Flexible Behavioral Capture–Recapture Modeling

Danilo Alunni Fegatelli1,* and Luca Tardella2,**

1Dipartimento di Sanità Pubblica e Malattie Infettive – Sapienza Università di Roma (Italy)
2Dipartimento di Scienze Statistiche – Sapienza Università di Roma (Italy)

*email: danilo.alunnifegatelli@uniroma1.it
**email: luca.tardella@uniroma1.it

Summary. We develop alternative strategies for building and fitting parametric capture–recapture models for closed populations which can be used to address a better understanding of behavioral patterns. In the perspective of transition models, we first rely on a conditional probability parameterization. A large subset of standard capture–recapture models can be regarded as a suitable partitioning in equivalence classes of the full set of conditional probability parameters. We exploit a regression approach combined with the use of new suitable summaries of the conditioning binary partial capture histories as a device for enlarging the scope of behavioral models and also exploring the range of all possible partitions. We show how one can easily find unconditional MLE of such models within a generalized linear model framework. We illustrate the potential of our approach with the analysis of some known datasets and a simulation study.

Key words: Behavioral response; Markov models; Mark–recapture; Memory effect; Memory-related summary statistics; Population size.

1. Introduction

Multiple capture–recapture models are successfully employed to infer the unknown size and characteristics of a finite population whose complete enumeration is difficult due to the elusive nature of its units. These models are also used in fields other than ecology such as software engineering, social sciences, and epidemiology. Much progress has been made to enlarge the scope of available models and refine inferential techniques. There are now many available monographs and review articles which can offer a wide perspective of the state of the art (White et al., 1982; Seber, 1987; Schwarz and Seber, 1999; Borchers et al., 2002; Amstrup et al., 2005; Böhning, 2008; Royle et al., 2013; McCrea and Morgan, 2014).

In this article, we develop tools for a better understanding of the behavioral response to capture within a general model framework. Empirical studies have provided evidence that mice, voles, small mammals, and butterflies, among others, often exhibit a response to capture (Yang and Chao, 2005; Ramsey and Severns, 2010). However, relevant response to capture is also at stake in studies involving human population (Farcomeni and Scacciatelli, 2013). The classical way to account for behavioral response is to assume that once a unit is captured, its probability of being recaptured in all future trapping occasions is modified permanently. This enduring effect is called trap-happiness or trap-shyness according to whether the recapture probability becomes larger or smaller. This very simple one-parameter model flexibility sheds some light on the population under study and the presence of a behavioral effect can have a great impact on the estimate of the unknown population size (Lee and Chen, 1998; Chao et al., 2000; Yip et al., 2000; Hwang et al., 2002; Lee et al., 2003; Ghosh and Norris, 2005; Hwang and Huggins, 2011; Alunni Fegatelli and Tardella, 2013). However, this specific type of behavioral effect is only a limited device to approach the understanding of complex behavioral patterns in multiple capture–recapture designs. In fact, a wider perspective has been introduced in Yang and Chao (2005), where an ephemeral behavioral effect is modeled by using Markov chains. Bartolucci and Pennoni (2007) accounted for dependence and heterogeneity with a hidden Markov model for the sequence of capture probabilities. Recently, new ideas have been put forward by Ramsey and Severns (2010) and Farcomeni (2011, 2015) to enlarge the scope of behavioral patterns with new instances of behavioral effects.

In order to provide a general framework to deal with behavioral effects, we start as in Farcomeni (2011) and Huggins (1989) by reparameterizing the joint probabilities of the observable capture histories in terms of conditional probabilities. Their approach can be regarded within the perspective of transition models for longitudinal binary data (Fitzmaurice and Molenberghs, 2009). Hence, we define suitable numerical summaries of partial capture histories and show how the appropriate handling of these summaries as explanatory variables within a generalized linear model (GLM) framework can help to analyze capture–recapture experiments with interpretable behavioral patterns and improved fit. Unlike the aforementioned articles, we prefer the use of the unconditional likelihood for making inference. Moreover, we show how the proposed modeling approach can be related to the broad class of marginal models in Bartolucci and Forcina (2006). The article is organized as follows: in Section 2, we introduce the basic notation for our set up; the saturated parameterization for the probability of observing a sequence of consecutive binary outcomes corresponding to all capture occasions and the subset of all possible reduced models. This reviews previous approaches starting from the pioneering...
articles by Huggins (1989) and Alho (1990) and more recently reinterpreted and extended in Farcomeni (2011) and Alunni Fegatelli and Tardella (2013). In Section 3, we define alternative numerical summaries of partial histories to be thought of as a memory-effect quantification and exploited in a GLM framework as time-dependent individual behavioral explanatory variables. This allows us to (i) enlarge the scope of available behavioral effect models and (ii) recover many of the existing ones by appropriate partitioning of the range of the behavioral summary statistics. For one of the proposed numerical summaries, we will show its relation with the class of Markov models of any order. In Section 4, we infer on unknown parameters through the maximization of unconditional likelihood and explain how easily this can be done by recycling standard GLM routines. In Section 5, we show the usefulness of our approach in discovering better parsimonious models with real data. In Section 6, we verify the ability of a model selection criterion to discriminate among different behavioral patterns with a simulation study. Section 7 closes with some remarks and a discussion on future developments.

2. Models Based on Partitions of Conditional Probabilities

Let us consider a discrete-time closed capture–recapture experiment in which the unknown population size \( N \) is constant and individual trappings are recorded in \( t \) consecutive times. We suppose that all units act independently, there is no misclassification and no tag loss. For notational convenience, one can assume that units captured during the study are labeled with generic-ordered \( 1 \) dimensional vector denoted with \( \boldsymbol{x} = (x_1, x_2, \ldots, x_t) \) such that \( x_i = 1 \) corresponds to unit \( i \) being captured at occasion \( j \). If we denote with \( X = \{0, 1\} \) the space of capture outcomes, the space of all possible capture histories for each unit is denoted by \( X^t \) and represents the \( t \)-fold Cartesian product of \( X \) where a generic-ordered \( t \)-tuple is \( \boldsymbol{x} = (x_1, x_2, \ldots, x_t) \). The set of all observable capture histories, denoted by \( X^* \), does not contain the \( t \)-tuple corresponding to a no capture history and can be written in terms of set difference \( X^* = X^t \setminus \{(0, \ldots, 0)\} \). Curly braces will be always used for set notation while round parentheses will be used for generic-ordered \( k \)-tuples. As a starting point, no individual heterogeneity is assumed for the probability of being captured at each time. At the end of Section 4, we will discuss the relaxation of this assumption.

