymbol{\phi}, \boldsymbol{p}, \boldsymbol{U}; \{r_{gt}\}, \{R_g\}, \{u_t\}) &= \mathrm{L}_m(\boldsymbol{\phi}, \boldsymbol{p}_{2:T}; \{r_{gt}\}, \{R_g\}) \\ &\quad \times \mathrm{L}_u(\boldsymbol{p}, \boldsymbol{U}; \{u_t\}).\end{aligned} \tag{5}$$

This is an illustration of integrated population modeling; see Besbeas et al. (2002). It is possible to maximize $\mathrm{L}_{\mathrm{j}}(\boldsymbol{\phi}, \boldsymbol{p}, \boldsymbol{U}; \{r_{gt}\}, \{R_g\}, \{u_t\})$, after some minor adjustments described later. However, if the $\{U_t\}$ have no structure such a model is still too general, and in particular there is no involvement of $\boldsymbol{\phi}$ in $\mathrm{L}_u(\boldsymbol{p}, \boldsymbol{U}; \{u_t\})$. An alternative structural approach is described below.

4.2.2. Open N-mixture approach. Here, we introduce structure, by relating the $\{U_t\}$ over time in a stochastic manner, using an open N-mixture model for the unmarked individuals. This model adopts a first-order Markov dependence; see Dail and Madsen (2011). The likelihood for the unmarked individuals is then given by

$$\begin{aligned}\mathrm{L}_u(\boldsymbol{\phi}, \boldsymbol{p}, \lambda, \eta; \{u_t\}) = &\sum_{U_1=u_1}^{\infty} \cdots \sum_{U_T=u_T}^{\infty} \prod_{t=1}^{T} \mathrm{Bin}(u_t; U_t, p_t) \\ &\times P(U_1) \prod_{t=2}^{T} P(U_t|U_{t-1}),\end{aligned} \tag{6}$$

where $P(U_1)$ is the probability function for U_1 and $P(U_t|U_{t-1})$ is the probability of U_t conditional upon the value of U_{t-1}. The total number of unmarked individuals in the population at time t is unknown, and following Dail and Madsen (2011), after experimentation we assume this to be less than some large number, U_{max}, for all t. Thus, U_{max} can be taken as the upper limit of all of the summations in equation (6). For related discussion, see Dennis et al. (2015). One might similarly introduce an appropriate U_{min} term, which could be beneficial in some cases. We can see how the binomial expression of the Student model, in equation (4), is the basis for the binomial term in equation (6), although the interpretation of $\{U_t\}$ is different between the two models. Once again we have a HMM, due to the Markov structure, and the computationally expensive summations in the likelihood expression can be seen to be a consequence of the unknown numbers of uncaptured individuals at each occasion; cf equation (1).

In order to obtain $P(U_t|U_{t-1})$, we write $U_t = A_t + C_t$ where A_t is the number of individuals that have survived

from occasion $t-1$ to t, and C_t is the number of individuals recruited into the population at occasion t. We model $A_t|U_{t-1} \sim \text{Binomial}(U_{t-1} - S_{t-1}, \phi_{t-1})$ and $C_t|U_{t-1} \sim \text{Poisson}(U_{t-1}\eta)$, although other distributions are possible, for example to account for overdispersion. The conditional distribution of C_t provides a form of density dependence of recruitment, and an alternative is simply to use Poisson(η); we have used both in the case study of the next section. The initial number of unmarked individuals in the population is modeled as $U_1 \sim \text{Poisson}(\lambda)$, and $P(U_t|U_{t-1})$ is given by the convolution sum,

$$P(U_t = j|U_{t-1} = i) = \sum_{c=0}^{min(i-S_{t-1},\, j)} \text{Bin}(c; i - S_{t-1}, \phi_{t-1}) \times \text{Pois}(j - c; U_{t-1}\eta), \text{ for } i \geq S_{t-1}, \quad (7)$$

and zero otherwise, where $\text{Bin}(x; n, p)$ and $\text{Pois}(x; \mu)$ denote the probability functions of the binomial and Poisson distributions, respectively. In view of the typically very large number of terms in the summations of equation (7) the Fast Fourier Transform was used to evaluate the summations efficiently.

