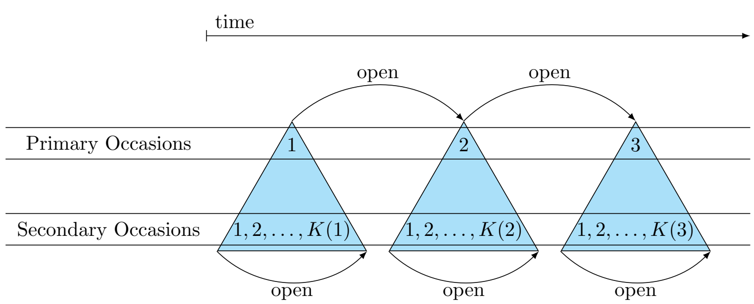
# SUPPLEMENT TO "ESTIMATING ABUNDANCE FROM MULTIPLE SAMPLING CAPTURE-RECAPTURE DATA VIA A MULTI-STATE MULTI-PERIOD STOPOVER MODEL"

By Hannah Worthington[1], Rachel McCrea[2], Ruth King[3] and Richard Griffiths[2]

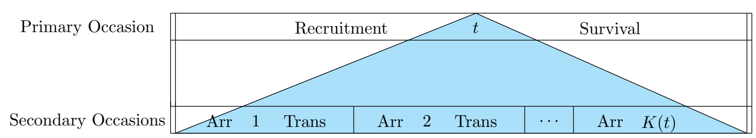
*University of St Andrews*[1], *University of Kent*[2] *and University of Edinburgh*[3]

## APPENDIX A: MODEL DIAGRAM

The diagrams in Figures 1 and 2 display the structure of the model. Recruitment occurs immediately prior to a primary occasion with survival immediately after a primary occasion. Within secondary occasions, arrivals occur immediately prior to a capture occasion with transitions between the different states of the HMM (this includes transitions to having left the site, or between observable states) immediately after a capture event.




Supplementary Figure. 1. *Diagram depicting the structure of the primary and secondary occasions.*




Supplementary Figure. 2. *Diagram depicting the order of events within a single primary period.*

## APPENDIX B: PRIMARY PERIOD ABUNDANCE ESTIMATION

We implement a forward-backward-type algorithm to estimate primary period abundance. In order to estimate this abundance we need to estimate the conditional state probabilities (conditional on the observed capture history on the primary level) where state refers to whether an individual has been recruited to the population. In general, the states of an HMM are hidden, however, in the case of capture-recapture the states are partially observed. When an individual is captured their state is observed and so there is only uncertainty when an individual is missed. For this reason we call this approach a forward-backward-type algorithm. Additionally, the forward-backward algorithm is typically used for state decoding, it is used to identify the most probable hidden state on each occasion given the full capture history. We have chosen not to decode the state sequence as this would imply that individuals with the same capture history also share the same state sequence and this is unlikely to be true in reality. Therefore, to estimate abundance we take a probabilistic approach and sum over the conditional state probabilities for all $N$ individuals estimated to be in the population ($N$ total population size estimated in the model likelihood).

The forward-backward algorithm finds the conditional state probabilities through a combination of forward and backward probabilities. These probabilities are estimated for each observed individual $i = 1, \dots, n$ along with an all-zero capture history to account for missed individuals. The model includes total abundance $N$ as a parameter and so the number of missed individuals is $n_m = N - n$.

Let $Z_i(t)$ be the random variable associated with observing individual $i$ on primary occasion $t = 1, \dots, T$, then $Z_i(t) \in \{0, 1\}$ for $i = 1, \dots, n$ and $t = 1, \dots, T$ (for $T$ the total number of primary occasions). Let $H_i(t)$ be the random variable associated with the state of individual $i$ on primary occasion $t$ then $H_i(t) \in \{1, 2, 3\}$ where

$$H_i(t) = \begin{cases} 1 & \text{not yet recruited} \\ 2 & \text{recruited} \\ 3 & \text{departed from population} \end{cases}$$

for $i = 1, \dots, n$ and $t = 1, \dots, T$. In this derivation we assume no time since recruitment effects on survival (as per the simulation study in section 3). If time since recruitment effects were included the hidden states would match those in the model derivation in section 2.2 and abundance could be estimated for each primary occasion across different time since recruitment groups.

