The Annals of Applied Statistics
2019, Vol. 13, No. 4, 2043–2064
https://doi.org/10.1214/19-AOAS1264

ESTIMATING ABUNDANCE FROM MULTIPLE SAMPLING CAPTURE-RECAPTURE DATA VIA A MULTI-STATE MULTI-PERIOD STOPOVER MODEL

BY HANNAH WORTHINGTON[*,1], RACHEL MCCREA[†,2], RUTH KING[‡] AND RICHARD GRIFFITHS[†]

University of St Andrews[*], University of Kent[†] and University of Edinburgh[‡]

Capture-recapture studies often involve collecting data on numerous capture occasions over a relatively short period of time. For many study species this process is repeated, for example, annually, resulting in capture information spanning multiple sampling periods. To account for the different temporal scales, the robust design class of models have traditionally been applied providing a framework in which to analyse all of the available capture data in a single likelihood expression. However, these models typically require strong constraints, either the assumption of closure within a sampling period (the closed robust design) or conditioning on the number of individuals captured within a sampling period (the open robust design). For real datasets these assumptions may not be appropriate. We develop a general modelling structure that requires neither assumption by explicitly modelling the movement of individuals into the population both within and between the sampling periods, which in turn permits the estimation of abundance within a single consistent framework. The flexibility of the novel model structure is further demonstrated by including the computationally challenging case of multi-state data where there is individual time-varying discrete covariate information. We derive an efficient likelihood expression for the new multi-state multi-period stopover model using the hidden Markov model framework. We demonstrate the significant improvement in parameter estimation using our new modelling approach in terms of both the multi-period and multi-state components through both a simulation study and a real dataset relating to the protected species of great crested newts, *Triturus cristatus*.

1. Introduction. In this paper we develop a model capable of analysing capture-recapture data from multiple sampling periods within a single likelihood expression. In comparison to existing models we retain the ability to estimate total population size through the likelihood. Hidden Markov model (HMM) methods are used to estimate state- and time-dependent abundance. Following the structure

Received August 2018; revised April 2019.

[1]Supported by a Carnegie Scholarship awarded to Hannah Worthington by the Carnegie Trust for the Universities of Scotland.

[2]Supported by a Natural Environment Research Council research fellowship grant NE/J018473/1.

Key words and phrases. Hidden Markov model, great crested newts, multi-state data, individual time-varying discrete covariate.

of existing models used to analyse multiple periods of capture-recapture data we allow parameters to be dependent both on time and time spent in the population.

Standard capture-recapture studies consist of several capture occasions where attempts are made to capture individuals from the population of interest. When an individual is captured for the first time it is marked, or unique physical marks recorded, to permit unique identification of each individual. At subsequent capture occasions it is then possible, using these unique marks, to identify new individuals (which are subsequently marked) or recaptured individuals (those that have been previously captured). In this paper, we assume that all sampled individuals are returned to the population after capture, that is, that there are no removals upon capture from the population. By repeating this process at each capture occasion it is possible to identify on which occasions each unique individual was recorded. This information is stored in the form of individual capture histories. Typically these capture histories are of binary form, for example,

$$0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0$$

where 0, and 1, indicate an individual was not captured, or captured, at each capture occasion, respectively. During some capture-recapture studies it may be possible to collect additional individual covariate information. We consider the case where an individual time-varying discrete covariate is recorded corresponding to the state of the individual upon capture. This additional information is recorded in the capture history where nonzero entries now indicate the observed state. This discrete state information may refer, for example, to behavioural states such as breeding or foraging, or alternatively it may refer to a discrete location such as which pond in a study site.

The Schwarz–Arnason (SA) model (Schwarz and Arnason (1996)) estimates the size of the super-population which includes both individuals that are captured at least once as well as those that are never captured (but are available for capture on at least one occasion). The inclusion of the super-population in the SA model, denoted N, allows for births to be modelled within the likelihood expression. The stopover model presented by Pledger et al. (2009) is an extension of the SA model in which the capture and retention probabilities are dependent both on time and time since arrival. In the stopover model the term "age" is used to refer to the time since joining the population (not necessarily physical age) and is generally unknown due to the unknown arrival time (an individual may have joined the population on an occasion before their first capture). We note that when collecting data to which we wish to fit a stopover model it is advisable to sample both before the first arrivals and after the final departures (this results in capture histories with leading and trailing zeros). Whilst the easiest approach is to analyse only the subset of data corresponding to when the site is occupied, the extended sampling before and after occupation can verify the implicit assumptions that those present on the first occasion a capture occurs have only recently arrived and those present on the

final occasion where captures occur are imminently about to depart. This second assumption can be partially relaxed and sampling need not necessarily continue until all individuals have departed; if this is the case the probability an individual is present at the site for the whole period the site is occupied (e.g., present for the entire breeding season) must be negligible (Kendall et al. (2018)).

Multi-state capture-recapture models allow the inclusion of individual time-varying discrete covariates. Dupuis and Schwarz (2007) consider a multi-state extension of the SA model for estimating abundance in open populations fitted within a Bayesian (data augmentation) framework. This model allows for time- and state-dependence in the capture probabilities but not the age-dependence of the stopover model. Typically these models assume a first-order Markov model for the transition of individuals among the different discrete states. King and Langrock (2016) relax this assumption through a semi-Markov model where the dwell-time distribution (the time spent in the state) has some parametric form. Brownie et al. (1993) use a second-order Markov approach and suggest analysis be based on the conditional likelihood.

The SA and stopover models consider only a single group of capture occasions. However, for many studies, capture occasions are spaced closely in time, for instance during the breeding season, and the sampling process is repeated many times, for example, every year. The robust design class of models consider the data at these two sampling levels; primary and secondary sampling periods. In general the robust design models assume that the capture-recapture data of the secondary periods are collected over a relatively short period of time, while the duration between the primary sampling periods is much larger. The closed robust design model (Pollock (1982), Kendall, Pollock and Brownie (1995)), uses closed-population capture-recapture models to estimate abundance within each primary occasion. The more generally applicable (since animal populations are rarely closed) open robust design model (Schwarz and Stobo (1997), Kendall and Bjorkland (2001)) retains the open primary occasions, but also permits the secondary occasions to be open to arrivals and departures. Kendall et al. (2018) extend this class of models to include a multi-state open robust design (MSORD) model which includes state information on the primary level and includes the estimation of abundance for each primary period.

We develop a general multi-state multi-period stopover model and formulate the explicit likelihood expression as an HMM. Similar to the MSORD model, this model permits time dependence in the survival (probability of remaining in the population on the primary level) and capture probabilities and dependence on time and time since arrival (within each sampling period) in the retention probabilities (probability of remaining at the site between capture occasions in the secondary level). In contrast to the MSORD model, our model also includes dependence on time since recruitment (time since arrival in the primary level) in the survival probabilities, time since arrival dependence in the capture probabilities, and permits

transitions between states within the primary periods. The model retains the flexibility of movement into and out of the population, assuming an open population both between and within each sampling period. In particular, by allowing survival to depend on the time since recruitment, transient individuals who spend very little time in the population may be accounted for differently to those more permanent (longer staying) individuals. We apply a similar argument to the stopover and SA models, assuming a total population across all the periods consisting of both those individuals that are observed and those that are not observed but available for capture. This approach allows the size of the total population to be estimated as a parameter in the model. Subsequently, algorithms for HMMs can be implemented to estimate the size of the population in each primary period along with state-dependent abundance on occasions within each primary period. Doing so allows the investigation of trends over time and comparisons of population size in the different states. This is demonstrated in the simulation study and newt application below. The multi-state aspect of the model is included so that additional information can be incorporated such as different mark types, location information or breeding status. We focus in particular on allowing the capture probabilities to be state-dependent thus allowing for heterogeneity in the population. This new multi-state multi-period stopover model can be considered a general model for capture-recapture data from which many of the existing models can be obtained by placing appropriate restrictions on the model parameters.