In order to setup a natural framework for characterizing the fundamental set of probabilities of all possible complete capture histories, we follow Farcomeni (2011) and we rely upon capture probabilities conditioned on each possible partial capture history as follows: \( p_1() = \Pr(X_1 = 1) \) and \( p_j(x_1, \ldots, x_{j-1}) = \Pr(X_j = 1|x_1, \ldots, x_{j-1}) \) for all \( j > 1 \) and \( (x_1, \ldots, x_{j-1}) \in X^{j-1} \). The initial empty brackets () are understood as the absence of previous capture history at time 1. Conditional probabilities can be arranged with a conventional order in a \( 2^t - 1 \) dimensional vector denoted with \( \boldsymbol{p} = (p_1(), p_2(0), p_2(1), p_3(0, 0), p_3(0, 1), p_3(1, 0), \ldots, p_i(1, \ldots, 1)) \) where, for example, the element \( p_3(0, 1) \) represents the probability of being captured at time 3 given that the unit is not captured in the first occasion, while it is captured in the second occasion. For notational convenience, we will often remove commas between binary digits when representing partial capture histories. The vector \( \boldsymbol{p} \) is a reparameterization of the joint probabilities corresponding to all \( 2^t \) complete capture history configurations in \( X^t \). Conditional probabilities, rather than the joint probabilities, are more easily interpreted in modeling the consequences determined by the change of behavior due to a particular previous history. Under the saturated reparameterization, the probability of never being observed during the experiment is

\[
P_0 = \left(1 - p_1()\right) \prod_{j=2}^{t} \left(1 - p_j(0, \ldots, 0)\right).
\]

This is one of the main quantities for the estimation of the parameter of interest via likelihood maximization since the MLE \( \hat{N} = \frac{M}{\sum_{j=1}^{t} p_j(0, \ldots, 0)} \) where \( P_0 \) is the MLE of \( P_0 \). This is of course true by definition of the conditional likelihood approach, but it is still true with the unconditional likelihood provided that \( P_0 \) is jointly determined with \( N \) according to the definition of the unconditional likelihood (see Farcomeni and Tardella (2012); Sanathanan (1972)).

From the saturated parametrization based on \( \boldsymbol{p} \), one can specify a parsimonious nested model based on a suitable partition of the conditional probabilities in \( \boldsymbol{p} \) in terms of equivalence classes. Let \( H \) be the set of all partial capture histories: \( H = \{(), (0), (1), (00), (10), (01), (11), \ldots\} = \cup_{j=0}^{t-1} X^j \) where \( X^0 = \{()\} \). Let us denote with \( \mathcal{H}_B \), one of the possible partitions of \( H \) in \( B \) subsets \( \mathcal{H}_B = \{H_1, H_2, \ldots, H_B\} \) where each \( H_b \subset H \). The role of the set \( H \) is to list all the partial capture histories which may yield possible changes in the conditional capture probability. Let us denote a generic partial capture history as follows \( \boldsymbol{x} = (x_1, \ldots, x_k) \) where \( k \) is the length of the binary vector. When there is absence of previous capture history \( (x = ()) \) we have \( k = 0 \). For each partition \( \mathcal{H}_B \), we consider a corresponding reduced parameter vector of probabilities denoted with \( \boldsymbol{p}_{\mathcal{H}_B} = (p_{H_1}, \ldots, p_{H_B}) \). The partition of capture histories in equivalence classes is such that if we have two histories \( \boldsymbol{x}, \boldsymbol{x}' \in H_b \) then \( p_{\mathcal{H}_B+1}() = p_{\mathcal{H}_B+1}(\boldsymbol{x}') = p_{\mathcal{H}_B}(\boldsymbol{x}) \). The equivalence class partition \( \mathcal{H}_B \) represents a different way of looking at the linear-constrained approach in Farcomeni (2011) which indeed can be seen as stemmed from the works of Huggins (1989) and Alho (1990). As a simple example of our formalization based on partitions of subsets of \( H \) as opposed to the linear constraint approach, one can consider model \( M_b \). It can be defined using two blocks of equality constraints: \( p_1() = p_2(0) = p_3(0, 0) = \cdots = p_i(0, \ldots, 0) = \pi_f \) and \( p_2(1) = p_3(1, 0) = p_3(0, 1) = \cdots = p_i(1, \ldots, 1) = \pi_r \). It is easy to verify that if we interpret \( \pi_f \) as the probability of first capture and \( \pi_r \) as the probability of being recaptured, one gets the most simple form of behavioral model with enduring effect usually denoted with \( M_b \). Equivalently, in our partition notation, model \( M_b \) corresponds to the bipartition

\[
\mathcal{H}_2(M_b) = \begin{cases} 
H_1 = \{(), (0), (00), (01), \ldots, (000)\} = X^0 \cup \left\{x \in \cup_{j=1}^{t-1} X^j : \sum_{j=1}^{t} x_j = 0\right\} \\
H_2 = H \setminus H_1
\end{cases}
\]
and the vector of parameters \((\pi_f, \pi_r)\) is represented in our notation as \(p_{H2}(M_6) = (p_{H1}, p_{B2})\).

Farcomeni (2011) shows that many other models proposed in the literature can be recovered as special cases of the model with saturated parameterization \(p\) subject to specific linear constraints. In the following, we prefer to index parameters with the partition notation and we refer to the reduced parametrization with \(\bar{p}_{H}(M_6) = (\bar{p}_{H1}, \cdots, \bar{p}_{B2})\) corresponding to the uniquely identified conditional probabilities associated to the partition \(H_2\). In Supplementary Web Material SWM-3, other partitions are illustrated with emphasis on Markov models of orders 1 and 2. Unfortunately, the number of all possible reduced models exponentially grows with \(t\), namely with the Bell number of \(2^t - 1\), with more than \(10^{25}\) alternatives when \(t = 5\). Most of the models resulting from all possible partitions are meaningless but one may find interesting behavioral patterns when there is a rationale to endow partial capture histories with a total order and group conditional probabilities in a meaningful way. In Section 3, we introduce alternative summaries quantifying the accumulation of trapping experience and exploit them as explanatory variables in a GLM framework yielding new parsimonious behavioral models. These summaries can be used as a tool for defining and exploring partition models within the huge collection of possibilities described above.

3. Regression Approach Based on Memory-Related Summaries

Similarly to Huggins (1989) and Alho (1990), we consider a logistic regression model viewing each capture occurrence of unit \(i\) at occasion \(j\) as a binary outcome whose probability can be modeled as a function of one or more explanatory variables. Our idea is to build up summary statistics regarded as time-varying behavioral explanatory variables. Formally this can be embedded in a general regression framework as follows

\[
\text{logit}(p(x_{1i}, \ldots, x_{ij-1})) = r(z_{ij}, v_{ij}; \theta) \quad (3)
\]

where \(\theta\) is a generic vector of unknown parameters, \(z_{ij} = (z_{ij}^{(1)}, \ldots, z_{ij}^{(m)})\) is a vector of time-varying behavioral explanatory variables, and \(v_{ij}\) denotes a vector of other available explanatory variables which may be time-dependent and/or unit-specific. In this way, we might incorporate other available individual covariates such as sex or weight as well as time-dependent numerical variables or factors. We start with a simple linear regression with a single \((m = 1)\) memory-related summary \(z_{ij}\) and no available individual covariates \(v_{ij}\)

\[
\text{logit}(p(x_{1i}, \ldots, x_{ij-1})) = r(z_{ij}; \alpha, \beta) = \alpha + \beta z_{ij} \quad \forall i = 1, \ldots, N \quad \forall j = 1, \ldots, t \quad (4)
\]

To simplify the notation, the unit index \(i\) will be omitted in the following when it is not needed. Recall that a partial capture history \(x = (x_1, \ldots, x_j)\) is a binary string taking values in \(H = \bigcup_{1 \leq r < l} \mathbb{X}^r\) and has a length \(l_x = r\) which can take values in \(0, 1, \ldots, t - 1\).