Let $\boldsymbol{f} = \{f_i(t,h) : i = 1,\dots,n,\ t = 1,\dots,T,\ h \in \{1,2,3\}\}$ denote the forward probabilities for each observed individual. The forward probabilities find the joint probability of the capture history up to occasion $t$ with the state on occasion $t$,

$$f_i(t,h) = (Z_i(1) = z_i(1), Z_i(2) = z_i(2), \dots, Z_i(t) = z_i(t), H_i(t) = h)$$

where $z_i(t)$ is the observation or not of individual $i$ in primary occasion $t$. The parameters of the model associated with movement between the primary states are the recruitment probabilities $\boldsymbol{r} = \{r(t) : t = 1,\dots,T\}$ (and conditional recruitment probabilities $r^*(t)$) and survival $s$ (constant across all primaries). We initialise the forward probabilities on occasion 1 using,

$$f_i(1,h) = \begin{cases} (1 - r(1)) \times \mathbb{P}(Z_i(t) = z_i(t) \mid H_i(t) = 1) & h = 1 \\ r(1) \times \mathbb{P}(Z_i(t) = z_i(t) \mid H_i(t) = 2) & h = 2 \\ 0 \times \mathbb{P}(Z_i(t) = z_i(t) \mid H_i(t) = 3) & h = 3. \end{cases}$$

For the remaining primary occasions $t = 2,\dots,T$ we use a recursive approach,

$$f_i(t,h) = \sum_{j \in 1,2,3} f_i(t-1,j)\mathbf{\Gamma}(t-1)[j,h]\mathbb{P}(Z_i(t) = z_i(t) \mid H_i(t) = h)$$

where

$$\mathbf{\Gamma}(t) = \begin{pmatrix} 1 - r^*(t+1) & r^*(t+1) & 0 \\ 0 & s & 1 - s \\ 0 & 0 & 1 \end{pmatrix}$$

is the transition matrix between the hidden states. The conditional probability $\mathbb{P}(Z_i(t) \mid H_i(t))$ describes whether an individual is seen or not seen in primary $t$ given the hidden state of the individual. These probabilities are taken from the observation probability matrices of the HMM,

$$\mathbf{P}(t, z_i(t)) = \begin{cases} \text{diag}(1, L_0(t), 1) & z_i(t) = 0 \\ \text{diag}(0, L_i(t), 0) & z_i(t) = 1 \end{cases}$$

where $L_0(t)$ and $L_i(t)$ are respectively the probabilities of observing an all-zero or individual capture history $i = 1,\dots,n$ across the secondary occasions within primary $t$, i.e. the likelihood contribution for a single-period stopover model.

Let $\boldsymbol{b} = \{b_i(t,h) : i = 1,\dots,n,\ t = 1,\dots,T,\ h \in \{1,2,3\}\}$ be the backward probabilities for all observed individuals. The backward probabilities find the conditional probability of observing the remaining capture history from occasion $t+1$ onwards given the state on occasion $t$,

$$b_i(t,h) = \mathbb{P}(Z_i(t{+}1) = z_i(t{+}1), Z_i(t{+}2) = z_i(t{+}2), \dots, Z_i(T) = z_i(T) \mid H_i(t) = h).$$

The backward probabilities are also calculated in a recursive manner, we initialise for the final occasion $T$,

$$b_i(T,h) = 1$$

for $h \in \{1,2,3\}$. For occasions $t = 1,\dots,T-1$,

$$b_i(t,h) = \sum_{j\in\{1,2,3\}} \mathbf{\Gamma}(t)[h,j]\mathbb{P}(Z_i(t+1) = z_i(t+1) \mid H_i(t+1) = j) b_i(t+1,j).$$