The motivation for developing this new multi-period stopover model is a long-term study on great crested newts, a protected species in Europe. Although up to £43 million is spent on mitigating the impacts of development on this species in England alone (Lewis, Griffiths and Wilkinson (2017)), current population assessment protocols for this species are inadequate (Griffiths et al. (2015)). There is consequently a need for more reliable statistical models that take account of the seasonal dynamics of this species. The study population considered here is unique in that it is based on replicated ponds that have been intensively monitored for nearly two decades. Individuals in this population visit the study site for the breeding season. Arrivals and departures occur over a number of weeks with the traps being set once per week during the breeding season with the process repeated annually. The additional state information for this population is the pond in which each individual newt is captured. Geographically, the area of the field study site has remained consistent. Originally the site contained four constructed ponds, a further four ponds were constructed in 2009 and first colonised in the 2010 breeding season. Given that pond creation is regarded as a fundamental component of amphibian conservation, of particular biological interest is how these new ponds have been colonised, whether capture probabilities differ between the old well-established ponds and the new ponds and the trap effectiveness at capturing the newts. The old and new ponds may exhibit differences due to differing amounts of vegetation, with these differences perhaps disappearing as the new ponds become established. The ponds are located in close proximity to one another and so newts

can freely move between them throughout each breeding season. For this population of newts there is particular interest in the total number of newts visiting the site and the number of newts using, and moving between, the old and new ponds.

We express the multi-state multi-period stopover model using an HMM. HMMs provide a flexible way of modelling series of observations collected through time that depend on underlying and often unobserved correlated states. After the initial capture and marking of an individual, the capture history can be considered as a combination of two processes: the observation process which depends on the availability of an individual for capture; and an underlying availability process. The HMM separates the underlying state process (i.e., availability for capture) from the observation process (i.e., capture process, conditional on availability). For further discussion see, for example, Gimenez et al. (2007), King (2012, 2014), King et al. (2009), Langrock and King (2013), Royle (2008), Schofield and Barker (2008), Zucchini, MacDonald and Langrock (2016).

In Section 2 we derive the multi-state multi-period stopover model. In Section 3 we perform a simulation study before applying the new model to a data set on great crested newts in Section 4. We conclude with a discussion in Section 5.

2. Model derivation. In this section we derive the multi-state multi-period stopover model. Within each primary period we assign an individual with time since arrival value 1 on the first occasion they attend the site to indicate that they have spent one capture occasion in the population. On the primary level we assign an individual with time since recruitment value 1 on the first primary period they join the population. We use the general terms "arrival" and "recruitment" to indicate an individual becoming available for capture in the different levels of the model. These will in practice have different interpretations and could refer to births, recruitment to the breeding population, or arrival at a specific colony for migratory species. Likewise departures may refer to different ways of leaving a site, including deaths or permanent emigration from the study area. In this derivation we incorporate the state-dependence in the capture probabilities. We allow for movement between the states within the primary periods to be first-order Markov with initial choice of state in the next primary period independent of the states used in the current primary period; generalisation of the initial choice of state is discussed further in Section 5. We also assume that the state of an individual is recorded without misclassification when an individual is observed, though this assumption can be relaxed (King and McCrea (2014)) in a multi-state capture-recapture setting.

2.1. *Notation*. In defining the notation of the multi-state multi-period stopover model we extend, where possible, the notation of Pledger et al. (2009). Let N denote the total population (to be estimated) consisting of all individuals who visit the study site for at least one capture occasion during the study period (all capture occasions and periods). Further, let n denote the number of observed individuals (those captured on at least one capture occasion) and n_m the number of

individuals that are missed (those that are never captured). Thus $N = n + n_m$. Let the entire study period consist of T primary periods, labelled $t = 1, \ldots, T$, with $K(t)$ secondary capture occasions in primary period t. We let the capture history for individual i be denoted by $\boldsymbol{x}_i = \{x_i(t,k) : k = 1, \ldots, K(t); t = 1, \ldots, T\}$ and let the set of capture histories for all observed individuals be denoted by $\boldsymbol{x} = \{\boldsymbol{x}_i : i = 1, \ldots, n\}$. Note that from the histories we can easily extract in which primary periods each observed individual was captured at least once.

We now define the set of model parameters (in addition to N above). We define the recruitment probabilities to be the set $\boldsymbol{r} = \{r(t) : t = 1, \ldots, T\}$ where $r(t)$ is the probability of being recruited into the population and first becoming available for capture in primary period t. Since an individual belonging to the total population must visit the site during at least one primary period, $\sum_{t=1}^{T} r(t) = 1$. For the HMM formulation of the model we define $r^*(t) = r(t)/\sum_{j=t}^{T} r(j)$ for $t = 2, \ldots, T$ which denotes the conditional recruitment probability (probability of being recruited in primary period t given the individual has not been recruited in any primary periods $1, \ldots, t-1$). We define the set of arrival probabilities to be $\boldsymbol{\beta} = \{\beta(t,k) : k = 1, \ldots, K(t); t = 1, \ldots, T\}$ where $\beta(t,k)$ is the probability of arriving at the study site and first being available for capture from occasion k within primary period t, given the individual is in the population and available for capture in primary period t. By definition, within each primary period $t = 1, \ldots, T$, $\sum_{k=1}^{K(t)} \beta(t,k) = 1$. Similarly to the recruitment probabilities, the HMM formulation requires conditional arrival probabilities which we define as $\beta^*(t,k) = \beta(t,k)/\sum_{j=k}^{K(t)} \beta(t,j)$ for $k = 2, \ldots, K(t)$ and $t = 1, \ldots, T$ (probability of arriving on occasion k in primary period t given the individual has not arrived on occasions $1, \ldots, k-1$ in primary period t). We let $\boldsymbol{s} = \{s_A(t) : A = 1, \ldots, t; t = 1, \ldots, T-1\}$ denote the set of survival probabilities, where $s_A(t)$ is the probability an individual is available for capture in primary period $t+1$ given in primary period t they are available for capture and have been present in the population for A primary periods. We let $\boldsymbol{\phi} = \{\phi_a(t,k) : a = 1, \ldots, k; k = 1, \ldots, K(t)-1; t = 1, \ldots, T\}$ denote the set of retention probabilities where $\phi_a(t,k)$ is the probability that an individual is available for capture on occasion $k+1$ in primary period t given on occasion k in primary period t the individual is available for capture and has been present in the population for a secondary capture occasions within primary period t.

In order to model the movement of individuals between the different observable discrete states we first need to consider the discrete state that an individual enters when they first arrive at the site within each primary period. In this derivation we assume initial choice of state within a primary is independent of previous state occupation, see Section 5 for a discussion of alternative structures. We denote these initial discrete state probabilities by $\boldsymbol{\alpha} = \{\alpha_g(t) : t = 1, \ldots, T; g = 1, \ldots, G\}$ where $\alpha_g(t)$ is the probability of being in state $g = 1, \ldots, G$ (where G is the total number of observable states) on the first occasion an individual is available for

capture in primary period t. In this derivation we assume these initial discrete state probabilities are constant over time and so regardless of when an individual arrives at the site the probability they enter each of the observable states remains the same.