We will first provide a rationale to define a memory-related summary by means of parametric classes of transformations and we then focus on few specific instances. According to the natural and intuitive interpretation of the accumulation of trapping experience and the fact that occurrence of trapping in the last occasions can have a different (probably greater) impact on the future capture probability than those occurred in the previous ones, one can grade a behavioral effect according to a weighted sum of previous captures as follows:

\[
q_x(x) = q_x(x_1, \ldots, x_n) = \sum_{j=1}^{l_x} w_j(\lambda, l_x)x_j \quad (5)
\]

where we assume that the partial capture history has length \(l_x \geq 1\) and \(w_j(\lambda, l_x) > 0\) is a positive weight system which may possibly depend on suitable parameters \(\lambda\) and the history length \(l_x\). We will consider weight system constructions which rely on the following one-parameter weights

\[
w_j(\lambda, l_x) = w_j(\lambda) = \lambda^{j-1} \quad (6)
\]

When \(\lambda = 1\), we get a very natural summary \(f_1(x) = \sum_{j=1}^{l_x} x_j\) which simply counts the number of previous captures. Values \(\lambda > 1\) can be used to increase the impact of the most recent captures on the memory effect summary. However, one can find it unsuitable that the range of the summary statistics varies with the length of the partial history and/or that the maximum value of the weights is larger than 1 and depends on \(l_x\). One can overcome this problem by rescaling the basic weight \(w_j(\lambda, l_x)\) \((j = 1, \ldots, l_x)\) so that the rescaled \(\bar{w}_j(\lambda, l_x)\) sums up to 1. In this way, the proxy for a memory effect can be regarded as a weighted average of the past experience and hence it is always within the interval \([0, 1]\). When we use the previous one-parameter weight \(w_j(\lambda, l_x) = \lambda^{j-1}\), the corresponding weighted average summary statistics will be denoted by

\[
g_x(x) = \sum_{j=1}^{l_x} \bar{w}_j(\lambda, l_x)x_j = \sum_{j=1}^{l_x} \frac{w_j(\lambda)}{\sum_{h=1}^{l_x} w_h(\lambda)} x_j. \quad (7)
\]

We emphasize that, unlike the memory-related summary \(f_1(x)\) in (6), the notation \(g_1(x)\) is used for a weight system \(\bar{w}_j(\lambda, l_x)\) \((j = 1, \ldots, l_x)\) which has been rescaled to have unit sum. A referee pointed out the similarity of the construction of \(g_1(x)\) and the exponential smoothing technique in time series.

Conventionally, for \(x = ()\) we set the generic \(q_x(x) = 0\) and, similarly, this holds also for (6) and (7). In this work, we will focus only on the following instances of memory-related summaries:

\[
f_1(x), \quad g_1(x), \quad \text{and} \quad g_2(x).
\]

In fact, both \(f_1(x)\) and \(g_1(x)\) have a very natural interpretation since they represent, respectively, the total number and the relative frequency of the previous capture occurrences. This implies that the most recent experience has the same relative weight as the least recent. This may be considered more plausible in those capture–recapture contexts where the time elapsed between consecutive trapping occasions is shorter and the overall recapture experiment has limited duration. On the other hand, one can
Quantifications $f_1(x)$, $g_1(x)$, and $g_2(x)$ of all partial capture histories corresponding to the complete individual capture history $(0, 0, 1, 0, 0, 1, 1, 0, 1, 0)$ in a capture–recapture experiment with $t = 10$ capture occasions. Recall that at time $j = 1$, there is no previous capture history and that the corresponding partial capture history is denoted with $()$.

| Time | Current occurrence $x_{ij}$ | Partial capture history $x = (x_1, \ldots, x_{ij-1})$ | Numeric summary $z_{ij} = f_1(x)$ | $z_{ij} = g_1(x)$ | $z_{ij} = g_2(x)$ |
|------|-----------------------------|------------------------------------------------|-----------------------------|-----------------------------|-----------------------------|
| 1    | 0                           | ()                                            | 0                           | 0.000                       | 0.000                       |
| 2    | 0                           | (0)                                          | 0                           | 0.000 $= 0/1$              | 0.000 $= 0/1$              |
| 3    | 1                           | (0, 0)                                        | 0                           | 0.000 $= 0/2$              | 0.000 $= 0/3$              |
| 4    | 0                           | (0, 0, 1)                                     | 1                           | 0.333 $= 1/3$              | 0.571 $= 4/7$              |
| 5    | 0                           | (0, 0, 1, 0)                                  | 1                           | 0.250 $= 1/4$              | 0.267 $= 4/15$             |
| 6    | 1                           | (0, 0, 1, 0, 0)                               | 1                           | 0.200 $= 1/5$              | 0.129 $= 4/31$             |
| 7    | 1                           | (0, 0, 1, 0, 0, 1)                            | 2                           | 0.300 $= 2/6$              | 0.571 $= 36/63$            |
| 8    | 0                           | (0, 0, 1, 0, 0, 1, 1)                         | 3                           | 0.429 $= 3/7$              | 0.787 $= 100/127$          |
| 9    | 0                           | (0, 0, 1, 0, 0, 1, 1, 1)                      | 3                           | 0.375 $= 3/8$              | 0.392 $= 100/255$          |
| 10   | 1                           | (0, 0, 1, 0, 0, 1, 1, 0, 0)                   | 3                           | 0.333 $= 3/9$              | 0.196 $= 100/511$          |

for $j = 1, \ldots, t$. It is clear from these examples that, whereas $g_1(x)$ and $g_2(x)$ generate the same capture probabilities in any given partial capture history, the first quantification $f_1(x)$ is a strict subset of the last two, $g_1(x)$ and $g_2(x)$, which codify the same information. It is also clear from the above examples that the probability of never being captured during the whole experiment is $P_0 = \left(1 - \frac{e^\alpha}{1 + e^\alpha}\right)^t$ and depends only on one parameter as model $M_b$. In fact, in the partition model framework of Alunni Fregatali and Tardella (2013) highlighted that all behavioral models for which the first equivalence class $H_1$ is composed exclusively by all partial capture histories with no capture contain the same unconditional likelihood factor involving only $N$ and $p_0$ and they yield the same estimates for $N$. Note that, unlike $M_b$, in the linear logistic models (4) also the recapture probabilities depend on $\alpha$ and this is why the new models end up with different estimates of $P_0$ and $N$. For further insights see SWM-7.

We remark that the saturated conditional probability parameterization can be derived within the marginal model framework proposed in Bartolucci and Forcina (2006), since one can get conditional distributions of the $j$-th binary outcome as a linear function of interaction parameters for the marginal distribution of the previous $j - 1$ binary outcomes (see Bartolucci et al., 2007), Section 2.4). However, in our opinion, when patterns of conditional probabilities are sought, the transition model perspective is more intuitive and direct.