The product of the forward and backward probabilities gives the joint probability of the entire capture history and the current state,

$$f_i(t,h) {\times} b_i(t,h) = \mathbb{P}(Z_i(1) = z_i(1), Z_i(2) = z_i(2), \dots, Z_i(T) = z_i(T), S_i(t) = h)$$

for $i = 1,\dots,n$, $t = 1,\dots,T$ and $h \in \{1,2,3\}$. To estimate the abundance in each primary we need to find the conditional probabilities for each hidden state given the capture histories of each observed individual along with an all-zero capture history (to account for missed individuals). $f_0(t,h)$ and $b_0(t,h)$, the forward and backward probabilities for an all zero-history, are found in the analogous way to the observed histories above. We use the result

$$\mathbb{P}(H_i(t) = h \mid Z_i(1) = z_i(1), Z_i(2) = z_i(2), \dots, Z_i(T) = z_i(T)) = \frac{f_i(t,h) \times b_i(t,h)}{f_i(t,\cdot) b_i(t,\cdot)^T}$$

for $f_i(t,\cdot) = \{f_i(t,h) : h \in \{1,2,3\}\}$, similarly for $b_i(t,\cdot)$ and $b_i(t,\cdot)^T$ denoting the transpose.

Taking a probabilistic approach the estimators for abundance in each hidden state and primary period, $N(t,h)$ are given by,

$$N(t,h) = \left( \sum_{i=1}^{n} \frac{f_i(t,h) b_i(t,h)}{f_i(t,\cdot) b_i(t,\cdot)^T} \right) + \left( n^m \times \frac{f_0(t,h) b_0(t,h)}{f_0(t,\cdot) b_0(t,\cdot)^T} \right).$$

Through these state abundances it is possible to see the progression of the population through recruitment, attendance and then departure from the population.

## APPENDIX C: SECONDARY OCCASION STATE ABUNDANCE ESTIMATION

The estimation of abundance per state and secondary occasion within a primary period follows in a similar manner to primary abundance.

Given an estimate for the number of individuals in the recruited population for primary occasion $t$ ($N(t,2)$ from appendix A), the forward-backward algorithm can estimate the abundance per state and secondary occasion. In the following we take the definitions from the model derivation in section 2.2.

Let $X_i(t,k)$ be the random variable associated with $x_i(t,k)$ the capture history for individual $i = 1,\dots,n(t)$ on occasion $k = 1,\dots,K(t)$ in primary period $t = 1,\dots,T$. Let $H_i(t,k)$ be the random variable associated with the hidden state of individual $i$ on occasion $k$ of primary period $t$. Then $X_i(t,k) \in \{0,1,2\}$ corresponding to a non-capture or capture in one of the two observable discrete states and $H_i(t,k) \in \mathcal{H}$ where $\mathcal{H}$ is the set of hidden states of the secondary level of the HMM for $t = 1,\dots,T$ and $k = 1,\dots,K(t)$. These states include progression through time since arrival and movement between the two observable states (this approach can be generalised to more observable states).

We initialise the forward probabilities for each hidden state $h \in \mathcal{H}$,

$$f_i(t,1,h) = \boldsymbol{\pi}(t,1)[h] \times \mathbb{P}(X_i(t,k) = x_i(t,k) \mid H_i(t,k) = h).$$

For the remaining occasions,

$$f_i(t,k,h) = \sum_{j\in\mathcal{H}} f_i(t,k{-}1,j)\boldsymbol{\Gamma}(t,k{-}1)[j,h]\mathbb{P}(X_i(t,k) = x_i(t,k) \mid H_i(t,k) = h).$$

The conditional probability of the observation $X_i(t,k)$ given a hidden state is taken from $\mathbf{P}(t,k,x_i(t,k))$.