As the open N-mixture model is also a HMM, we can write the likelihood of equation (6) in the form,

$$\text{L}_u(\boldsymbol{\phi}, \boldsymbol{p}, \lambda, \eta; \{u_t\}) = \boldsymbol{\delta^u P}^u_1(u_1)\boldsymbol{\Gamma}^u_1\boldsymbol{P}^u_2(u_2)\boldsymbol{\Gamma}^u_2 \ldots \boldsymbol{\Gamma}^u_{T-1}\boldsymbol{P}^u_T(u_T)\mathbf{1}' \quad (8)$$

where

$$\boldsymbol{P}^u_t(u_t) = \begin{bmatrix} P(u_t|U_t = 0) & & & \\ & P(u_t|U_t = 1) & & 0 \\ & & \ddots & \\ 0 & & & P(u_t|U_t = U_{max}) \end{bmatrix}$$

and $P(u_t|U_t = i) = 0$ for $i < u_t$, and otherwise we have the binomial forms: $P(u_t|U_t = i) = \binom{i}{u_t} p_t^{u_t}(1 - p_t)^{i-u_t}$, for $i \geq u_t$. This is clearly the same approach as that adopted in Section 4.1.

The state transition probability matrix for the unmarked population $\boldsymbol{\Gamma}^u_t$, describing both the survival and immigration of individuals, has elements given by equation (7), and the initial state distribution, $\boldsymbol{\delta^u}$, is Poisson(λ), as stated above.

The joint likelihood in this case has the form,

$$\text{L}_\text{j}(\boldsymbol{\phi}, \boldsymbol{p}, \lambda, \eta; \{r_{gt}\}, \{R_g\}, \{u_t\}) = \text{L}_m(\boldsymbol{\phi}, \boldsymbol{p}_{2:T}; \{r_{gt}\}, \{R_g\}) \times \text{L}_u(\boldsymbol{\phi}, \boldsymbol{p}, \lambda, \eta; \{u_t\}). \quad (9)$$

We thank a referee for pointing out similar work in Schmidt et al. (2015). In that article data are collected on known-fate individuals which are radio marked, and also unmarked individuals. Integrated modeling is employed: an open N-mixture model is used to model the information on unmarked individuals, and that likelihood component is also formed using a HMM.

5. Estimation, Model Selection, and Goodness-of-Fit

Hidden Markov modeling is challenging in the work of this article in view of the very large numbers of states that might be involved. We discuss this issue in Section 7.2. However, the computer programs written were effective and efficient. Parameters are estimated using numerical maximum likelihood methods, and we use the observed information matrix and the bootstrap to provide estimates of uncertainty. Confounded parameters (e.g., the product $\phi_{T-1}p_T$) can be dealt with by introducing single parameters denoting such a product. To obtain estimated standard errors in general, we need to be able to deal with the Hessian being singular due to boundary estimates (e.g., $\hat{\phi}_7 = 1$). We obtain the singular-value decomposition (SVD) of the Hessian, $\boldsymbol{H}$, which in standard notation can be written as $\boldsymbol{H} = \boldsymbol{WDV}^T$; see Searle (1982, p. 316). We then determine which values of the diagonal matrix $\boldsymbol{D}$ are close to zero, and remove the corresponding rows and columns of the SVD, resulting in $\boldsymbol{H}^*$. We then obtain standard errors from the inverse of $\boldsymbol{H}^*$ using $\boldsymbol{H}^{*-1} = \boldsymbol{V}^*\boldsymbol{D}^{*-1}\boldsymbol{W}^{*T}$ (Searle 1982, p. 318), and set standard errors equal to zero for the estimates corresponding to the deleted rows.

Model selection can be achieved using standard likelihood methods, see Besbeas et al. (2015), and we illustrate the model selection process for the case study using the Akaike information criterion (AIC) in Section 7. The only model fitted by HWK, to data on marked individuals, was that where survival probability varies by time and capture probability is constant, (ϕ_t, p), which they justify in terms of how the sampling was carried out.

To study goodness-of-fit for marked individuals, observed and expected cell counts of the marked individuals can be compared straightforwardly, where expected cell counts are calculated as $\mathbb{E}(r_{gt}) = R_g Q_{gt}(\boldsymbol{\phi}, \boldsymbol{p})$, as done by HWK, who showed that for the weather-loach data the Pearson goodness-of-fit statistic for model (ϕ_t, p) has the value $X^2 = 21.62$, which is insignificant (p-value $> 10\%$) when referred to χ^2 tables on 16 degrees of freedom, after appropriate pooling of cells. For the model for unmarked individuals, we plot normal pseudo-residual segments. These arise as we are dealing with discrete random variables, so that the standard pseudo residuals, which are points defined for continuous random variables, are replaced by intervals (Zucchini et al., 2016, p. 101). Ordinary normal pseudo residuals (Zucchini et al., 2016, p. 102) could also be used for the model for marked individuals, but cohort lengths are typically short, and so instead we plot observed versus expected values in that case.