3.1. Covariate Representation and Markovian Structure

In this subsection, we go back to the topic of building behavioral models based on interesting partitions of the subset $H$ as in Section 2. We show how the summary $z$ can also be used to order partial capture histories and derive partitions of $H$ accordingly. Moreover, we explain how, from the ordering relying on the specific $z = g_2(x)$, one can recover partitions corresponding to Markov models. In fact, if we fix a positive integer $k < t$ we can partition the set $H$ of all partial capture histories according to the value of $g_2(x)$ into appropriate subintervals namely $I_i = \left[\left(h, \frac{h}{2}\right]\right], \ldots, I_{2^k - 1} = \left[\frac{2^{k-1}}{2}, 1\right)$. Hence, we get the partition $\mathcal{H} = \{H_1, \ldots, H_{2^k}\}$ where

$$x \in H_r \Leftrightarrow z = g_2(x) \in I_r \quad \forall r \in \{1, \ldots, 2^k\} \quad (9)$$

so that the equivalence classes of binary subsequences depend only on the last $k$ binary events. Indeed, one can prove that
the mapping \( g_2 \) defined in (8) is such that, for each \( x \in H \), the set \( I_\lambda \), to which \( z = g_2(x) \) belongs is determined by the last \( k \) digits of \( x \). Hence, the definition of equivalence classes of conditional probabilities given partial capture histories in these partitions yields the Markov property of order \( k \). The formal proof is provided in SWM-§2 with further details on the appropriate correspondence of partitions \( H_1, \ldots, H_M \) with \( I_1, \ldots, I_A \) deriving from (9) also for partial capture histories with less than \( k \) digits.

We highlight that the partition defined in (9) is also equivalent to considering the general logistic regression as in (3) where the regression function \( r(z) \) is nonlinear and corresponds to a step function as follows \( s(z) = \log(p_{H_\lambda}) \) for all \( z \in I_\lambda \) and \( r = 1, \ldots, 2^k \), where \( p_{H_\lambda} = P(X_{ij} = 1|x_{11}, \ldots, x_{ij-1}) \) for any \( (x_{11}, \ldots, x_{ij-1}) \in H_\lambda \) according to the notation used in Section 2. In the logistic regression setup, this is equivalent to converting the numerical covariate into a categorical factor according to which subinterval \( I_a \) the covariate falls in. More details and examples of the correspondence are included in SWM-§3.

3.2. Alternative Partitioning and Construction of Memory-Related Summaries

More generally, in the logistic regression perspective we believe that an alternative partition of the range of \( z \) into \( A \) consecutive subintervals \( I_1 = [0, \epsilon_1], \ldots, I_A = (\epsilon_{A-1}, 1] \) can represent an alternative meaningful behavioral model corresponding to the regression step function

\[
s(z) = \log(p_{H_\lambda}) \quad \forall z \in I_\lambda \quad a = 1, \ldots, A.
\]

In fact, one can prove that, for suitable fixed choices of the number and position of the cutoffpoints different from those represented in the intervals \( I_\lambda = \left[\frac{\epsilon_1}{2}, \frac{1}{2}\right) \cup \left[\frac{1}{2}, \frac{2-\epsilon_1}{2}\right) \cup \left[\frac{2-\epsilon_1}{2}, 1\right] \), one can recover the classical behavioral model \( M_0 \), the more recent \( M_{A\lambda} \) proposed in Farcomeni (2011), and many others (see SWM-§4). Moreover, whenever the partition \( \{I_1, \ldots, I_A\} \) is a partition coarser than the Markovian partition \( \{[0, \frac{1}{2}), \ldots, (\frac{2^{k-1}-1}{2}, \frac{1}{2})\} \), this would be a more parsimonious-constrained Markov model of order \( k \) where some conditional probabilities associated to subintervals adjacent with respect to the \( z \) ordering are restricted to be equal. We pursue this further in Section 5 and show how these particularly flexible instances of partitioning looking for optimal data-dependent choice of the cutoffs \( \epsilon_1, \ldots, \epsilon_{A-1} \) yield interesting parsimonious models and represent an alternative perspective for exploring a range of meaningful variable order Markov chain models or approximation thereof. Note that when we apply the same strategy of partitioning with different memory-related summaries, we end up with different partitions. In our applications, we have implemented this idea with alternative memory-related summaries such as \( z = g_1(x) \) and \( z = g_2(x) \) improving sometimes model fit.

As possible alternative way of constructing useful mappings, one can consider a different rescaling of the basic weight system so that the maximum rescaled weight is one and corresponds to the last occasion. One can define \( z = \sum_{j=1}^k w_j(z) \), which in the case of \( w_j(z) = \lambda^{j-1} \) with \( \lambda > 1 \) would yield another possibility of grading the accumulation of memory with an alternative decreasing impact of the less recent occasions \( h_j(x) = \lambda^{j-1} \).

As already mentioned, the most critical parameter for the estimation of \( N \) is the probability \( p_0 \) as in (1). Indeed, when we partition the set of conditional probabilities through partitioning the generic summary \( z \) into intervals \( I_1, \ldots, I_A \) we have that, as long as \( z \in I_1 = [0, \epsilon_1] \) (and this is certainly true for all histories with no capture) we get \( p_j(0, \ldots, 0) = p_{H_\lambda} \) for all \( j \) so that the fundamental probability \( P_0 = \int \prod_{l=2}(1 - p_j(0, \ldots, 0)) \) depends only on the first component of the parameter vector \( \theta_{H_\lambda} \).

4. Unconditional Maximum Likelihood Inference

In this section, we show how exploiting memory-related summaries yields a simple-to-implement inference based on unconditional likelihood when there are no available covariates \( x_j \). Indeed, one can basically recycle consolidated standard GLM routines in our capture–recapture context. Let \( L(N, \theta) \) be the likelihood function for the regression model (3)

\[
L(N, \theta) \propto \prod_{l=1}^{N} \left[ \frac{\exp(r(z_i; \theta))}{1 + \exp(r(z_i; \theta))} \right]^{x_i} \prod_{l=1}^{N} \left[ 1 - \frac{\exp(r(z_i; \theta))}{1 + \exp(r(z_i; \theta))} \right]^{1-x_i}.
\]

In order to make inference on \( N \), one can first look at \( L(N, \theta) \) as a function of \( \theta \) only for a fixed value of \( N \). Let us denote with \( \hat{L}(N) = L(N, \hat{N}(N)) \) the maximum likelihood w.r.t. \( \theta \) obtained as a result of a standard logistic model fitted by using \( N \times t \) binary observations \( x_{ij} \) with their corresponding numerical covariates \( z_{ij} \). Unconditional maximum likelihood estimate for \( N \) will then be \( \hat{N} = \arg\max_{N \in \mathbb{N}}(\prod_{l=1}^{M} L(N)) \), where \( N_{upp} \) is a suitably high-fixed upperbound for the population size. The joint unconditional likelihood for all parameters involved in the model is globally maximized at the UMLE value \( (\hat{N}, \hat{\theta}) \). Hence, the estimating procedure requires iterative fitting of the logistic regression for each \( N \in \{M, \ldots, N_{upp}\} \).

Standard GLM-mixed models (GLMM) routines allow to fit more flexible models incorporating unobserved heterogeneity as in latent trait analysis. In fact, similarly to the pioneering work of Coull and Agresti (1999) one can add, on a logit scale, an individual random effect \( U_i \sim N(0, \sigma^2_U) \) as follows: \( \logit(p_j(x_{ij1}, \ldots, x_{ij(t-1)})) = r(z_{ij}; \theta) + U_i \).