The backward probabilities are initialised on the final secondary occasion within a primary period

$$b_i(t,K(t),h) = 1.$$

For occasions $k = 1,\dots,K(t)-1$,

$$b_i(t,k,h) = \sum_{j\in\mathcal{H}} \boldsymbol{\Gamma}(t,k{-}1)[h,j]\mathbb{P}(X_i(t,k{+}1) = x_i(t,k{+}1) \mid H_i(t,k{+}1) = j)b_i(t,k{+}1,j).$$

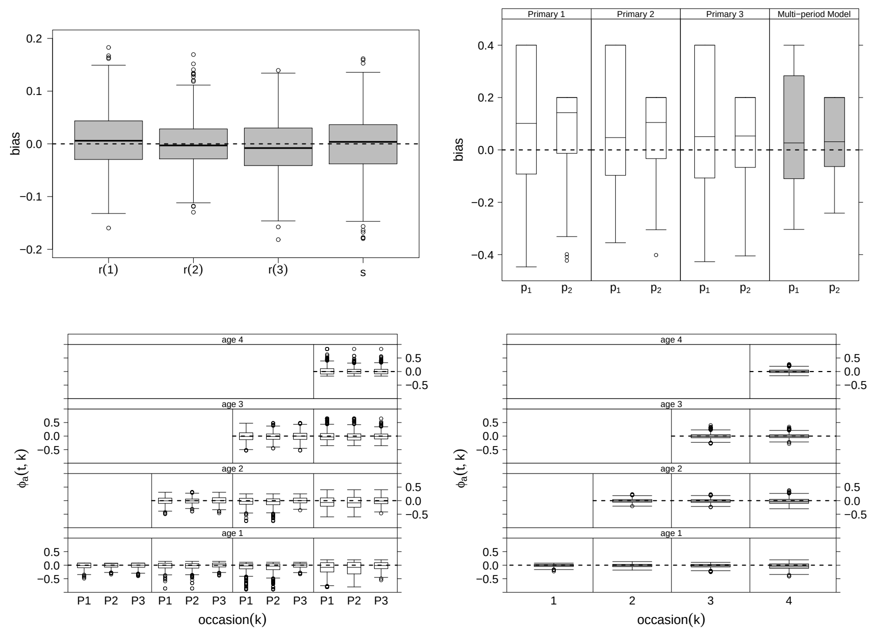
As above the product of the forward and backward probabilities give the joint probability of the entire capture history over the secondary occasions within a primary given the hidden state.

To estimate the abundance in each state we find the forward and backward probabilities for each observed individual $i = 1, \ldots, n(t)$ in primary period $t$ along with $f_0(t, k, h)$ and $b_0(t, k, h)$ the analogous probabilities for an all-zero history. Within a primary period, given an estimate $N(t)$ for the number of individuals in the recruited population, the number of missed individuals $n^m(t) = N(t) - n(t)$. As above we take a probabilistic approach to the abundance estimation to avoid assigning the same state sequence to every individual with the same capture history.