The set of transition probability matrices between the discrete states is denoted by $\boldsymbol{\Psi} = \{\boldsymbol{\Psi}(t) : t = 1, \ldots, T\}$. The transition probabilities in primary period t are given by,

$$\boldsymbol{\Psi}(t) = \begin{pmatrix} \psi_{11}(t) & \psi_{12}(t) & \ldots & \psi_{1G}(t) \\ \psi_{21}(t) & \psi_{22}(t) & \ldots & \psi_{2G}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{G1}(t) & \psi_{G2}(t) & \ldots & \psi_{GG}(t) \end{pmatrix}$$

such that $\psi_{ij}(t)$ denotes the probability of moving from state i to state j between consecutive secondary occasions in primary period t, conditional on the individual remaining available for capture in primary period t. For simplicity we have defined the transition probabilities to be constant across all occasions within a primary period. In general, this need not be the case, however, there are likely to be issues with parameter redundancy and identifiability in the fully time-dependent case. Finally, we define the capture probabilities to be $\boldsymbol{p} = \{p_{ga}(t, k) : a = 1, \ldots, k; k = 1, \ldots, K(t); t = 1, \ldots, T; g = 1, \ldots, G\}$ where $p_{ga}(t, k)$ is the probability an individual is captured given they are in state g and have been present at the site for a secondary occasions on occasion k in primary period t. We include a diagram in the Supplementary Material (Worthington et al. (2019), Appendix A) of the ordering of events to aid the understanding of the model parameters. The full set of model parameters for the multi-state multi-period stopover model is given by $\boldsymbol{\theta} = \{N, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{\alpha}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{p}\}$.

The choice of model structure developed here is motivated by the underlying ecology and biological understanding of the behaviour of great crested newts. For further discussion on how the model could be further generalised see Section 5.

2.2. *HMM formulation.* Following the convention of the robust design models we consider nested (or hierarchical) Markov chains, the first operating on the primary level and the second nested chain operating on the secondary capture occasions. Let $\boldsymbol{h} = \{h(t) : t = 1, \ldots, T\}$ be the hidden states in the primary level where

$$h(t) = \begin{cases} 1 & \text{not yet recruited into the attending population;} \\ 2 & \text{time since recruitment} = 1 \text{ (recruited in this period);} \\ \vdots & \vdots \\ A' + 1 & \text{time since recruitment } = A' \text{ } (A' \text{ periods spent in population);} \\ A' + 2 & \text{departed from the attending population;} \end{cases}$$

where A' is the maximum observable time since recruitment of individuals in the population on the primary level ($A' \leq T$) implying individuals move to the absorbing state of having departed from the population (a combination of death and permanant migration) after spending A' primary occasions in the population. Similarly, let $\boldsymbol{h}(t) = \{h(t,k) : k = 1, \ldots, K(t); t = 1, \ldots, T\}$ be the hidden states in the secondary level where

$$h(t,k) = \begin{cases} 1 & \text{not yet available for capture;} \\ 2 & \text{available for capture in primary occasion } t, \\ & \quad \text{time since arrival} = 1 \text{ and in state 1;} \\ 3 & \text{available for capture in primary occasion } t, \\ & \quad \text{time since arrival} = 1 \text{ and in state 2;} \\ \vdots & \vdots \\ G+1 & \text{available for capture in primary occasion } t, \\ & \quad \text{time since arrival} = 1 \text{ and in state } G; \\ G+2 & \text{available for capture in primary occasion } t, \\ & \quad \text{time since arrival} = 2 \text{ and in state 1;} \\ \vdots & \vdots \\ 2G+1 & \text{available for capture in primary occasion } t, \\ & \quad \text{time since arrival} = 2 \text{ and in state } G; \\ \vdots & \vdots \\ a'(t)G+1 & \text{available for capture in primary occasion } t, \\ & \quad \text{time since arrival} = a'(t) \text{ and in state } G; \\ a'(t)G+2 & \text{departed from the site in primary occasion } t; \end{cases}$$

where $a'(t)$ is the maximum observable time since arrival of individuals in the secondary level ($a'(t) \leq K(t)$) again implying that once an individual has spent $a'(t)$ secondary occasions at the site in primary period t, they move to the absorbing state of having departed the site for that period. We note that it is possible that $a'(t)$ and G could be different in each primary period. This would change the size of the matrices used within the secondary level of the model but no other changes are necessary. We also note that the age need not increment deterministically by one on each occasion but could more generally refer to age classes, for example, immature, adult and senior, we discuss this further below.

Let the initial hidden state distribution of the primary level HMM,

$$\boldsymbol{\pi}(1) = \big(\mathbb{P}\big(h(1) = 1\big) \quad \mathbb{P}\big(h(1) = 2\big) \quad \ldots \quad \mathbb{P}\big(h(1) = A' + 2\big),\big)$$

be the probabilities of entering each primary hidden state for primary period 1. Similarly, for the secondary level HMM,

$$\boldsymbol{\pi}(t,1) = \big(\big(\mathbb{P}\big(h(t,1) = 1\big) \quad \mathbb{P}\big(h(t,1) = 2\big) \quad \ldots \quad \mathbb{P}\big(h(t,1) = a'(t)G + 2\big)\big)\big),$$

for $t = 1, \ldots, T$, describes the probabilities of entering each secondary hidden state on occasion 1 in each primary period $t = 1, \ldots, T$. Then, by definition of the model parameters above,

$$\boldsymbol{\pi}(1) = \begin{pmatrix} 1 - r(1) & r(1) & 0 & \ldots & 0 \end{pmatrix},$$

$$\boldsymbol{\pi}(t, 1) = \begin{pmatrix} 1 - \beta(t, 1) & \beta(t, 1)\boldsymbol{\alpha}(t) & 0 & \ldots & 0 \end{pmatrix},$$

where $\boldsymbol{\alpha}(t) = \begin{pmatrix} \alpha_1(t) & \alpha_2(t) & \ldots & \alpha_G(t) \end{pmatrix}$ is the set of initial discrete state probabilities for primary period t. Next, we consider the transition matrices which describe the movement between the states of the Markov chains. In the primary level this concerns the survival between the primary periods whilst in the secondary level it is the retention within the given primary period. Let $\boldsymbol{\Gamma}(t)$ be an $(A' + 2) \times (A' + 2)$ matrix where

$$\boldsymbol{\Gamma}(t)[a, b] = \mathbb{P}\big(h(t + 1) = b | h(t) = a\big)$$

for $t = 1, \ldots, T - 1$, $a = 1, \ldots, A' + 2$ and $b = 1, \ldots, A' + 2$. Similarly, let $\boldsymbol{\Gamma}(t, k)$ be an $(a'(t)G + 2) \times (a'(t)G + 2)$ matrix where

$$\boldsymbol{\Gamma}(t, k)[a, b] = \mathbb{P}\big(h(t, k + 1) = b | h(t, k) = a\big)$$

for $k = 1, \ldots, K(t) - 1$, $t = 1, \ldots, T$, $a = 1, \ldots, a'(t)G + 2$ and $b = 1, \ldots, a'(t)G + 2$. By definition,

$$\boldsymbol{\Gamma}(t) = \begin{pmatrix} 1 - r^*(t+1) & r^*(t+1) & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & s_1(t) & 0 & \ldots & 0 & 1 - s_1(t) \\ 0 & 0 & 0 & s_2(t) & \ldots & 0 & 1 - s_2(t) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & s_{A'-1}(t) & 1 - s_{A'-1}(t) \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{\Gamma}(t, k) = \begin{pmatrix} 1 - \beta^*(t, k+1) & \beta^*(t, k+1)\boldsymbol{\alpha}(t) & 0 & \ldots & 0 & 0 \\ 0 & 0 & \phi_1(t, k)\boldsymbol{\Psi}(t) & \ldots & 0 & \big(1 - \phi_1(t, k)\big)_G \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \phi_{a'(t)-1}(t, k)\boldsymbol{\Psi}(t) & \big(1 - \phi_{a'(t)-1}(t, k)\big)_G \\ 0 & 0 & 0 & \ldots & 0 & \mathbf{1}_G \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix},$$

where $(1 - \phi_a(t, k))_G$ is a column vector of length G with each entry equal to $(1 - \phi_a(t, k))$ and $\mathbf{1}_G$ is a column vector of ones of length G.