6. Estimating Population Size

Population size was estimated by HWK using a Horvitz–Thompson-like approach, resulting in

$$\hat{N}_t = \frac{S_t}{\hat{p}}, \quad (10)$$

where $\hat{p}$ was estimated from the data on marked individuals. However, when we also model the number of unmarked

individuals, we can form alternative estimates of population size. This is because we can estimate the numbers of both marked and unmarked individuals in different ways. For the open N-mixture model the estimated expected numbers of unmarked individuals are derived variables, functions of the estimated model parameters, and given by recursion; see Dail and Madsen (2011, p. 4). For instance, for the case of constant survival and density-dependent recruitment, we have

$$\hat{\mathbb{E}}(U_t) = \hat{\lambda}(\hat{\phi} + \hat{\eta})^{t-1}. \tag{11}$$

We shall refer back to this equation later in the article. We could use the multivariate delta method to estimate the variances of the estimated expectations in equation (11), but in the following we shall instead use the bootstrap. An alternative approach for estimating the numbers of unmarked individuals arises because of the HMM nature of the model for unmarked individuals, and results from application of the Viterbi algorithm, a dynamic programming algorithm which produces the most likely set of $\{\hat{U}_t\}$; see Zucchini et al. (2016, p. 89). In the following, we shall refer to $\{\hat{U}_t\}$, but when the Dail/Madsen approach is used, rather than the Viterbi algorithm, this is to be interpreted as $\{\hat{\mathbb{E}}(U_t)\}$.

In addition, whichever estimate of unmarked individuals is used, by adopting a Horvitz–Thompson-like approach, the number of individuals alive at occasion t can be obtained from,

$$\hat{N}_1 = \hat{U}_1,$$

$$\hat{N}_t = \hat{U}_t + \frac{\sum_{g=1}^{t-1} r_{gt}}{\hat{p}_t}, \quad \text{for} \quad t > 1. \tag{12}$$

Cf equation (10). Alternatively, we can estimate the numbers of marked individuals alive directly from the numbers of individuals marked in appropriate samples:

$$\tilde{N}_1 = \hat{U}_1$$

$$\tilde{N}_t = \hat{U}_t + \sum_{i=1}^{t-1} R_i \prod_{\ell=i}^{t-1} \hat{\phi}_\ell, \quad \text{for} \quad t > 1. \tag{13}$$

We shall consider the relative merits of alternative methods for estimating population sizes in the analyses of the next section. However, we note here that our experience has been that the expression of equation (13) results in smaller standard errors than those of equation (12), and we only illustrate the former in the case study which follows. A further approach, suggested by a referee, would result from posterior sampling of numbers of marked and unmarked individuals following a Bayesian analysis, such as that outlined in Schmidt et al. (2015).

7. Case Study

We use the data of Table 1 to illustrate the methods of the article.

7.1. *Analysis of Marked Data*

To compare the HMM likelihood method with the pseudo-likelihood (PL) method of HWK we present in Table 2 parameter estimates for the model (ϕ_t, p) for data on marked individuals only, the sole case for which results were presented by HWK. We found that parameter estimates were in good agreement between the two methods for this case study; however, standard errors for the likelihood method were usually smaller, as independently suggested by the simulation study of Cowen et al. (2014). We note that our estimates of error for PL differ substantially from what HWK reported, partly due to errors both in their R code for obtaining standard error estimates and in the delta method that HWK used, and partly due to the likelihood method being more efficient. Note also that the sandwich standard errors are harder to compute than those resulting from inverting a Hessian in the normal way. This is due to the need to construct manually a model-dependent derivative matrix. Additionally, although most of the HMM standard errors are smaller than those from the PL approach, we have found the difference to be far greater, with HMM resulting in smaller errors, when smaller data sets are analyzed, as found, for example, in Haynes and Robinson (2011).

Several models were fitted using the HMM and their AIC values were compared (Table 3a). Interestingly, we found model (ϕ_t, p) to be the best fitting model, the only model fitted by HWK, and the model (ϕ, p_t) performs well. Note that model (ϕ_t, p_t) is parameter redundant, as only the product $\phi_{T-1}p_T$ can be estimated, rather than either of the component terms.