Alternatively, one can account for unobserved heterogeneity in terms of latent classes providing possibly more efficient fitting routines as in Atkin (1996), Einbeck et al. (2014), Bartolucci and Forcina (2006) and Grün and Leisch
Great Copper Butterfly data: point and interval estimates together with AIC index of alternative fitted models.

Confidence intervals at level 1 − α = 0.95. Linear logistic models as in (4) with \( z = g_2(x) \) is denoted with \( M_{g2} \); with \( z = g_1(x) \) is denoted with \( M_{g1} \); with \( z = f_1(x) \) is denoted with \( M_{f1} \). Model \( M_{g1b} \) is \( k \)-th order Markov model with a specific first capture probability which differs from the re-capture probability conditioned on the absence of capture in the last \( k \) occasions. Model \( M_{g2 \text{ cut}(j)} \) denotes the model corresponding to a step-function regression with respect to the summary \( z = g_2(x) \) with \( j \) jumps located by an optimal search.

| Model          | # Parameters | \( \hat{N} \) (\( N^- \), \( N^+ \)) | AIC |
|----------------|-------------|---------------------------------|-----|
| \( M_{g2} \)   | 2+1         | 170 (87,448)                    | 321.46 |
| \( M_{g2 \text{ cut}(2)} \) | 5+1         | 198 (85,1039)                   | 324.00 |
| \( M_{g2 \text{ cut}(3)} \) | 7+1         | 62 (48,223)                     | 324.06 |
| \( M_{g2b} \)  | 5+1         | 62 (48,223)                     | 325.46 |
| \( M_{f1} \)   | 2+1         | 154 (82,367)                    | 325.99 |
| \( M_{g2} \)   | 2+1         | 90 (63,152)                     | 326.01 |
| \( M_{g2 \text{ cut}(1)} \) | 3+1         | 91 (64,154)                     | 326.41 |
| \( M_{g2} \)   | 4+1         | 176 (78,896)                    | 327.20 |
| \( M_{g2} \)   | 6+1         | 96 (64,184)                     | 330.92 |
| \( M_{g2} \)   | 2+1         | 97 (64,181)                     | 330.93 |
| \( M_{g2b} \)  | 3+1         | 62 (48,223)                     | 331.24 |
| \( M_{g2} \)   | 3+1         | 95 (64,184)                     | 332.84 |
| \( M_{f1} \)   | 2+1         | 96 (62,184)                     | 338.30 |
| \( M_{g2} \)   | 1+1         | 64 (53,85)                      | 342.80 |
| \( M_{g2} \)   | 2+1         | 62 (48,223)                     | 344.77 |
| \( M_{g2} \)   | 8+1         | 64 (52,84)                      | 352.85 |

(2008). Latent trait and latent class analysis are two commonly used latent variable models for categorical data and an updated overview with comparative features can be found in Bartholomew et al. (2011).

5. Examples

5.1. Great Copper Butterflies

We consider the Great Copper data originally analyzed in Ramsey and Severns (2010) and also reviewed in Farcomeni (2011) and Alunni Fegatelli and Tardella (2013). There are \( t = 8 \) capture occasions and \( M = 45 \) observed butterflies. Ramsey and Severns (2010) explain that butterflies tend to congregate near favorable habitat, which is readily recognized by the observer and this may yield a persistence related to the characteristics of the subject animals, the environment, and/or the observational pattern. In their first attempt to model and understand the dependence structure in these data, they show that there can be a great impact of the modeled temporal dependence and behavioral pattern on the final estimates with possible large uncertainty on the magnitude of the population size. In fact, Farcomeni (2011) provided evidence of alternative ephemeral effects which correspond to possibly larger population estimates. We now show the ability of our approach to improve model fit and gain an alternative simple parsimonious understanding of the dependence pattern. From the maximization of the unconditional likelihood, we get the results displayed in Table 2 where models are listed according to the increasing value of the AIC. We have considered the linear logistic model in (4) with \( z = g_2(x) \) denoted with \( M_{g2} \) and, with similar notation, linear logistic models \( M_{g1} \) and \( M_{f1} \). Besides standard models in the Otis class such as \( M_o \), \( M_t \), or \( M_b \), we have considered alternative benchmark competing models which account for behavioral effects such as those considered in Yang and Chao (2005) (denoted with \( M_{g1}, M_{g1b} \)) Ramsey and Severns (2010) (denoted with \( M_{g1}, M_{g2b} \)) and Farcomeni (2011) (denoted with \( M_{g2b}, M_{g2} \), and \( M_{g2} \)). For further details on these models see SWM-G. From Table 2, it is apparent that the use of \( z = g_2(x) \) allows for a sensible improvement of the AIC which is, however, accompanied by a larger point estimate and width of the confidence interval. More precisely, model \( M_{g2} \) yields \( \hat{N} = 170 \), with \( \hat{a} = -3.243 \) and \( \hat{\beta} = 3.179 \). A significantly positive \( \hat{\beta} \) (\( p \)-value < \( 10^{-10} \)) highlights an initial trap-happiness effect which gradually decreases when the memory-effect covariate \( z = g_2(x) \) decreases.

We have also implemented the idea of partitioning the range of the memory-related summary \( z = g_2(x) \) in order to get further improvement and understanding. In fact, one can see from Table 2 that we could not get a better fit looking for an appropriate number of optimal cutpoints (ranging from 1 to 3). However, these models allow us to discuss alternative models based on the same memory-related summary looking at the different shape of the regression function (Figure 1) and possibly appreciating deviations from linearity and monotonicity. More details are reported in SWM-G where the grid search of optimal cutpoints is illustrated. Notice that \( M_{g2 \text{ cut}(2)} \) and \( M_{g2 \text{ cut}(3)} \) are among the best fitting models. The shape of their regression functions looks similar and in agreement with the pattern corresponding to the regression curve fitted with \( M_{g2} \) with an initial trap-happiness response. If we focus on \( M_{g2 \text{ cut}(2)} \), we have trap-happiness since \( p_{H1} < p_{H5} \) and we still have a decreasing discontinuous pattern of recapture probabilities \( (p_{H5} > p_{H2} > p_{H1}) \) with the decreasing memory effect corresponding to the vanishing value of \( z \). If one looks at the step function corresponding to \( M_{g2 \text{ cut}(3)} \) in Figure 1, we point out that the first interval \( I_1 \) determines the same subset \( H_1 \) of the partition \( H_2(M_0) \) in (2). As already argued in Section 3 even if the regression functions provide a similar overall pattern (yielding a similar AIC), it is their different local behavior at \( z = 0 \) which provides different values of \( \hat{N} \). This depends on \( a \) in the linear regression case and on \( p_{H1} \) in the step-function case. Finally, note the awkward pattern resulting from \( M_{g2 \text{ cut}(4)} \). We feel that similar wildly fluctuating regression functions should be disregarded from the final analysis and this is why we have not reported the resulting estimates in Table 2. The somewhat remarkable difference among the estimates of the best fitting models may suggest taking into account model uncertainty and providing a model-averaged estimate \( \hat{N}_{MA} \) based on the Akaike weights following Burnham and Anderson (2002, Chapter 4) which leads to \( \hat{N}_{MA} = 90 \) and a wider confidence interval [45, 380].