Here we assume a fully deterministic progression in time since recruitment (it increases by 1 on each primary period) and time since arrival (it increases by 1 on each secondary occasion). This implies that once an individual has attained a time since recruitment of A' they necessarily move to the absorbing state of having departed from the population (a combination of death, migration, or if $A' = T$ the end of study so the probability of further sighting is 0). If time since recruitment were instead grouped into classes (e.g., first time attenders and then repeat attenders), but progression was still deterministic, then the above transition matrix

would stay of the same form but parameters within a class would be shared (e.g., $s_1(t)$, $s_2(t) = s_3(t) = \cdots$). Alternatively, transitions between classes, for example, young, adult and senior, may be stochastic in nature to include, for example, individual variability in maturation rates. In this case a semi-Markov approach could be used in which a distribution is specified for the time spent in each age-class with the transition probabilities being determined from the chosen distribution.

Finally we consider the observation process which connects the observed data to the hidden states. The primary level relates to the probability of observing the capture histories within each primary period and the secondary level relates to the probability of capture on each occasion. We first consider the secondary level and work with unique capture histories $y = 1, \ldots, Y$ rather than considering each individual in turn ($Y \leq n$). Let $\mathbf{P}(t, k, x_y(t, k))$ be an $(a'(t)G + 2) \times (a'(t)G + 2)$ diagonal matrix for $k = 1, \ldots, K(t)$ and $t = 1, \ldots, T$ where $\mathbf{P}(t, k, a)[b, b] = \mathbb{P}(x_y(t, k) = a | h(t, k) = b)$ for $a = 0, 1, \ldots, G$, $b = 1, \ldots, a'(t)G + 2$ and all off-diagonal entries are zero. Then,

$$
\begin{aligned}
&\mathbf{P}\big(t, k, x_y(t, k)\big) \\
&\quad = \begin{cases}
\operatorname{diag}\big(1, 1 - p_{11}(t, k), 1 - p_{21}(t, k), \ldots, 1 - p_{G1}(t, k), \\
\quad \ldots, 1 - p_{1a'(t)}(t, k), \ldots, 1 - p_{Ga'(t)}(t, k), 1\big) & x_y(t, k) = 0, \\
\operatorname{diag}\big(0, p_{11}(t, k), 0, \ldots, 0, p_{12}(t, k), 0, \\
\quad \ldots, 0, p_{1a'(t)}(t, k), 0, \ldots, 0\big) & x_y(t, k) = 1, \\
\operatorname{diag}\big(0, 0, p_{21}(t, k), 0, \ldots, 0, p_{22}(t, k), 0, \\
\quad \ldots, 0, p_{2a'(t)}(t, k), 0, \ldots, 0\big) & x_y(t, k) = 2, \\
\vdots & \vdots \\
\operatorname{diag}\big(0, \ldots, 0, p_{G1}(t, k), 0, \ldots, 0, p_{G2}(t, k), 0, \\
\quad \ldots, 0, p_{Ga'(t)}(t, k), 0\big) & x_y(t, k) = G.
\end{cases}
\end{aligned}
$$

Let $L_0(t)$ and $L_y(t)$ denote the likelihood contribution for a single-period stopover model (i.e., considering the secondary occasions within one primary occasion only) for an all zero capture history (i.e., an individual that is not captured) and a nonzero capture history in primary period t respectively. Then for each primary period $t = 1, \ldots, T$,

$$
L_0(t) = \boldsymbol{\pi}(t, 1)\mathbf{P}(t, 1, 0)\left(\prod_{k=2}^{K(t)} \boldsymbol{\Gamma}(t, k-1)\mathbf{P}(t, k, 0)\right)\mathbf{1}_{a'(t)G+2},
$$

$$
L_y(t) = \boldsymbol{\pi}(t, 1)\mathbf{P}\big(t, 1, x_y(t, 1)\big)\left(\prod_{k=2}^{K(t)} \boldsymbol{\Gamma}(t, k-1)\mathbf{P}\big(t, k, x_y(t, k)\big)\right)\mathbf{1}_{a'(t)G+2},
$$

where $\mathbf{1}_{a'(t)G+2}$ is a column of ones of length $a'(t)G + 2$ (the number of states in the secondary level of the HMM). We can now consider the observation process in

the primary level. Let $z_y(t)$ indicate whether capture history $y = 1, \ldots, Y$ contains a capture in primary period t. Then $z_y(t) = 0$ if $x_y(t,k) = 0$ for all $k = 1, \ldots, K(t)$ and conversely $z_y(t) = 1$ if $x_y(t,k) \neq 0$ for at least one occasion $k = 1, \ldots, K(t)$. Let $\mathbf{P}(t, z_y(t))$ be an $(A'+2) \times (A'+2)$ diagonal matrix for $t = 1, \ldots, T$ where $\mathbf{P}(t,a)[b,b] = \mathbb{P}(z_y(t) = a | h(t) = b)$ for $a = 0, 1$, $b = 1, \ldots, A'+2$ and all off-diagonal entries are zero. Then,

$$\mathbf{P}\big(t, z_y(t)\big) = \begin{cases} \operatorname{diag}\big(1, L_0(t), \ldots, L_0(t), 1\big) & z_y(t) = 0, \\ \operatorname{diag}\big(0, L_y(t), \ldots, L_y(t), 0\big) & z_y(t) = 1. \end{cases}$$

Let L_0 denote the probability an individual is never captured and L_y the probability of observing the unique (nonzero) capture history y, then the primary level expressions for the HMM are

$$L_0 = \boldsymbol{\pi}(1)\mathbf{P}(1,0)\left(\prod_{t=2}^{T} \boldsymbol{\Gamma}(t-1)\mathbf{P}(t,0)\right)\mathbf{1}_{A'+2},$$

$$L_y = \boldsymbol{\pi}(1)\mathbf{P}\big(1, z_y(1)\big)\left(\prod_{t=2}^{T} \boldsymbol{\Gamma}(t-1)\mathbf{P}\big(t, z_y(t)\big)\right)\mathbf{1}_{A'+2},$$

where $\mathbf{1}_{A'+2}$ is a column of ones of length $A'+2$ (the number of states in the primary level of the HMM).

The expression for the full likelihood is of multi-nomial form where individuals with the same capture history are grouped together. We let n_y denote the frequency of each unique capture history $y = 1, \ldots, Y$ where Y is the total number of unique nonzero capture histories. The likelihood expression is given by:

$$L(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{N!}{(N-n)!\prod_{y=1}^{Y} n_y!} L_0^{N-n} \prod_{y=1}^{Y} L_y^{n_y}.$$

Thus we have an explicit likelihood expression.

3. Simulation study. To demonstrate the ability to estimate the parameters of the multi-state multi-period stopover model we perform a simulation study. To explore the advantages of the new approach we compare the results of fitting a multi-state multi-period stopover model against the results of fitting separate multi-state stopover models independently to each primary period of data.