Table 2

Parameter estimates and estimated standard errors (given below in parentheses) produced by the pseudo-likelihood method (PL) and the likelihood method (HMM) for model (ϕ_t, p) for the marked weather-loach data. In the PL case we also give, in square parentheses, the incorrect standard errors from HWK.

	Parameter estimates										
Method	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	p
PL	0.59	0.84	0.92	0.27	0.53	0.48	1.00	0.75	0.79	0.36	0.18
	(0.09)	(0.12)	(0.14)	(0.07)	(0.13)	(0.10)	(0.00)	(0.16)	(0.17)	(0.12)	(0.02)
	[0.18]	[0.34]	[0.51]	[0.15]	[0.26]	[0.21]	[0.05]	[0.36]	[0.41]	[0.25]	[0.06]
HMM	0.59	0.86	0.90	0.26	0.56	0.47	1.00	0.75	0.81	0.35	0.18
	(0.09)	(0.12)	(0.13)	(0.05)	(0.11)	(0.09)	(0.00)	(0.15)	(0.18)	(0.11)	(0.02)

Table 3
Model-selection results for the weather-loach data:(a) for marked data only. (b) for marked and unmarked data combined, using first the Student approach and then the open N-mixture approach (with $U_{max} = 1800$ and $w = 20$). Here, κ denotes the number of estimable model parameters.

(a)

Model	$-\log(L_m)$	κ	AIC	ΔAIC
(ϕ_t, p_t)	91.03	19	220.03	6.5
(ϕ_t, p)	95.78	11	213.55	0.0
(ϕ, p_t)	97.29	11	216.58	3.0
(ϕ, p)	124.98	2	253.97	40.4

(b)
Student approach

Model	$-\log(L_j)$	κ	AIC	ΔAIC
(ϕ, p, U_t)	158.11	13	342.22	46.62
(ϕ, p_t, U_t)	126.83	22	297.66	2.06
(ϕ_t, p, U_t)	129.01	22	302.03	6.43
(ϕ_t, p_t, U_t)	117.85	30	295.60	0.00

Open N-mixture approach; density-dependent model

Model	$-\log(L_j)$	κ	AIC	ΔAIC
(ϕ, p, λ, η)	202.90	4	413.80	110.92
$(\phi, p_t, \lambda, \eta)$	140.18	14	308.36	5.48
$(\phi_t, p, \lambda, \eta)$	150.56	13	327.12	24.24
$(\phi_t, p_t, \lambda, \eta)$	128.44	23	302.88	0.00

7.2. *Combining Marked and Unmarked Data*

A complication with the analysis of the unmarked data is that a large value of U_{max} was required. We experimented with several values and found $U_{max} = 1800$ to work well for these data. However, this requires operations with matrices of dimension 1801×1801 for the computation of L_u, which is memory (RAM) intensive for a personal computer. To deal with this, we allocate the states to bins of equal size w: we let $\zeta_1, \zeta_2, \ldots, \zeta_n$ be a partition of the state space such that each ζ_i is of length w; the midpoint of the interval is taken to represent the state; see Zucchini et al. (2016, p. 158). For this case study, we experimented with values of $w = 4$, 10, 20, and 50, and found that taking $w = 20$ appeared to be satisfactory, though there was little difference between taking $w = 4$ and $w = 20$. This resulted in 90 hidden states, whereas, for example, taking $w = 4$ increased the number of states to 450. Binning in this way introduces an element of approximation to the formation of the likelihood component for unmarked individuals, the approximation increasing with w. For illustration, we present results for the density-dependent model for new individuals, as in equation (7), as these resulted in slightly better AIC values than the alternative, of constant augmentation. However, which alternative is more realistic might also depend on which agrees better with the biology of the study.