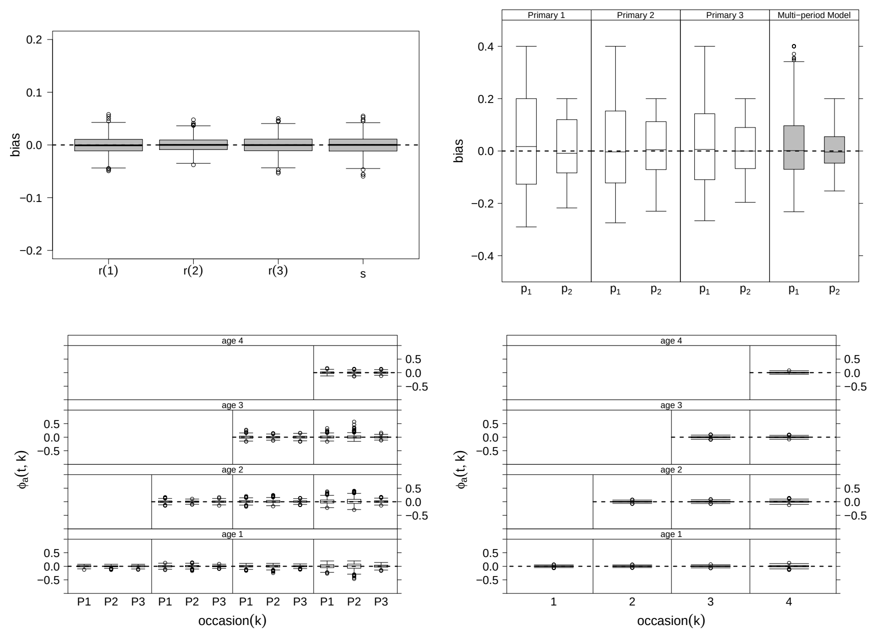
$$N(t,k,h) = \left( \sum_{i=1}^{N(t)} \frac{f_i(t,k,h) b_i(t,k,h)}{f_i(t,k,\cdot) b_i(t,k,\cdot)^T} \right) + \left( n^m(t) \times \frac{f_0(t,k,h) b_i(t,k,h)}{f_0(t,k,\cdot) b_i(t,k,\cdot)^T} \right).$$

To achieve estimates for the abundance in each observable state (old and new ponds in the great crested newt application) sum over all ages (time since arrival) within the same state.