These results show that, although the enduring effect of the classical behavioral model \( M_b \) yields one of the worst fitting models, a novel more gradual behavioral effect is highlighted in \( M_{g2} \) by smoothly modeling the subsequent changes in the longitudinal pattern of capture probabilities after the first capture in terms of a suitable memory-related summary. This model could confirm a kind of persistence effect conjec-
Flexible Behavioral Capture–Recapture Modeling

5.2. Giant Day Geckos

Capture–recapture sampling on the giant day gecko has been conducted in the Masoala rainforest exhibit at the Zurich Zoo and data have been analyzed in Wanger et al. (2009). It is an interesting dataset where a large number of capture occasions (t = 30) are available and the closed population assumption is valid since it is a captive population. More details on the sampling process can be found in Wanger et al. (2009).

We are interested in analyzing behavioral patterns possibly originated by feeding habits of the geckos and/or by the human presence. We have fitted the same standard models as well as new ones as in the previous example. We have also repeated the analysis with the addition of an individual Gaussian random effect to see whether there might be some other source of unobserved heterogeneity which could affect the conditional probabilities. Results are displayed in Table 3. In this dataset, the AIC highlights a different kind of behavioral response based on the logistic model (4) with summary \( z = g_2(x) \) as in (7). If we focus on model \( M_{g2} \) (AIC = 1126.4), we find that \( M_{g2,b} \) accounting for unobserved heterogeneity in terms of a Gaussian random effect achieves a much better fit (AIC = 1120.3).

On the other hand, there is more evidence in favor of a nonlinear regression in terms of the same covariate with the best fitting model \( M_{g2,cut(3)} \) which is based on an increasing step function. In Table 3, we reported also the results of the model \( M_{g2,cut(3)}b \) with the same step function regression and the inclusion of a Gaussian random effect which in this case does not improve the fit. A graphical comparison of the best fitting models based on \( g_3(x) \) is in Figure 2. The regression function corresponding to \( M_{g2,a} \) represents an average individual regression with a lower mean intercept and a reduced
models with unobserved heterogeneity are in SWM-based on the behavioral covariates. Motivated by the real data comparing some classical behavioral models and new models In this section, we present a simulation study aimed at com-

A Simulation Study for Model Selection

6. A Simulation Study for Model Selection

Table 3

Giant Day Gecko data: point and interval estimates together with AIC index of alternative fitted models. Confidence intervals at level $1 - \alpha = 0.95$. Linear logistic models as in (4) with $z = g_2(x)$ is denoted with $M_{g_2}$; with $z = g_1(x)$ is denoted with $M_{g_1}$; with $z = f_1(x)$ is denoted with $M_{f_1}$. Model $M_{i+1}$ is $k$-th order Markov model with a specific first capture probability which differs from the re-capture probability conditioned on the absence of capture in the last $k$ occasions.

Model $M_{i+1, cut(j)}$ denotes the model corresponding to a step-function regression with respect to the summary $z = g_1(x)$ with $j$ jumps located by an optimal search. Models with an $h$ suffix in the subscript account for unobserved heterogeneity in terms of an additional Gaussian random effect.

| Model          | # Parameters | $\hat{N}$ (N−, N+) | AIC |
|----------------|--------------|----------------------|-----|
| $M_{g_1, cut(3)}$ | 7+1          | 105 (83,154)         | 114.76 |
| $M_{g_1, cut(2)}$ | 5+1          | 89 (77,108)          | 114.80 |
| $M_{g_1, cut(3), h}$ | 8+1          | 105 (83,161)         | 116.74 |
| $M_{g_1, cut(1)}$ | 3+1          | 89 (77,108)          | 116.81 |
| $M_{t, h}$ | 3+1          | 110 (84,206)         | 112.09 |
| $M_{h, b}$ | 2+1          | 490 (68,∞)           | 1124.31 |
| $M_{h, b}$ | 3+1          | 127 (89,322)         | 1125.21 |
| $M_{h, h}$ | 3+1          | 161 (96,1066)        | 1125.46 |
| $M_{h, b}$ | 2+1          | 134 (91,388)         | 1125.47 |
| $M_{b}$ | 3+1          | 171 (98,670)         | 1125.55 |
| $M_{b}$ | 2+1          | 86 (76,101)          | 1126.36 |
| $M_{b}$ | 2+1          | 87 (76,105)          | 1141.09 |
| $M_{g_2}$ | 2+1          | 80 (73,91)           | 1147.36 |
| $M_{g_2}$ | 5+1          | 107 (79,449)         | 1150.25 |
| $M_{g_2}$ | 3+1          | 107 (79,449)         | 1153.18 |
| $M_{g_2}$ | 4+1          | 79 (72,89)           | 1154.70 |
| $M_{g_2}$ | 2+1          | 107 (79,449)         | 1155.73 |
| $M_{p}$ | 3+1          | 76 (71,85)           | 1159.77 |
| $M_{p}$ | 2+1          | 76 (71,85)           | 1160.32 |
| $M_{p}$ | 2+1          | 75 (70,84)           | 1160.60 |
| $M_{p}$ | 3+1          | 76 (71,85)           | 1161.26 |
| $M_{p}$ | 30+1         | 74 (70,81)           | 1164.72 |
| $M_{p}$ | 1+1          | 74 (70,82)           | 1166.18 |

behavioral response if compared to $M_{g_1}$. More details on models with unobserved heterogeneity are in SWM.§8.

The comparative analysis of the two real-data examples provides some evidence of the improved understanding of the behavioral pattern. This can be seen in terms of the ability of our approach to detect the relevance of the past history in such a way that for some capture-recapture contexts (Great Copper data), the most recent trapping experience matters, relatively much more than the less recent one whereas in some other contexts (Giant Day Gecko data) the most recent past has, relatively, the same weight as the less recent occasions.