When the unmarked individuals are incorporated into the likelihood, model selection may be done in the same way as before, using AIC; see Table 3b. Here, we find that model (ϕ_t, p_t) fits the data best in both the Student and N-mixture approaches. However, uncritical use of AIC can result in overly complex models, and the models (ϕ, p_t, U_t) and $(\phi, p_t, \lambda, \eta)$ perform well in terms of AIC. Therefore, we shall use these for ease of illustration. We see in particular that models with constant recapture probability are not supported, in contrast to the findings of Table 3a, and the belief of HWK that this would be an acceptable assumption for the weather-loach data, due to the use of constant-effort electrofishing. Table 4 presents parameter estimates and estimated standard errors for the models (ϕ, p_t, U_t) and $(\phi, p_t, \lambda, \eta)$. We note that the average of the estimates of recapture probability from the N-mixture model is 0.19, in comparison with the value of 0.18 in Table 2, and conversely, from Table 2 the average estimate of ϕ is given by 0.65, in comparison with the value of 0.63 from Table 4. Standard errors are presented from using an estimated Hessian in the usual way, and are very similar to those obtained from a parametric bootstrap approach (not shown). The estimates of $\boldsymbol{U}$ from the Viterbi analysis of the N-mixture model are rounded due to the binning of states. They are in agreement with the predictions from equation (11), as they should be. As observed earlier, the Student estimates of $\boldsymbol{U}$ differ in interpretation compared with those obtained from the N-mixture model. The differences are generally small in comparison with the standard errors. The Student models generally are hard to fit, being subject to confounding, boundary estimates and convergence to local optima. For example, it is not possible to estimate p_1 and U_1 separately, nor ϕ_{10} and p_{11} in the (ϕ_t, p_t, U_t) model, which can only be estimated as a product, as observed also by HWK. However, they are estimable separately for model $(\phi_t, p_t, \lambda, \eta)$. We note in Table 4 the much smaller standard errors for the components of $\{\boldsymbol{U}\}$ from using the N-mixture approach, compared with the Student model, as expected, due to the chaining of the components of $\boldsymbol{U}$ in the open N-mixture model; standard errors are formed by using a parametric bootstrap, based on simulating HMMs. We can see that what is driving the need for the recapture probabilities to change with time is the estimate for the 7th sample occasion. It is interesting to note from Table 4 that the Dail/Madsen estimates of error are slightly larger than those from the Viterbi algorithm. In addition the Viterbi approach is more general. The very good agreement between the Viterbi and Dail/Madsen estimates, obtained by quite different methods, provides a validation for the methods used.

We show in Table 5 how we can estimate $\{N_t\}$, in this case making use of the Viterbi estimates of the numbers of unmarked individuals for illustration. We see that the standard errors for N_t and the population size estimates using "HT with p-vals," resulting from using the Horvitz–Thompson-like approach of equation (10) with estimated time-varying recapture probabilities, are similar in size, and generally different from those resulting from HWK. A remarkable feature of Table 5 is the very close agreements of both the estimates of N_t and their corresponding estimated standard errors, whether one forms the estimates by evaluating the marked and unmarked components of N_t separately, as in this article, or simply uses a HT approach. We note also the curious feature that the estimated standard errors are close to the sums

Table 4

Combining both marked and unmarked data: parameter estimates and estimated standard errors (below in parentheses) for the weather-loach data, produced by the Student approach, for model (ϕ, p_t, U_t), and the N-mixture approach for model $(\phi, p_t, \lambda, \eta)$ (with $N_{max} = 1800$ and $w = 20$): V indicates Viterbi and D indicates Dail/Madsen; † indicates that p_1 was fixed to 1. Standard errors marked with a $*$ are obtained using a Hessian, while without result from 100 bootstraps.

	ϕ	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	λ	η
Student	0.62	†	0.18	0.25	0.31	0.11	0.12	0.07	0.16	0.21	0.28	0.13		
$(SE)^*$	(0.03)		(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.04)	(0.05)	(0.06)	(0.04)		
N-mixture	0.63	0.32	0.22	0.21	0.25	0.15	0.15	0.05	0.15	0.20	0.22	0.15	944.8	0.24
$(SE)^*$	(0.03)	(0.05)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.03)	(0.04)	(0.05)	(0.04)	(133.0)	(0.03)

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}
Student	306	1012	545	477	797	652	297	332	297	194	289
$(SE)^*$		(182)	(89)	(76)	(197)	(173)	(118)	(102)	(83)	(49)	(99)
N-mixture (V)	950	830	710	610	530	470	410	350	310	270	230
(SE)	(143)	(108)	(82)	(68)	(61)	(58)	(57)	(57)	(56)	(56)	(55)
N-mixture (D)	945	821	713	619	538	467	406	352	306	266	231
(SE)	(143)	(109)	(85)	(71)	(63)	(59)	(58)	(57)	(57)	(56)	(55)

Table 5

Components of estimated population sizes. The Viterbi values are taken from Table 4. The estimated marked values are the result of using the last term of the expression of equation (13). HWK gives the estimates of $\{N_t\}$ from HWK, using equation (10), while HT with p-vals is the same structure, but when there are time-varying probabilities of detection, $\{p_t\}$, estimated from marked and unmarked data. The estimates of $\{N_t\}$ are obtained from equation (13), using the values of $\hat{U}_t$ obtained from the Viterbi algorithm. All standard errors are estimated in this table from the (parametric) bootstrap.

