A General Framework for Modeling Memory in Capture–Recapture Data

L. ROUAN, R. CHOQUET, and R. PRADEL

In classical multisite capture–recapture (CR) models, the probability of moving to a new location depends only on the current site occupied. Yet, it is known that some species, such as Canada geese (*Branta canadensis*), have a strong tendency to return to sites visited earlier during their life. To account for this "phenomenon of memory," several authors have considered CR models in which transition probabilities depend not only on the current location of the individuals but also on the sites previously visited. In this article, we clarify the differences between these previous "memory" models and provide a general framework for the study of memory using CR data. We illustrate this study with the reanalysis of the movements of Canada geese among three wintering sites. This article has supplementary material online.

Key Words: Canada geese (*Branta canadensis*); Hidden Markov model; Hybrid symbolic-numeric method; Memory model; Multievent model; Parameter redundancy; Transients.

1. INTRODUCTION

If animals take any advantage of their past experience, their movements should be preferentially oriented along previously visited routes. Capture–recapture (CR) data, which track the locations of recognizable individuals over several occasions, allow one to investigate this phenomenon. With this type of data, when an animal is encountered, its current location among a set of several discrete sampling sites is recorded. Then, the application of specific models allows the estimation of a probability of movement between each pair of sites, even if the animals are not systematically reencountered (CR multistate models: Arnason 1973; Schwarz, Schweigert, and Arnason 1993). It has generally been found that animals tend to come back year after year to the same site, hence exhibiting memory, rather than settling randomly among the sites available. Whether memory extends beyond

L. Rouan is Research and Teaching Assistant (E-mail: *lauriane.rouan@cefe.cnrs.fr*), R. Choquet is Research Engineer, and R. Pradel is Research Scientist, Centre d'Écologie Fonctionnelle et Évolutive, Centre National de la Recherche Scientifique, Equipe Biométrie et Biologie des Populations, 1919 Route de Mende—34293 Montpellier Cedex 5, France.

one episode is a more challenging question and to address it requires both a large dataset and a more complex model. Hestbeck, Nichols, and Malecki (1991) were the first to investigate this extended phenomenon. They studied the choice of wintering sites made each year by 21,435 Canada geese along the North American Atlantic flyway, divided for this purpose into three regions, for the time period 1984–1989. They found that the geese wintering in each particular region (year t) tended to join at the next wintering season ($t + 1$) the region where they had wintered one year earlier ($t − 1$). A test for the detection of long-term memory in animal movements (Pradel, Wintrebert, and Gimenez 2003) has since confirmed this fact for both Canada geese and for the greater flamingo *Phoenicopteus roseus*, another species for which a large dataset was available (unpublished data). Other studies using different approaches (e.g. Ridout 1999) also concur that memory strongly influences the way animals move. Hence, models that describe extended memory will undoubtedly be needed more and more in the future.

The model of Hestbeck, Nichols, and Malecki (1991) was the first "two-step memory model" to appear; it has since been followed by two others (Brownie et al. 1993; Pradel 2005). The Brownie et al. model is in fact a straightforward generalization of that of Hestbeck et al. While Hestbeck et al. chose to deal with the simpler alternative: the previous site (time $t − 1$) is or is not the same as the site of arrival (time $t + 1$), Brownie et al. (1993) took into account the exact site previously occupied, thus considering a greater variety of transitions. As in Brownie et al. Pradel considered full dependency on the previous site. It is only in the treatment of the initial transition—that following the first time an animal is contacted—that the Brownie et al. and Pradel models differ. While Pradel did not distinguish the initial transition from subsequent transitions, Brownie et al. used specific parameters at this stage. These specific parameters do not depend on the previous (unknown) location. Although Brownie et al. gave no explanation for this last choice, this is clearly a way to avoid identifiability problems. The full implication of this choice is as yet unclear.

A main difficulty is that the Brownie et al. and Pradel models cannot be compared directly at the moment. Our aim, in this article, is to formally study these two models and to cast them in the most recent framework of CR modeling, the multievent formalization of CR models (Pradel 2005). In this way, alternative hypotheses about memory will become comparable (through model selection procedures). Section 2 introduces the parameters involved in the multievent formulation. Rewritten as multievent models, the models of Arnason–Schwarz, Brownie et al., and Hestbeck et al. will hereafter be called Models AS, B, and H, respectively, while the Pradel model will be denoted as Model P. These models have potentially many parameters and identifiability is an issue. Also, knowing how many parameters (or functions of the parameters) are estimable is crucial in the interpretation of estimates and model selection. In Section 3, we identify the redundant parameters in Models B and P using a symbolic-numeric version of the Catchpole–Morgan method for studying parameter redundancy (Catchpole and Morgan 1997). Section 4 clarifies the relationships between Models B and P, emphasizing their biological meaning. Finally, Section 5 revisits the study of the movements of the Canada geese among wintering sites through a set of biologically motivated models among which Models B and P, as well as AS are just some possible alternatives.