We consider two different total population sizes, $N = 100$ and $N = 1000$, to determine the effect of population size on the ability to estimate the model parameters. We expect that for small total population sizes (where the number of individuals captured in any one primary period will be relatively small) the multi-period model will perform better than the single-period models by taking strength from sharing parameters across the different primary periods. As the population size

increases we expect the variation of parameter estimates to decrease and the performance of the single-period models to improve. Alternatively, a joint likelihood over the primary periods could be formed which could also share parameters but such an approach would not include the between primary modelling, thus would be unable to estimate survival or recruitment probabilities.

We generate three primary periods of data ($T = 3$) with each primary period having five capture occasions ($K(t) = 5$ for all $t = 1, \ldots, T$). We let the number of individuals joining the population at each primary period follow a multi-nomial distribution with probabilities $r(1) = 0.4$, $r(2) = 0.2$ and $r(3) = 0.4$. The probability of survival between each primary period is assumed to be constant with value $s = 0.7$. We define a logistic regression for the cumulative arrival probabilities within each primary period, $B(t,k) = \sum_{j=1}^{k} \beta(t,j)$, from which the arrival probabilities for each occasion, $\beta(t,k)$, can be obtained. We include a primary-dependent intercept, such that

$$\operatorname{logit}\big(B(t,k)\big) = \eta k + \delta(t),$$

where $\eta = 1$, $\delta(1) = -1$, $\delta(2) = 0$ and $\delta(3) = -2$. The arrival probabilities are then

$$\beta(t,k) = \frac{B(t,k) - B(t,k-1)}{B(t,K(t))},$$

where $B(t,0) = 0$ and division by $B(t,K(t))$ ensures the arrival probabilities sum to one. For the retention probabilities we include time effects and a linear age term but no primary period effects,

$$\operatorname{logit}\big(\phi_a(t,k)\big) = \tau(k) + \gamma(a-1),$$

where $\tau(1) = 2.5$, $\tau(2) = 1.8$, $\tau(3) = 2.1$, $\tau(4) = 1.4$, $\gamma = -1$ and a is the time since arrival within primary t.

For the state-dependent parameters we assume them to be constant across all primary periods. The initial discrete state probabilities are $\alpha_1 = 0.35$ and so $\alpha_2 = 0.65$. The capture probabilities we assume to be dependent on state only with $p_1 = 0.6$ and $p_2 = 0.8$. Finally we let the transition probability matrix between the observable discrete states be

$$\boldsymbol{\Psi} = \begin{pmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{pmatrix}.$$

For each population size we generate 1000 data sets. For each data set we fit the multi-state single-period stopover model to each period of data and then the multi-state multi-period stopover model to the full data set using the `nlm` function in `R` to maximise the likelihoods. The results for population size $N(t)$ (estimated using a forward-backward-type algorithm for the multi-period model; details in the Supplementary Material, Worthington et al. (2019), Appendix B), initial discrete state probabilities $\boldsymbol{\alpha}$, transition probabilities $\boldsymbol{\Psi}$ and the arrival probabilities $\boldsymbol{\beta}$ are

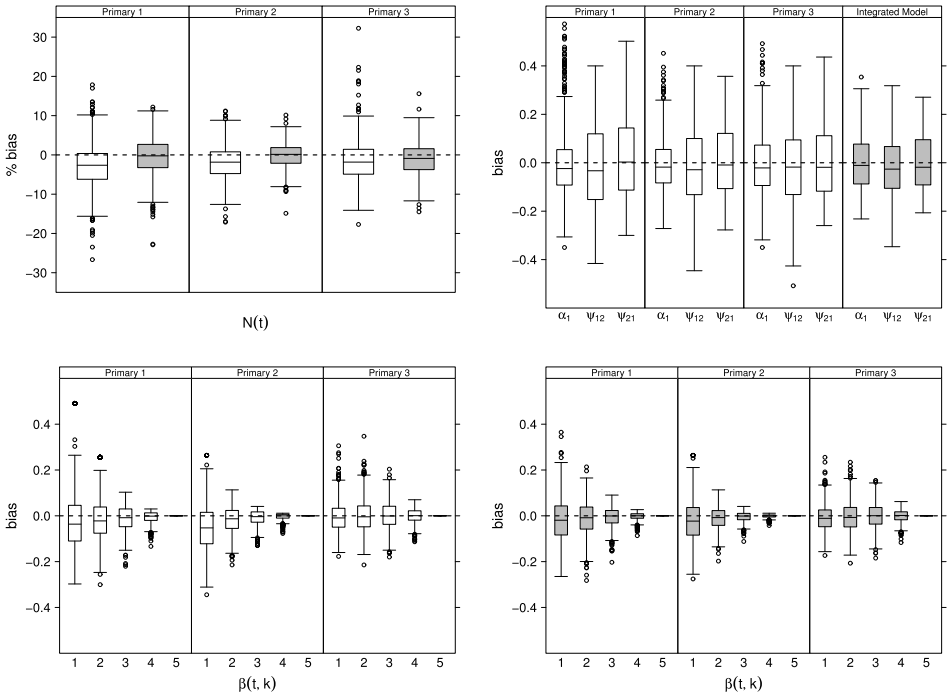

FIG. 1. *Results from the simulation study where* $N = 100$: (*top*, *left*) *percentage bias of the population size estimates in each primary period for the single-period model* (*white*) *and multi-period model* (*grey*); (*top*, *right*) *bias of the initial discrete state and transition probabilities in each primary period for the single-period and multi-period model*; (*bottom*, *left*) *bias of the arrival probabilities for each primary period for the single-period model and*; (*bottom*, *right*) *bias of the arrival probabilities for each primary period for the multi-period model*.

displayed in Figures 1 and 2 for $N = 100$ and $N = 1000$ respectively (output for the remaining parameters and state-dependent abundance within each primary is available in the Supplementary Material, Worthington et al. (2019), Appendix D).

From the simulation study we can clearly see the improved performance through using the multi-period approach. All of the parameters are estimated well and appear to be unbiased (or close to unbiased). We particularly note the improvement in the bias of the estimates for the population size in each primary period and the decrease in variability of the transition and arrival probabilities. Of particular interest is the ability of the models to correctly estimate the state-dependent parameters. For these, when $N = 100$, the multi-period model does provide lower variability in the MLEs than the single-period approach though uncertainty in these parameters is quite large compared to the other parameters in the model. Similar improvements to the precision of transition parameters were found by McCrea et al. (2010) where a multi-state integrated population modelling approach is used to jointly analyse