Sample occasion (t)	1	2	3	4	5	6	7	8	9	10	11
$\hat{U}_t$ (Viterbi)	950	830	710	610	530	470	410	350	310	270	230
(SE)	(143)	(108)	(82)	(68)	(61)	(58)	(57)	(57)	(56)	(56)	(55)
Estimated marked		177	200	199	206	181	155	107	99	97	96
(SE)		(7)	(12)	(15)	(17)	(18)	(17)	(15)	(13)	(12)	(11)
$\hat{N}_t$	950	1007	910	809	736	651	565	457	409	367	326
(SE)	(143)	(110)	(88)	(77)	(73)	(72)	(70)	(67)	(65)	(64)	(63)
$\hat{N}_t$:HT with p-vals	945	997	914	821	745	650	565	463	408	365	330
(SE)	(143)	(111)	(91)	(80)	(75)	(73)	(71)	(69)	(67)	(65)	(63)
$\hat{N}_t$:HWK	1689	1209	1043	1143	613	530	166	375	458	447	276
(SE)	(233)	(171)	(150)	(163)	(94)	(84)	(35)	(63)	(74)	(73)	(50)

of the estimated standard errors for the marked/unmarked components. Consequently, the corresponding variances are all slightly greater than what would be obtained by treating the components of the sum as independent, suggesting that the components are slightly positively correlated. A further interesting aspect of Table 5 is the difference in using equation (10) depending on whether the recapture probability is time varying or not. The big difference occurs at the 7th sampling occasion. HWK comment that, "The estimated population size of the weather-loach was low on the 7th occasion, which happened to be in the winter period. However, as spring came, the numbers of weather-loach increased. These estimates indicate that the population is quite seasonal and can be low at some times of the year." However, when the capture probabilities are allowed to vary with time then the estimates of population size suggest a declining population over time.

Figure 1 illustrates goodness-of-fit plots for model $(\phi, p_t, \lambda, \eta)$, with $w = 20$ and density dependence, separately for each likelihood component, as recommended by Besbeas and Morgan (2014). Overall there does not appear to be a serious lack of fit for the models fitted. The patterns in Figure 1a are a simple consequence of the repeated values of observed counts.

We note finally that the expected number of new unmarked individuals entering the population at occasion t can be derived using

$$\hat{B}_t = \hat{U}_{t+1} - (\hat{U}_t \hat{\phi}_t), \tag{14}$$

as in the Jolly–Seber model; see McCrea and Morgan (2014, p. 150). We would expect $\hat{B}_t \approx \mathbb{E}[C_t | U_{t-1}]$, where $\hat{B}_t$ is estimated in equation (14), and this relationship is found to hold.