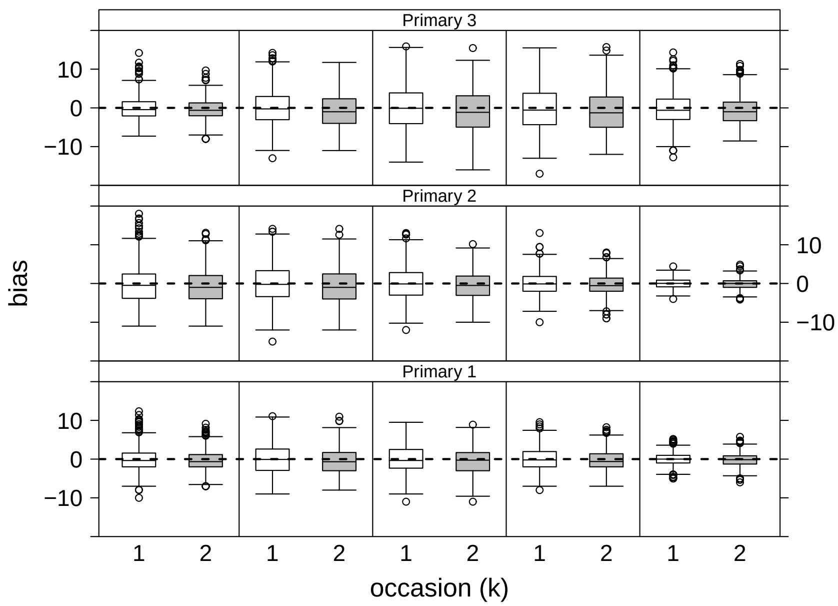
# APPENDIX D: ADDITIONAL SIMULATION STUDY RESULTS



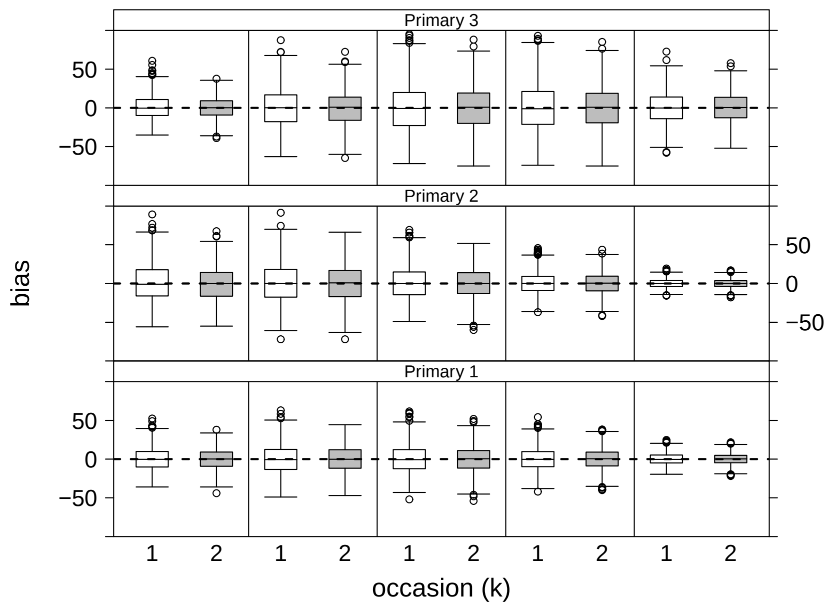

Supplementary Figure. 3. *Results from the simulation study where $N = 100$ for: (top, left) bias of the recruitment and survival probabilities for the multi-period model; (top, right) bias of the capture probabilities in each primary period for the single-period model (white) and the multi-period model (grey); (bottom, left) bias of the retention probabilities for each primary period, each capture occasion within primaries and ages for the single-period model and; (bottom, right) bias of the retention probabilities for each capture occasion and ages (shared across primaries) for the multi-period model.*




Supplementary Figure. 4. *Results from the simulation study where* $N = 1000$ *for: (top, left) bias of the recruitment and survival probabilities for the multi-period model; (top, right) bias of the capture probabilities in each primary period for the single-period model (white) and the multi-period model (grey); (bottom, left) bias of the retention probabilities for each primary period, each capture occasion within primaries and ages for the single-period model and; (bottom, right) bias of the retention probabilities for each capture occasion and ages (shared across primaries) for the multi-period model.*

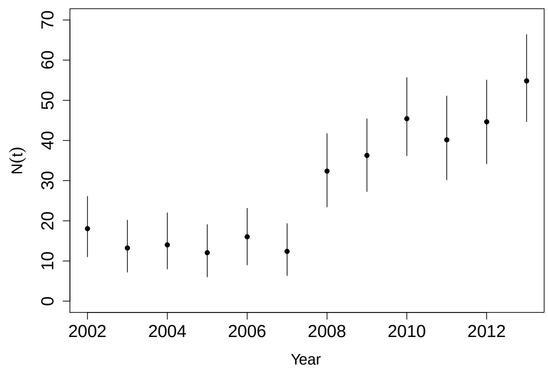



Supplementary Figure. 5. *Bias of the population size estimates for each secondary occasion and discrete state (state 1 white, state 2 grey) for the multi-period model from the simulation study where* $N = 100$.