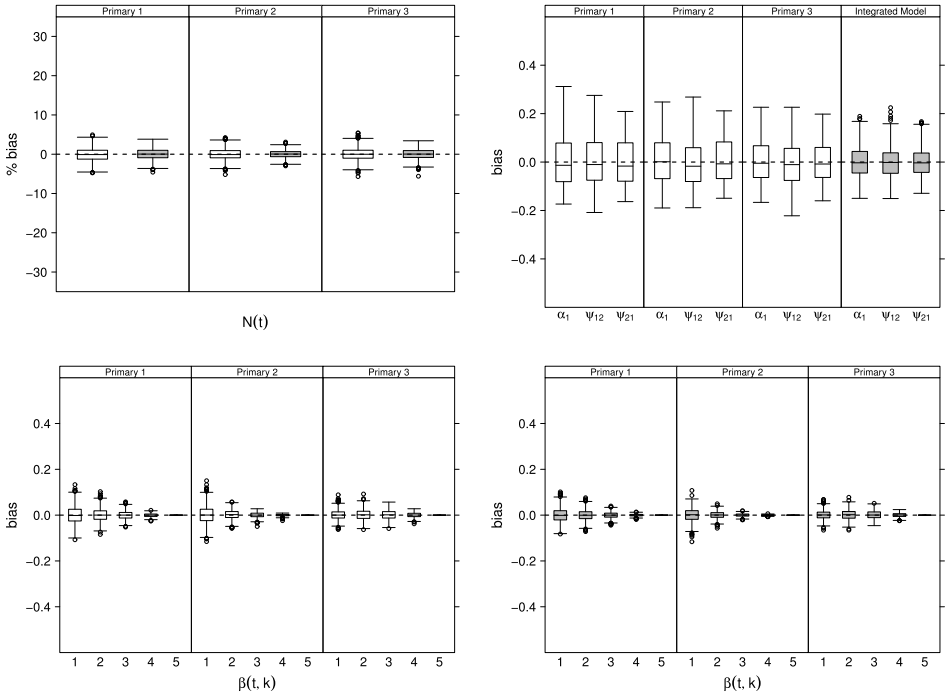

FIG. 2. *Results from the simulation study where* $N = 1000$: (*top*, *left*) *percentage bias of the population size estimates in each primary period for the single-period model* (*white*) *and multi-period model* (*grey*); (*top*, *right*) *bias of the initial discrete state and transition probabilities in each primary period for the single-period and multi-period model*; (*bottom*, *left*) *bias of the arrival probabilities for each primary period for the single-period model and*; (*bottom*, *right*) *bias of the arrival probabilities for each primary period for the multi-period model*.

capture-recapture and census data. We also note the strong improvement in the estimation of the retention probabilities ($\boldsymbol{\phi}$) using the multi-period approach.

When the total population size is increased to $N = 1000$ the variation in estimates for all parameters in both models is reduced. The multi-period model still performs better than the single-period model with the most obvious improvement now in the capture, initial discrete state and transition probabilities. The variation in parameter p_1 in the multi-period model remains greater than the other parameters, this is most likely due to the low capture probability and the probability of remaining in the state being 0.4. This results in only a small number of captures in this state and so larger uncertainty than the equivalent parameter in the other state. However, we do note that the estimates are unbiased and so in general the parameters appear to be estimated well.

4. Application. Data on a population of great crested newts are collected from a field study site on the University of Kent campus. The data have been

collected since 2002 on a weekly basis throughout the breeding season following a standard and repeatable sampling routine. Whilst all captured newts are recorded we analyse only the adult newt data since the natural markings used to uniquely identify individuals may still be developing in juvenile newts. We consider the data collected between 2002–2013 inclusive, a total of 12 years of data (available in the Supplementary Material). In total there are 241 capture occasions across the 12 years. The number of capture occasions each year varies; traps are set from the final week in February, which is typically before any newts arrive, and continue to be set until no further newts are captured or the water level in the ponds falls making trapping problematic. We format the data such that the first capture occasion occurs within the same week every year (this may require truncating leading zeros from the capture histories within some years). Originally consisting of four ponds the site was extended in 2009 to a total of eight ponds which were then first colonised during the 2010 breeding season. We define the observable capture states to be the type of pond (old or new) the individual is captured in; the "old" ponds were available in all years 2002–2013 whilst the second state, "new" ponds, were available in years 2010–2013. The ponds are all located close together at the field study site (1–12 m apart) and so movement between all eight ponds is possible and it is the environmental differences (e.g., the amount of vegetation) between the old and the new ponds that is likely to affect the choice of pond. In total $n = 106$ unique individuals were captured with a total of 1536 recaptures (ranging from 0 to 68 recaptures per individual, median 6.5 recaptures) during the 12 years of sampling.

To consider the choice of model we first model the capture-recapture data, without considering the additional state information, using the HMM formulation of the multi-period stopover model. We perform a systematic search through a series of models of varying complexity in terms of the parameter dependencies. We start with the most basic model where all the parameters are considered to be constant and shared across all years. Improvement in the model fit is determined through the AIC statistic using a "step-up" approach in order to avoid choosing an overly complex model (McCrea and Morgan (2011)). Due to the large number of capture occasions, we use a logistic regression over the secondary occasions for both the arrivals and retentions within each primary period (rather than estimating probabilities for each secondary occasion separately, this approach would require a very large number of parameters and the sample size here is comparatively small). The model chosen by AIC (where the state information is ignored) includes year-dependent recruitment probabilities, a constant survival probability between each breeding season, and capture probabilities that are both year- and occasion-dependent, that is, a different capture probability on every capture occasion. For the logistic regressions on arrival and retention, the intercepts are constant and shared across all years whilst the gradients are year-dependent, with the gradient estimated separately for each year. Assessing goodness-of-fit is challenging for complex capture-recapture models and appreciable pooling is generally required

to be able to classically compare observed and expected values. Due to the very sparse histories in this application cell numbers are very small, the majority smaller than 5. Pooling would make the identification of areas of poor model fit difficult to diagnose. For this reason we rely on the stepwise search and prior understanding of the species to support our model selection.

We now consider the additional observable states (old or new ponds). This additional information is available for the 2010–2013 breeding seasons (all ponds in 2002–2009 are old ponds and so the multi-state parameters are not required for these years). Due to the large number of capture occasions, and very small population size, we remove the occasion-dependence from the capture probabilities and instead allow them to be dependent on both year and state. We also estimate the initial discrete state probabilities and transition probabilities between the different observable states for each year where the multi-state data is available (2010–2013). The results from fitting the multi-state multi-period stopover model are given in Figures 3 and Table 1. Figure 4 shows the estimated abun-

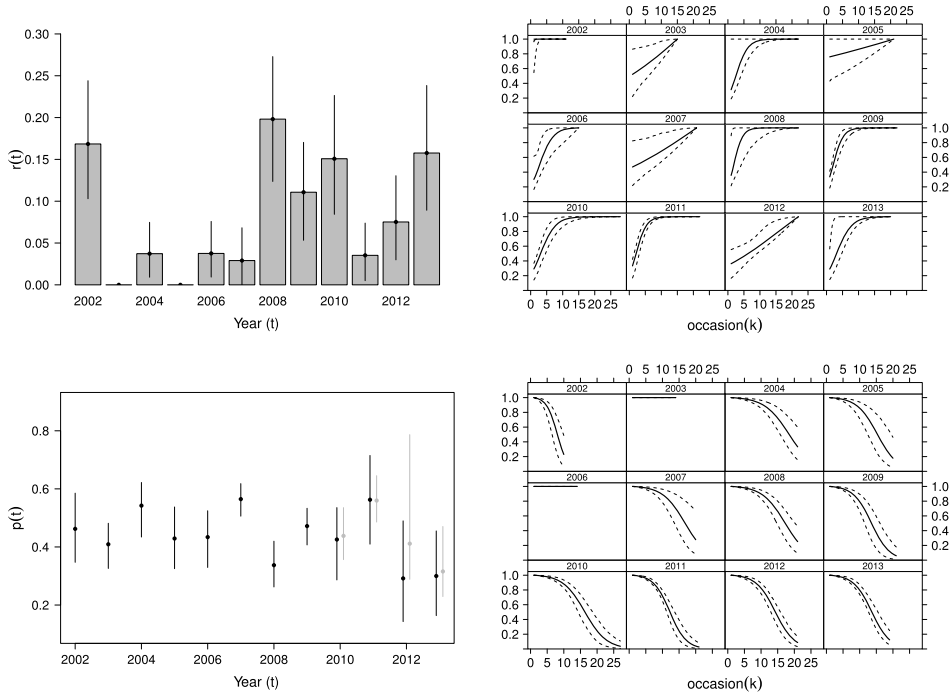

FIG. 3. *Maximum likelihood estimates and* 95% *bootstrap confidence intervals from the multi-state multi-period stopover model for*: (*top*, *left*) *recruitment probabilities for years* 2002–2013; (*top*, *right*) *logistic regression on arrival probabilities for years* 2002–2013; (*bottom*, *left*) *capture probabilities for years* 2002–2013 *for the old ponds* (*black*) *and years* 2010–2013 *for the new ponds* (*grey*) *and*; (*bottom*, *right*) *logistic regression on retention probabilities for years* 2002–2013 *of the great crested newt study.*