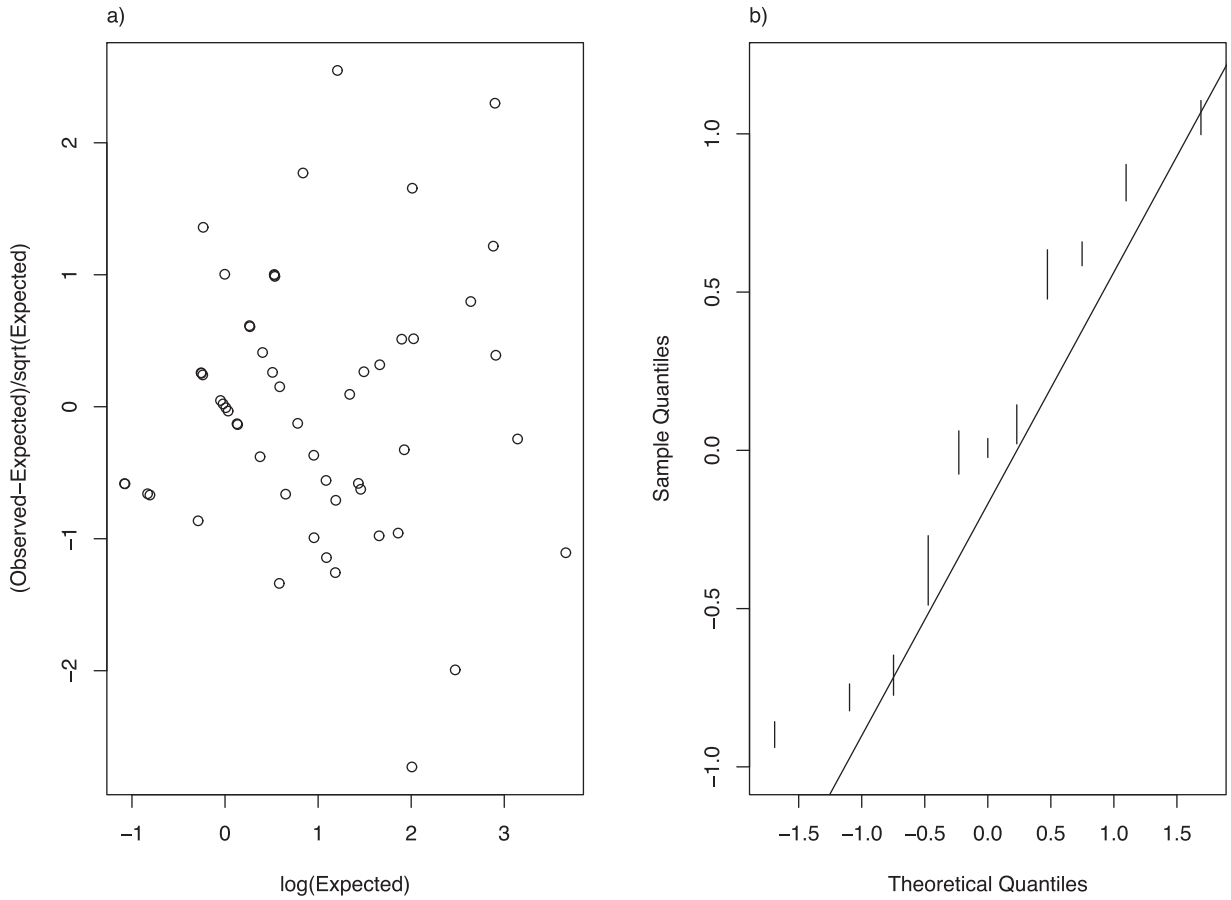

Figure 1. Goodness-of-fit plots for the weather-loach data, for model (ϕ, p_t, λ, η): (a) we plot the difference, observed minus expected recapture counts for the marked individuals, divided by the square root of the expected counts, against the logarithm of the expected counts, and (b) the QQ plot for the normal pseudo-residual segments of the unmarked counts.

8. Discussion

Batch marking is fundamentally important in ecology, providing an important tool for studying population dynamics. However, there is a pressing need for effective new methods of analysis. We have developed a comprehensive model for the extended batch-marking experiment, using an efficient hidden Markov formulation that supersedes previous analyses. The approach incorporates information on unmarked individuals, leading to increased precision of parameter estimates, and improved model-selection. It has been interesting to see, in the case study, that the selected model changes as a result of the joint analysis. In addition one can devise and compare different models for the unmarked individuals, and useful descriptions of the population result. For instance, we have examined whether constraining the population sizes for the unmarked individuals in the Student model to be constant over time results in a realistic model for the weather-loach data, and it was not found to be competitive. An added advantage of using HMMs is the ability to examine pseudo residuals to check goodness-of-fit. In addition the Viterbi algorithm provides a convenient, general method for estimating numbers of unmarked individuals, which can be bootstrapped to provide estimates of standard errors. As the model is fitted using standard likelihood optimization methods, it is possible to easily undertake model selection and compute estimates of parameter uncertainty.

The models presented provide a flexible platform for additional development, depending on what data are available. For instance, one can incorporate suitable covariates, which would reduce the number of parameters to be estimated and potentially increase biological understanding. Extensions to cope with age and size information, as discussed by HWK, would also be straightforward. Further, we know from discussions with batch marking users that tag loss can be non-negligible, though apparently this was not an issue with the weather-loach study. For some study designs, it might be possible to implement double marks on individuals and implement methods similar to Cowen and Schwarz (2006). Alternatively, one might be able to obtain auxiliary information on tag-loss rates through a holding study and adjust parameter estimates using methods of Seber and Felton (1981).