2. NOTATION AND MULTIEVENT FORMULATION OF THE MODELS

Let us consider a CR study with N sites and T capture occasions. We will distinguish the states, not directly observed, from the events (set Ω) coded in the capture histories (Pradel 2005). The events are what information is gathered about an individual when visiting the sites. Namely, it is either "not encountered" (code 0), or "encountered in Site 1" (code 1), or "encountered in Site 2" (code 2), etc.: $\Omega = \{0, 1, 2, \ldots, N\}$.

The states to be considered actually depend on the number of previous locations that are deemed relevant for the description of the transitions, that is, on the order of the Markov chain that governs these transitions. For instance, in the Arnason–Schwarz model, which is essentially a "one-step memory model," the transitions depend only on the site currently occupied; so the relevant states are the current location plus the "dead" state denoted $\dagger$, making up the set $E_0 = \{1, 2, \ldots, N, \dagger\}$. Unlike the event, the current state may remain unknown (if the animal is not encountered). In Models H, B, and P, which will concern us more directly in the following sections, transitions depend on both the current and the immediately previous site. The states to consider are thus the pairs of previous and current locations (see Figure 1). The state space is $E = \{11, 12, \ldots, NN, \dagger\}$. Here, the state will be known exactly only when the animal is encountered at both times $t - 1$ and t. (Note that for Model B, only the current location is relevant at the time of the first encounter. The state set can then be reduced to E_0.)

Having identified the event and state sets, it remains to specify the parameters of the memory Models AS, B, and P. Multievent models have three kinds of parameters (see Pradel 2005 for details):

1. Transition probabilities (shared with the multisite models)

   - in Model AS: ϕ^t_{jk}, probability of being present in site k at $t + 1$ for an animal seen in j at t;

   - in Model P: ϕ^t_{ijk}, probability of being present in site k at $t + 1$ for an animal that was in j at t and in i at $t - 1$; and
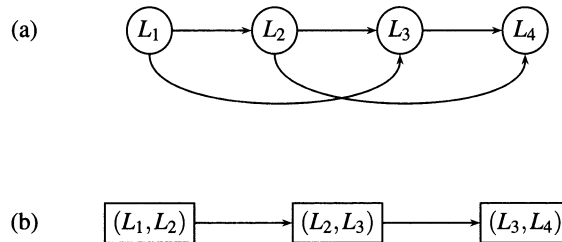

Figure 1. Memory modeling. In "two-step memory models" transitions between sites depend not only on the site currently occupied but also on the previous site occupied. These transitions can be modelled using: (a): a second-order Markov chain on the locations (L_t) (when the animal is observed, its location is known) or, as in multievent formalization, (b): a first-order Markov chain on the pairs of successive locations (when the animal is observed, its previous location may be unknown).

- in Model B: ϕ^t_{*jk}, probability of being present in site k at $t + 1$ for an animal seen for the first time in j at t and ϕ^t_{ijk}, probability of being present in site k at $t + 1$ for an animal that was in j at t and in i at $t - 1$ and was seen earlier than t.

2. Initial state probabilities

   - in Models AS and B: π^t_j, probability of being in site j when first captured at t; and

   - in Model P: π^t_{ij}, probability of being in site j at t and having been in site i at $t - 1$ when first captured at t.

3. Conditional event probabilities—this kind of parameter requires caution in its specification. Because the event is known and the state is what one wants to know, it is appealing to consider the probability of the state given the event. However, the probability appearing naturally in the model is the opposite in terms of conditioning:

   - $b^t_s(v)$, probability of event v (which here spans $\Omega = \{0, 1, 2, \ldots, N\}$) conditional on state s (which spans E or E_0 depending on the model).

For instance, if the states come from E_0, we will need among others $b^t_1(0)$, the probability of not being encountered when present on site 1; $b^t_†(0)$, the probability of not being encountered when dead; $b^t_1(1)$, the probability to be encountered on site 1 when present on site 1. If the states come from E, we will need $b^t_{21}(0)$, the probability of not being encountered at t when present on site 2 at $t - 1$ and on site 1 at t; $b^t_†(0)$, the probability of not being encountered when dead; $b^t_{21}(2)$, the probability of being encountered on site 2 at t when present on site 2 at $t - 1$ and on site 1 at t. Note that the last probability corresponds to an impossible event and is thus equal to zero. All of these parameters can actually be expressed as functions of the familiar encounter probabilities of the multisite models:

   - p^t_j, the probability of being encountered in site j at time t for an individual alive and in site j at that time.

Most of the $b^t_s(v)$ are equal to zero. The only nonnull parameters are: $b^t_j(j) = p^t_j$ and $b^t_j(0) = 1 - p^t_j$ for $j = 1, \ldots, N$, and $b^t_†(0) = 1$ when the state set is E_0; $b^t_{ij}(j) = p^t_j$ and $b^t_{ij}(0) = 1 - p^t_j$ for $i = 1, \ldots, N$ and $j = 1, \ldots, N$, and $b^t_†(0) = 1$ when the state set is E.