TABLE 1

Maximum likelihood estimates and bootstrap standard errors from the multi-state multi-period stopover model for the initial discrete state and transition probabilities for the old (state 1) and new (state 2) ponds for years 2010–2013 of the great crested newt study

	Year (t)			
Parameter	2010	2011	2012	2013
$\alpha(t, 1)$	0.69 (0.12)	0.28 (0.08)	0.48 (0.23)	0.33 (0.11)
$\psi_{12}(t)$	0.11 (0.03)	0.17 (0.06)	0.26 (0.08)	0.14 (0.05)
$\psi_{21}(t)$	0.06 (0.02)	0.10 (0.03)	0.17 (0.08)	0.10 (0.06)

dance (through a forward-backward-type algorithm; see the Supplementary Material, Worthington et al. (2019), Appendix C) for each occasion and observable state in years 2010–2013. Estimated abundance for each year 2002–2013 and estimated abundance on each occasion for years 2002–2009 are given in the Supplementary Material (Worthington et al. (2019), Appendix E). Standard errors and 95% confidence intervals are estimated through a nonparametric bootstrap (resampling indi-

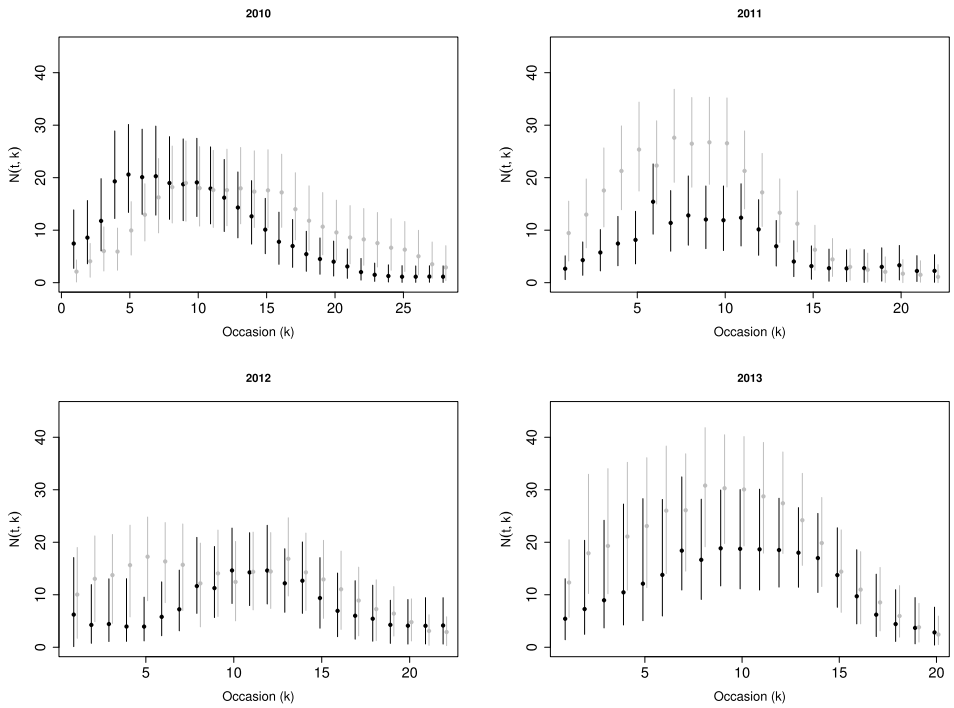

FIG. 4. *Estimated abundances for old ponds (black) and new ponds (grey) with 95% bootstrap confidence intervals for years 2010–2013 of the great crested newt study.*

vidual capture histories) to ensure intervals remain within permissible ranges and to avoid issues with boundary estimates (e.g., recruitment and arrival probabilities very close to 0).

The results indicate the data collection process is close to a complete census of individuals present at the site. The total population of newts that visited the site at least once during the 12 years of sampling is estimated to be $N = 107.2$ (SE 0.47) of which $n = 106$ were captured. The parameter estimates and estimated state-abundance in Figure 4 also indicate some possible differences between the old and the new ponds. When the new ponds were initially colonised in 2010 there appears to have been a preference for the old ponds as an initial choice when newts arrived at the site. This is likely due to the amount of vegetation and invertebrates in the ponds; the older ponds had longer to mature than the new ponds and therefore contained significantly more food, predators and plant cover. It appears that upon arrival at the site the newts have moved towards the new ponds as their initial choice as the new ponds have become more established. The capture probabilities indicate clear temporal variation and with the exception of 2012 the capture probabilities in both old and new ponds are very similar. The movement between the ponds is quite low, newts appear to show high fidelity to the type of pond they are in (old or new) with a consistently higher fidelity for the new ponds. The survival probability for this population of newts between breeding seasons, assumed to be constant between years, is estimated to be 0.82 (SE 0.027).

5. Discussion. In this paper we have developed a generalised multi-state multi-period stopover model. This global model for capture-recapture data offers extensions to many of the commonly applied capture-recapture models. In particular the multi-state single-period stopover model combines the AS model and stopover models to allow the capture probabilities to be time-, age- and state-dependent. The new model is a fully open population model able to estimate total abundance and therefore likely to resolve long-standing issues concerning the assumption of closure when sampling animal populations repeatedly over short time frames. The multi-state multi-period stopover model is a further extension of this multi-state stopover model considering multiple periods of capture occasions within a single tractable likelihood. Forming the model as an HMM offers the estimation of abundance for each primary period along with each occasion and state within primary periods using standard techniques. Models that allow for the combining of information, either across several years of data collection or different sources of information, for example, count data, are widely used in ecological applications (Besbeas et al. (2002)).

The model could be further generalised in a number of ways. As stated above the transition probabilities between states need not be constant within each primary period, additionally the transition, arrival and retention probabilities could have further dependence on the time since recruitment. For example, older individuals

in the population may demonstrate different time of arrival and state transition behaviour compared to newly recruited individuals. Similarly, retention probabilities may be different for more experienced breeders compared to first time breeders. Further consideration could also be given to the initial discrete state distribution. Currently this is modelled independently for each primary period with no information on state-preference brought forward from preceding primary periods. Possible extensions would be to allow choice of initial state to be related to the final state of the previous primary period, or as a function of the time spent in each state up to that point. A fully general model incorporating the above is likely to require a large volume of data to estimate the model parameters regardless of whether they are identifiable. Whilst it would be preferable to use model selection techniques to determine an appropriate model structure, it is reasonable that choice of model structure for such a complex model be motivated by the underlying ecology of the species under study. The model structure we have chosen to develop is motivated by the application on great crested newts.