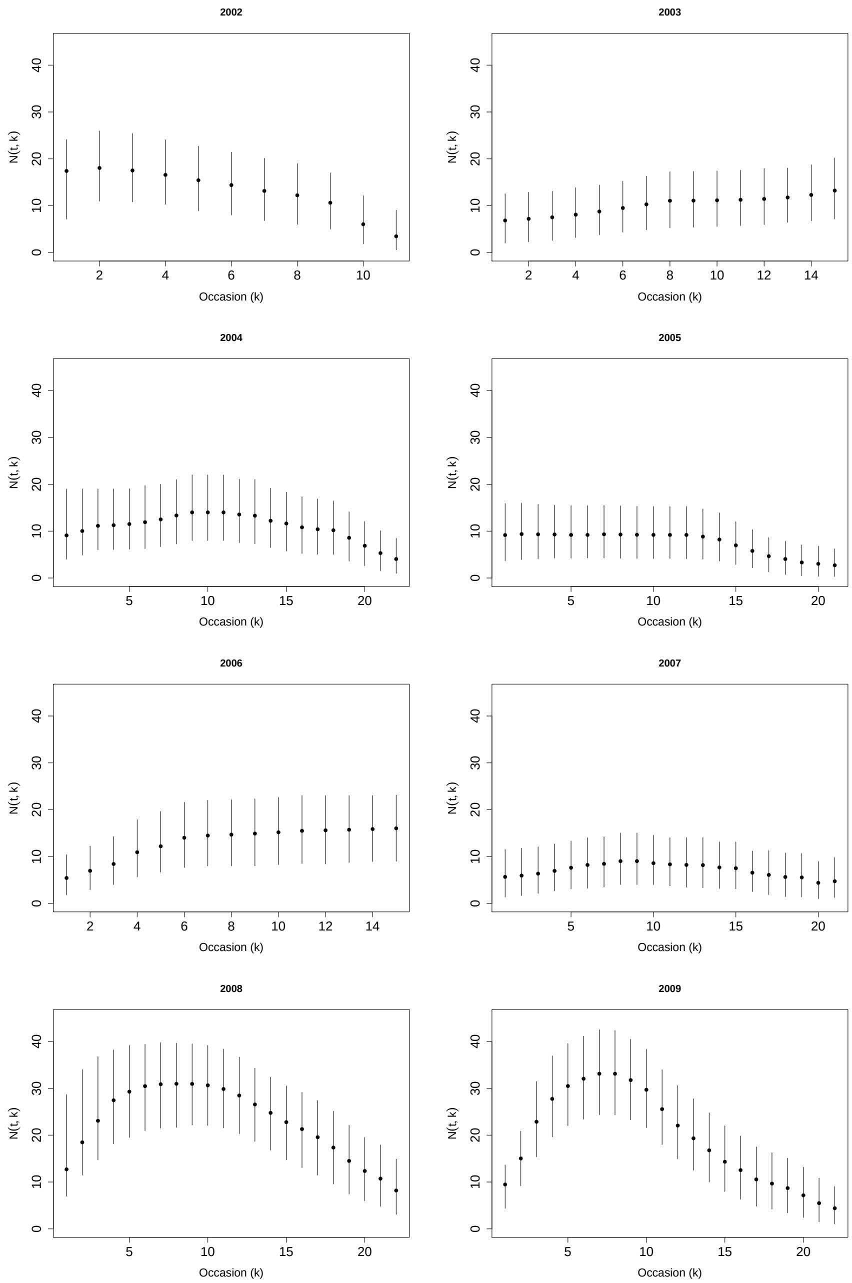



Supplementary Figure. 6. *Bias of the population size estimates for each secondary occasion and discrete state (state 1 white, state 2 grey) for the multi-period model from the simulation study where* $N = 1000$.

## APPENDIX E: ABUNDANCE ESTIMATES FOR THE GREAT CRESTED NEWT STUDY




SUPPLEMENTARY FIGURE. 7. *Estimated abundance and 95% bootstrap confidence intervals for each year 2002–2013 of the great crested newt study.*




Supplementary Figure. 8. *Estimated abundance for each secondary capture occasion and 95% bootstrap confidence intervals for each year 2002–2009 of the great crested newt study.*

School of Mathematics and Statistics
The University of St Andrews
The Observatory
Buchanan Gardens
St Andrews
Fife
KY16 9LZ
E-mail: hw233@st-andrews.ac.uk

School of Mathematics, Statistics and
Actuarial Science (SMSAS)
The University of Kent
Sibson Building
Parkwood Road
Canterbury
CT2 7FS
E-mail: R.S.McCrea@kent.ac.uk

School of Mathematics
University of Edinburgh
James Clerk Maxwell Building
The King's Buildings
Peter Guthrie Tait Road
Edinburgh
EH9 3FD
E-mail: Ruth.King@ed.ac.uk

Durrell Institute of Conservation
and Ecology
School of Anthropology and Conservation
University of Kent
Marlowe Building
The University of Kent
Canterbury
Kent
CT2 7NR
E-mail: R.A.Griffiths@kent.ac.uk