9. Supplementary Materials

MATLAB code for the analyses of the article is available with this article at the *Biometrics* website on Wiley Online Library. It is written for use in the Parallel Computing toolbox, but is readily translated into `R` if necessary.

Acknowledgements

This work was initiated while LC was on study leave at the University of Kent supported by both a NSERC Discovery grant and a University of Victoria Professional Development grant. We thank David Borchers, Ruth King, and Roland Langrock for discussing HMM methods. Comments by the two referees improved the article. This work was partly funded by EPSRC/NERC grant EP/1000917/1.

References

Besbeas, P., Freeman, S., Morgan, B. J. T., and Catchpole, E. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58**, 540–547.

Besbeas, P., McCrea, R. S., and Morgan, B. J. T. (2015). Integrated population model selection in ecology. *Kent Academic Repository* https://kar.kent.ac.uk/id/eprint/48039

Besbeas, P. and Morgan, B. J. T. (2014). Goodness of fit of integrated population models using calibrated simulation. *Methods in Ecology and Evolution* 1373–1382.

Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. S. (1985). Statistical inference from band recovery data: A handbook. Technical report, U.S. Fish and Wildlife Service Resource Publication, Washington, D.C.

Buckland, S. T. and Morgan, B. J. T. (2016). 50-year anniversary of papers by Cormack, Jolly and Seber. *Statistical Science* **31**, 141.

Cowen, L. and Schwarz, C. J. (2006). The Jolly-Seber model with tag loss. *Biometrics* **62**, 699–705.

Cowen, L. L. E., Besbeas, P., Morgan, B. J. T., and Schwarz, C. J. (2014). A comparison of abundance estimates from extended batch-marking and Jolly-Seber-type experiments. *Ecology and Evolution* **4**, 210–218.

Dail, D. and Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics* **67**, 577–587.

Dang, H., Huang, Y., Robertson, R., Diaz, E., de la Rosa, D., Viguernas, F., Tifenbach, R., Kashyap, H., and Cowen, L. (2009). Mark-recapture with batch marks but no remarking. In *Report on PIMS 2008 GIMMC and IPSW*, B. Alspach and S. Fallat (eds), 9–20. University of Regina. https://www.pims.math.ca/files/ipsw12.pdf

Dennis, E. B., Morgan, B. J. T., and Ridout, M. S. (2015). Computational aspects of N-mixture models. *Biometrics* **71**, 237–246.

Haynes, T. B. and Robinson, C. L. K. (2011). Re-use of shallow sediment patches by pacific sand lance (ammodytes hexapterus) in Barkley Sound, British Columbia, Canada. *Environmental Biology of Fishes* **92**, 1–12.

Huggins, R., Wang, Y., and Kearns, J. (2010). Analysis of an extended batch marking experiment using estimating equations. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 279–289.

King, R. (2012). A review of Bayesian state-space modelling of capture-recapture data. *Interface Focus* **2**, 190–204.

Laake, J. L. (2013). Capture-recapture analysis with hidden Markov models. AFSC Processed Report 2013-04, 34p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

McCrea, R. S. and Morgan, B. J. T. (2014). *Analysis of Capture-recapture Data*. Boca Raton: CRC Press, Chapman & Hall.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**, 3–135.

Pine, W. E., Pollock, K. H., Hightower, J. E., Kwak, T. J., and Rice, J. A. (2003). A review of tagging methods for estimating fish population size and components of mortality. *Fisheries Research* **28**, 10–23.

Rawson, K. (2009). Review of marking methods and release-recapture designs for estimating the survival of very small fish: Examples from the assessment of salmonid fry survival. *Reviews in Fisheries Science* **17**, 391–401.

Schmidt, J. H., Johnson, D. S., Lindberg, M. S., and Adams, L. G. (2015). Estimating demographic parameters using a combination of known-fate and open N-mixture models. *Ecology* **96**, 2583–2589.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York: Wiley.

Seber, G. A. F. and Felton, R. (1981). Tag loss and the Petersen mark-recapture experiment. *Biometrika* **68**, 211–219.

Skalski, J. R., Buchanan, R. A., and Griswold, J. (2009). Review of marking methods and release-recapture designs for estimating the survival of very small fish: Examples from the assessment of salmonid fry survival. *Reviews in Fisheries Science* **17**, 391–401.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series, An Introduction Using R, Second Edition*. Chapman & Hall/CRC.

Received July 2016. *Revised February* 2017.
Accepted March 2017.