In this section, we present a simulation study aimed at comparing some classical behavioral models and new models based on the behavioral covariates. Motivated by the real data results, we decided to focus on models $M_{g_2}$ and $M_{h, b}$, which represent different aspects of the behavioral reaction to capture. We use each one as a model to generate data with the same parameter settings ($\alpha = -3$ and $\beta = 4$) considering different values for the population size and the number of occasions: $N \in \{100, 200\}$ and $t \in \{10, 20, 30\}$. Notice that, for fixed $t$, the probability, $P_0$, of never being observed, will be the same in both models and for fixed $N$, we get the same expected number $E[M]$ of distinct units captured at least once. For each setting described in Table 2 in SWM-§6, $K = 100$ dataset are generated and, for each generated dataset, we calculate point and interval estimates using the unconditional likelihood. The AIC is used to compare models. We have focussed the attention mainly on classical and more recent approaches which aim to model behavioral or longitudinal patterns. Table 3 in SWM-§6 lists all the alternative models considered and reports the empirical mean and the root mean square error (RMSE) of the alternative estimates of $N$, empirical coverage and average length of the interval estimates, and the percentage of times that each model achieves the lowest AIC. Table 3 in SWM-§6 shows that the estimates of $N$ from the true model almost always yield best results in terms of both point and interval estimates, the empirical mean $\hat{N}$ is very close to the real values of $N$ and the RMSE is almost always the smallest one. Only in Trial 1 and 7 do the true models not achieve the smallest RMSE. However, they are the only ones which guarantee a confidence interval coverage close to the nominal level. Finally, one can verify that the AIC allows to identify the true model most of the times, especially when $t$ increases. Following the suggestion of anonymous referees, we have investigated the case where the true model is a Markov model of order 1 and 2 obtaining similar results. Details are provided in SWM-§6. We have also extended our study simulating data from model $M_0$, with homogeneous capture probability, and the beta-binomial model in Dorazio and Royce (2003), denoted by $M_{h, h, b}$, corresponding to a purely unobserved heterogeneity model. The goal is to verify whether there could be a tendency for AIC to support more complex models even when there is no underlying behavioral effect in the data. We considered 12 simulation settings with the same grid of alternative values of $N$ and $t$ and with the remaining model specific parameters set so that the value of $P_0$ is comparable to the previous settings. Details are reported in Table 2 in SWM-§6. Interestingly, when $M_0$ is the true model, we have from Table 4 some evidence of AIC failing to detect the underlying structure of the true model with a tendency to support more complex models. However, to address this issue we have considered a possible adjustment relying on the pairwise model comparison in terms of likelihood ratio test between each of the alternative models and the true model, we have from Table 4 some evidence of AIC failing to detect the underlying structure of the true model with a tendency to support more complex models. However, following the suggestion of anonymous referees, we have investigated the case where the true model is a Markov model of order 1 and 2 obtaining similar results. Details are provided in SWM-§6. We have also extended our study simulating data from model $M_0$, with homogeneous capture probability, and the beta-binomial model in Dorazio and Royce (2003), denoted by $M_{h, h, b}$, corresponding to a purely unobserved heterogeneity model. The goal is to verify whether there could be a tendency for AIC to support more complex models even when there is no underlying behavioral effect in the data. We considered 12 simulation settings with the same grid of alternative values of $N$ and $t$ and with the remaining model specific parameters set so that the value of $P_0$ is comparable to the previous settings. Details are reported in Table 2 in SWM-§6. Interestingly, when $M_0$ is the true model, we have from Table 4 some evidence of AIC failing to detect the underlying structure of the true model with a tendency to support more complex models. However, to address this issue we have considered a possible adjustment relying on the pairwise model comparison in terms of likelihood ratio test between each of the alternative models and the nested model $M_0$. We have verified that this would give the appropriate false positive error control in each single test. Of course the best AIC criterion would also involve a multiplicity issue, hence we implemented a Bonferroni correction to the comparison between the more complex model selected with lowest AIC and $M_0$. In Table 7, one can see that this correction detected the true model consistently as long as $t$ increases. Overall, the conclusions about the relevance of the behavioral models $M_{g_2}$ and $M_{h, b}$ in the real data examples in Section 5 are strengthened, since the likelihood ratio test with the Bonferroni correction is verified in both cases. However,
Flexible Behavioral Capture–Recapture Modeling

133

Figure 2. Comparison of regression functions underlying best fitting models in the Giant Day Geckos data example.

we note that when $t$ is low the percentage of times that the underlying $M_{hbb}$ model is selected can be a little low.

7. Concluding Remarks and Discussion
The main contribution of this article is the idea of suitably summarizing numerically some relevant features of partial capture histories and exploit these summaries as explanatory variables in a general transition model regression with the aim of modeling and understanding behavioral effects possibly decreasing or reinforcing during the course of the capture sessions. One can interpret such summaries as a proxy for a memory effect and a correspondence with the Markov dependence has been established for a specific summary. Standard behavioral component underlying $M_b$ can be an oversimplified understanding of the enduring change while the alternative (or additional) ephemeral component underlying $M_{kx}$ could require a too rapid increase of parameters as $k$ grows to achieve sufficient flexibility. Setting up and exploiting memory-related summaries possibly parametrizing in terms of partitions thereof allowed us to revisit well-known datasets and discover new flexible and parsimonious behavioral patterns that fit the observed data better than other classes of models proposed in the literature. Unconditional likelihood inference has been easily implemented by recycling consolidated standard GLM routines. An integrated suite of R functions has been developed and is available in SWM. The discernibility of new patterns with respect to already available enduring or ephemeral behavioral effects has been verified with a simulation study where minimizing AIC yields the true pattern in most simulated data.

In the transition model perspective, the interpretation of the proposed models hinges on how the conditional probability of capture changes when the possibly rescaled accumulation effect underlying the memory-related summary changes and this can be visualized in terms of the shape of the regression function. Our transition model is suitable when the accumulation of trapping experience happens within a recapture setting with equally spaced times although the adjustment of weighting systems for irregularly spaced time points could be possible. It should not be used for other kinds of recapture data such as those in epidemiological studies where multiple lists cannot be considered as time-dependent occasions with a prescribed order. In those cases, the marginal model perspective can be preferred.

Allowing for a different functional form of the regression function can lead to an extended flexibility with a possibly more parsimonious model. Further flexibility is also gained by using alternative memory-related summary statistics derived from the same rationale of weighted accumulation of experience. The behavioral features summarized in $g_k(x)$ can be more appropriate for situations in which it is plausible that the memory of capture fades rather rapidly, although gradually. Whether this is realistic will depend on what “capture” means in the particular context, and on the temporal proximity of the sampling occasions. In the Great Copper examples, where the trapping sessions took place once every 3 days the summary $z = g_1(x)$ is selected as best fitting while in the daily Giant Day Gecko sessions the summary $z = g_1(x)$ highlighted a memory effect with a less pronounced fading.

There are possible extensions outside the closed capture–recapture context of the quantification idea. The first natural extension is the use of the behavioral covariate in open population models. Such an extension can be straightforward within a complete data likelihood framework such as the one
in Schofield and Barker (2011) where difficulties of previous approaches in dealing with individual time-varying covariates are also discussed; see also Catchpole et al. (2004). However, trap response in open population studies is likely to be more of ephemeral in nature and, in any case, less frequent since consecutive occasions are usually separated by long intervals, typically a year (Pradel and Sanz-Agualar, 2012).

Another issue which can be addressed within the proposed regression framework for a more thorough understanding of real data is the inclusion of individual covariates, possibly time-varying. However, individual covariates are typically not available for unobserved units. On one hand, this will make it difficult to embed this precious extra source of information in the proposed inferential setting which relies on the unconditional likelihood. On the other hand, it can be easily integrated in the alternative inferential setting based on the conditional likelihood. Indeed, in SWM we have provided examples of fitting models based on the proposed memory-related summaries in the presence of available individual covariates through the conditional likelihood approach. Models have been fitted in the conditional likelihood setup with the software Mark/RMark (White and Burnham, 1999; Laake, 2013).

The presence of individual covariates and the full power of the unconditional likelihood can be alternatively exploited through data augmentation within a Bayesian framework following the approach in Royle (2009). We will develop this in future work. Alternative inferential approaches for standard models have been investigated in Alunni Fegatelli and Tardella (2013) and suggest that the Bayesian approach could be more promising. Here, we have focussed more specifically on understanding the role, meaning, and alternative uses of the new memory-effect proxies and their connections with already available models.