This likelihood is constructed using an HMM form leading to an efficient likelihood expression that can be maximised using standard optimisation algorithms and software. This structure also permits the extension to include additional complexities in a straightforward manner. For example, in this paper we assume that the state information is recorded with certainty. In practice this may not be the case but the model can be extended further to incorporate such state-uncertainty by introducing additional state assignment probabilities (King and McCrea (2014), King and Langrock (2016)).

In these derivations we assume the states are discrete. In the case of continuous state information the approach of the HMMs above could still be applied by using a fine discretization of the continuous states into a discrete form (Langrock and King (2013)). Care would need to be adopted in this instance to avoid the dimensions of the matrices involved becoming too large leading to computational issues.

Further extensions to these models could include the addition of a state-dependence to the retention probabilities. This would allow the departure of individuals to be modelled differently depending on their final state in a given year. To reduce the number of parameters estimated from the capture-recapture data alone, covariates could also be considered. As with the multi-period stopover model, consideration could also be given to temporary migration and the idea of individuals skipping attendance in some years. For instance the success or failure to breed in a given year may lead an individual to skip the following year to improve their body condition before returning in later years to reattempt breeding. In the case of the newts this behaviour is more likely in females as they have to invest more energy to produce eggs each year. This information would need to be incorporated in the primary level of the model where the behaviour in a given year is summarised into a state on the primary level. Again such extensions can be considered and the efficient HMM likelihood exploited.

Acknowledgements. The authors gratefully acknowledge all of the volunteers and students who collected the data from the newts at the University of Kent, particularly Sue Young, Amy Wright and Brett Lewis.

SUPPLEMENTARY MATERIAL

Supplement A: Supplementary appendices (DOI: 10.1214/19-AOAS1264SUPPA; .pdf). Model diagram. Details on the forward-backward-type algorithms used to estimate primary abundance and abundance for each occasion and state within primaries. Additional output from the simulation study conducted in Section 3. Abundance estimates for each year of the great crested newt study in Section 4.

Supplement B: Data (DOI: 10.1214/19-AOAS1264SUPPB; .zip). Data set on great crested newts used to illustrate the multi-state multi-period stopover model in Section 4.

Supplement C: Code (DOI: 10.1214/19-AOAS1264SUPPC; .zip). R code to perform the simulation study in Section 3 and the analysis on great crested newts in Section 4.

REFERENCES

BESBEAS, P., FREEMAN, S. N., MORGAN, B. J. T. and CATCHPOLE, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* **58** 540–547. MR1933532

BROWNIE, C., HINES, J. E., NICHOLS, J. D., POLLOCK, K. H. and HESTBECK, J. B. (1993). Capture-recapture studies for multiple strata including non-Markovian transitions. *Biometrics* **49** 1173–1187.

DUPUIS, J. A. and SCHWARZ, C. J. (2007). A Bayesian approach to the multistate Jolly–Seber capture-recapture model. *Biometrics* **63** 1015–1022, 1309. MR2414577

GIMENEZ, O., ROSSI, V., CHOQUET, R., DEHAIS, C., DORIS, B., VARELLA, H., VILA, J.-P. and PRADEL, R. (2007). State-space modelling of data on marked individuals. *Ecol. Model.* **206** 431–438.

GRIFFITHS, R. A., FOSTER, J., WILKINSON, J. W. and SEWELL, D. (2015). Science, statistics and surveys: A herpetological perspective. *J. Appl. Ecol.* **52** 1413–1417.

KENDALL, W. L. and BJORKLAND, R. (2001). Using open robust design models to estimate temporary emigration from capture-recapture data. *Biometrics* **57** 1113–1122. MR1973818

KENDALL, W. L., POLLOCK, K. H. and BROWNIE, C. (1995). A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics* **51** 293–308.

KENDALL, W. L., STAPLETON, S., WHITE, G. C., RICHARDSON, J. I., PEARSON, K. N. and MASON, P. (2018). A multistate open robust design: Population dynamics, reproductive effort, and phenology of sea turtles from tagging data. *Ecol. Monogr.* **89** e1329.

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

KING, R. (2014). Statistical ecology. *Annu. Rev. Stat. Appl.* **1** 401–426.

KING, R. and LANGROCK, R. (2016). Semi-Markov Arnason–Schwarz models. *Biometrics* **72** 619–628. MR3515788

KING, R. and MCCREA, R. S. (2014). A generalised likelihood framework for partially observed capture-recapture-recovery models. *Stat. Methodol.* **17** 30–45. MR3133584

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

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

LEWIS, B., GRIFFITHS, R. A. and WILKINSON, J. W. (2017). Population status of great crested newts (*Triturus cristatus*) at sites subjected to development mitigation. *Herpetological Journal* **27** 133–142.

MCCREA, R. S. and MORGAN, B. J. T. (2011). Multistate mark-recapture model selection using score tests. *Biometrics* **67** 234–241. MR2898835

MCCREA, R. S., MORGAN, B. J. T., GIMENEZ, O., BESBEAS, P., LEBRETON, J.-D. and BREGNBALLE, T. (2010). Multi-site integrated population modelling. *J. Agric. Biol. Environ. Stat.* **15** 539–561. MR2788639

PLEDGER, S., EFFORD, M., POLLOCK, K. H., COLLAZO, J. A. and LYONS, J. E. (2009). Stopover duration analysis with departure probability dependent on unknown time since arrival. In *Environmental and Ecological Statistics* **3** (D. L. Thomson, E. G. Cooch and M. J. Conroy, eds.) 349–363.

POLLOCK, K. H. (1982). A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management* **46** 752–757.

ROYLE, J. A. (2008). Modeling individual effects in the Cormack–Jolly–Seber model: A state-space formulation. *Biometrics* **64** 364–370, 664. MR2432405

SCHOFIELD, M. R. and BARKER, R. J. (2008). A unified capture-recapture framework. *J. Agric. Biol. Environ. Stat.* **13** 458–477. MR2590940

SCHWARZ, C. J. and ARNASON, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* **52** 860–873. MR1411736

SCHWARZ, C. J. and STOBO, W. T. (1997). Estimating temporary migration using the robust design. *Biometrics* **53** 178–194.

WORTHINGTON, H., MCCREA, R., KING, R. and GRIFFITHS, R. (2019). Supplement to "Estimating abundance from multiple sampling capture-recapture data via a multi-state multi-period stopover model." DOI:10.1214/19-AOAS1264SUPPA, DOI:10.1214/19-AOAS1264SUPPB, DOI:10.1214/19-AOAS1264SUPPC.

ZUCCHINI, W., MACDONALD, I. L. and LANGROCK, R. (2016). *Hidden Markov Models for Time Series*: *An Introduction Using R*, 2nd ed. *Monographs on Statistics and Applied Probability* **150**. CRC Press, Boca Raton, FL. MR3618333

H. WORTHINGTON
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF ST ANDREWS
THE OBSERVATORY
BUCHANAN GARDENS
ST ANDREWS
FIFE
KY16 9LZ
UNITED KINGDOM
E-MAIL: hw233@st-andrews.ac.uk

R. MCCREA
SCHOOL OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE (SMSAS)
UNIVERSITY OF KENT
SIBSON BUILDING
PARKWOOD ROAD
CANTERBURY
CT2 7FS
UNITED KINGDOM
E-MAIL: R.S.McCrea@kent.ac.uk

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

R. GRIFFITHS
DURRELL INSTITUTE OF CONSERVATION
AND ECOLOGY
SCHOOL OF ANTHROPOLOGY AND CONSERVATION
UNIVERSITY OF KENT
MARLOWE BUILDING
CANTERBURY
KENT
CT2 7NR
UNITED KINGDOM
E-MAIL: R.A.Griffiths@kent.ac.uk