8. Supplementary Material

Web Appendices, Tables, Figures, and R code for data analysis referenced in Sections 2, 3, 5, 6, and 7 are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

We thank three anonymous referees and the AE for insightful comments which allowed us to improve the original manuscript. Thanks also to Antonio Forcina for a stimulating discussion. This work has been supported by Sapienza Università di Roma, grant C26A125A5J.

References

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. Statistics and Computing 6, 251–262.

Alho, J. M. (1990). Logistic regression in capture-recapture models. Biometrics 46, 623–635.

Alunni Fegatelli, D. and Tardella, L. (2013). Improved inference on capture-recapture models with behavioural effects. Statistical Methods & Applications, 22, 45–66.

Amstrup, S. C., McDonald, T. L., and Manly, B. F. (eds) (2005). Handbook of Capture-Recapture Analysis. Princeton, New Jersey: Princeton University Press.

Bartholomew, D., Knott, M., and Moustaki, I. (2011) Latent Variable Models and Factor Analysis. 3rd edition. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons, Ltd., A unified approach.

Bartolucci, F. and Forcina, A. (2006). A class of latent marginal models for capture-recapture data with continuous covariates. Journal of the American Statistical Association 101, 786–794.

Bartolucci, F. and Pennoni, F. (2007). A class of latent Markov models for capture-recapture data allowing for time, heterogeneity, and behavior effects. Biometrics 63, 568–578.

Bartolucci, F., Colombi, R., and Forcina, A. (2007). An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints. Statistica Sinica 17, 691–711.

Böhning, D. (2008). Editorial - recent developments in capture-recapture methods and their applications. Biometrical Journal 50, 954–956.

Borchers, D., Buckland, S., and Zucchini, W. (2002). Estimating Animal Abundance: Closed Populations. Statistics for Biology and Health. London, United Kingdom: Springer-Verlag.

Burnham, K. P. and Anderson, D. R. (2002). Practical information-theoretic approach. Model Selection and Multimodel Inference. 2nd edition. New York: Springer-Verlag.

Catchpole, E. A., Morgan, B. J. T., and Coulson, T. (2004). Conditional methodology for individual case history data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 53, 123–131.

Chao, A., Chu, W., and Hsu, C.-H. (2000). Capture-recapture when time and behavioral response affect capture probabilities. Biometrics 56, 427–433.

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

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

Einbeck, J., Darnell, R., and Hinde, J. (2014). npmlreg: Nonparametric Maximum Likelihood Estimation for Random Effect Models. R package version 0.46-1.

Farcomeni, A. and Scacciatelli, D. (2013). Heterogeneity and behavioral response in continuous time capture-recapture, with application to street cannabis use in Italy. The Annals of Applied Statistics 7, 2293–2314.

Farcomeni, A. and Tardella, L. (2012). Identifiability and inferential issues in capture-recapture experiments with heterogeneous detection probabilities. Electronic Journal of Statistics 6, 2602–2626.

Farcomeni, A. (2011). Capture-recapture models under equality constraints for the conditional capture probabilities. Biometrika 98, 237–242.

Farcomeni, A. (2015). A general class of capture-recapture models based on the conditional capture probabilities. Biometrics (accepted) Article first published online: 10 SEP 2015 DOI: 10.1111/biom.12375.

Fitzmaurice, G. and Molenberghs, G. (2009). Advances in longitudinal data analysis: An historical perspective. In Longitudinal Data Analysis, pages 3–27. Chapman & Hall/CRC Handb. Mod. Stat. Methods. Boca Raton, Florida: CRC Press.

Ghosh, S. K. and Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects. Journal of Agricultural, Biological, and Environmental Statistics 10, 35–49.
Flexible Behavioral Capture–Recapture Modeling

Grün, B. and Leisch, F. (2008). Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software* 28, 1–35, 10.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* 76, 133–140.

Hwang, W.-H. and Huggins, R. (2011). A semiparametric model for a functional behavioural response to capture in capture-recapture experiments. *Australian & New Zealand Journal of Statistics* 53, 403–421.

Hwang, W.-H., Chao, A., and Yip, P. S. F. (2002). Continuous-time capture-recapture models with time variation and behavioural response. *Australian & New Zealand Journal of Statistics* 44, 41–54.

Laake, J. L. (2013). RMark: An R interface for analysis of capture-recapture data with MARK. AFSC Processed Rep. 2013-01, Seattle, Washington: Alaska Fisheries Science Center, NOAA, National Maritime Fisheries Services.

Lee, S.-M. and Chen, C. W. S. (1998). Bayesian inference of population size for behavioral response models. *Statistica Sinica* 8, 1233–1248.

Lee, S.-M., Hwang, W.-H., and Huang, L.-H. (2003). Bayes estimation of population size from capture-recapture models with time variation and behavior response. *Statistica Sinica* 13, 477–494.

McCrea, R. S. and Morgan, B. J. T. (2014) *Analysis of Capture-Recapture Data*. Chapman & Hall/CRC Interdisciplinary Statistics. Boca Raton, Florida: Taylor & Francis Group.

Pradel, R. and Sanz-Aguilar, A. (2012). Modeling trap-awareness and related phenomena in capture-recapture studies. *PLoS ONE* 7(3), e32666 03.

Ramsey, F. and Severns, P. (2010). Persistence models for mark-recapture. *Environmental and Ecological Statistics* 17, 97–109.

Royle, J., Chandler, R., Sollmann, R., and Gardner, B. (2013). *Spatial Capture–Recapture*. Amsterdam/Oxford/Waltham: Elsevier Science.

Royle, J. A. (2009). Analysis of capture-recapture models with individual covariates using data augmentation. *Biometrics* 65, 267–274.

Sanathanan, L. (1972). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics* 43, 142–152.

Schofield, M. and Barker, R. (2011). Full open population capture-recapture models with individual covariates. *Journal of Agricultural, Biological, and Environmental Statistics* 16, 253–268.

Schwarz, C. and Seber, G. (1999). Estimating animal abundance: Review iii. *Statistical Science* 14, 427–456.

Seber, G. A. F. (1987). *Estimation of Animal Abundance (Second Edition)*. London, United Kingdom: Charles Griffin.

Wanger, T. C., Motzke, I., Furrer, S. C., Brook, B. W. and Gruber, B. (2009). How to monitor elusive lizards: Comparison of capture-recapture methods on giant day geckos (Gekkonidae, Phelsuma madagascariensis grandis) in the masoala rainforest exhibit, Zurich Zoo. *Ecological Research* 24, 345–353.

White, G. C. and Burnham, K. P. (1999). Program mark: Survival estimation from populations of marked animals. *Bird Study* 46(sup001), S120–S139.

White, G. C., Anderson, D. R., Burnham, K. P., and Otis, D. L. (1982). *Capture–Recapture and Removal Methods for Sampling Closed Populations*. Los Alamos, New Mexico: Los Alamos National Laboratory.

Yang, H.-C. and Chao, A. (2005). Modeling animals’ behavioral response by Markov chain models for capture-recapture experiments. *Biometrics* 61, 1010–1017.

Yip, P. S. F., Xi, L., Chao, A., and Hwang, W.-H. (2000). Estimating the population size with a behavioral response in capture-recapture experiment. *Environmental and Ecological Statistics* 7, 405–414.

Received January 2014. Revised July 2015. Accepted August 2015.