A further conditioning takes place at the time of marking (the first encounter). Because the animal is necessarily encountered then, the only remaining uncertainty is the site of capture. Thus, we use specific event probabilities at that time:

   - $b^{0t}_s(v)$, probability of event v conditional on state s *for an animal first captured* at t.

This probability is easily derived from the corresponding $b^t_s(v)$ by replacing the p with 1 (see also Table 1).

Table 1. Matrices of parameters for the Arnason–Schwarz, Brownie and Pradel models under the multievent formulation. For ease of readability, we have chosen here to consider only 3 sites (site 1: Mid-Atlantic, site 2: Chesapeake, site 3: Carolinas). In the event distribution matrices, the rows correspond to the events (starting from the first row: not seen, seen in site 1, seen in site 2, or seen in site 3 at time t) and the columns to the sites and "dead state" for the Model AS (from left to right: site 1, site 2, site 3, "dead") and to the couples of sites and "dead state" for Models B and P (from left to right: 11, 12, 13, 21, ..., 32, 33, "dead").

	Initial state	Transition	Event

Model AS

Initial state:
$$\Pi_t = \begin{pmatrix} \pi_1^t & \pi_2^t & \pi_3^t & 0 \end{pmatrix}$$

Transition:
$$\Phi_t = \begin{pmatrix} \phi_{11}^t & \phi_{12}^t & \phi_{13}^t & \phi_{1\dagger}^t \\ \phi_{21}^t & \phi_{22}^t & \phi_{23}^t & \phi_{1\dagger}^t \\ \phi_{31}^t & \phi_{32}^t & \phi_{33}^t & \phi_{1\dagger}^t \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Event:
$$B_t^0 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$B_t = \begin{pmatrix} 1-p_1^t & 1-p_2^t & 1-p_3^t & 1 \\ p_1^t & 0 & 0 & 0 \\ 0 & p_2^t & 0 & 0 \\ 0 & 0 & p_3^t & 0 \end{pmatrix}$$

Model B

Initial state:
$$\Pi_t = \begin{pmatrix} \pi_1^t & \pi_2^t & \pi_3^t & 0 \end{pmatrix}$$

Transition:
$$\Phi_t = \begin{pmatrix}
\phi_{*11}^t & \phi_{*12}^t & \phi_{*13}^t & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{*1\dagger}^t \\
0 & 0 & 0 & \phi_{*21}^t & \phi_{*22}^t & \phi_{*23}^t & 0 & 0 & 0 & \phi_{*2\dagger}^t \\
0 & 0 & 0 & 0 & 0 & 0 & \phi_{*31}^t & \phi_{*32}^t & \phi_{*33}^t & \phi_{*3\dagger}^t \\
\phi_{111}^t & \phi_{112}^t & \phi_{113}^t & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{11\dagger}^t \\
0 & 0 & 0 & \phi_{121}^t & \phi_{122}^t & \phi_{123}^t & 0 & 0 & 0 & \phi_{12\dagger}^t \\
0 & 0 & 0 & 0 & 0 & 0 & \phi_{131}^t & \phi_{132}^t & \phi_{133}^t & \phi_{13\dagger}^t \\
\phi_{211}^t & \phi_{212}^t & \phi_{213}^t & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{21\dagger}^t \\
0 & 0 & 0 & \phi_{221}^t & \phi_{222}^t & \phi_{223}^t & 0 & 0 & 0 & \phi_{22\dagger}^t \\
0 & 0 & 0 & 0 & 0 & 0 & \phi_{231}^t & \phi_{232}^t & \phi_{233}^t & \phi_{23\dagger}^t \\
\phi_{311}^t & \phi_{312}^t & \phi_{313}^t & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{31\dagger}^t \\
0 & 0 & 0 & \phi_{321}^t & \phi_{322}^t & \phi_{323}^t & 0 & 0 & 0 & \phi_{32\dagger}^t \\
0 & 0 & 0 & 0 & 0 & 0 & \phi_{331}^t & \phi_{332}^t & \phi_{333}^t & \phi_{33\dagger}^t \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Event:
$$B_t^0 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{pmatrix}$$

$$B_t = \begin{pmatrix}
1-p_1^t & 1-p_2^t & 1-p_3^t & 1-p_1^t & 1-p_2^t & 1-p_3^t & 1-p_1^t & 1-p_2^t & 1-p_3^t & 1 \\
p_1^t & 0 & 0 & p_1^t & 0 & 0 & p_1^t & 0 & 0 & 0 \\
0 & p_2^t & 0 & 0 & p_2^t & 0 & 0 & p_2^t & 0 & 0 \\
0 & 0 & p_3^t & 0 & 0 & p_3^t & 0 & 0 & p_3^t & 0
\end{pmatrix}$$

Table 1. (Continued.)

	Initial state	Transition	Event

Model P

$$\Pi_t = (\pi_{11}^t \quad \pi_{12}^t \quad \pi_{13}^t \quad \pi_{21}^t \quad \pi_{22}^t \quad \pi_{23}^t \quad \pi_{31}^t \quad \pi_{32}^t \quad \pi_{33}^t \quad 0)$$

$$\Phi_t = \begin{pmatrix}
\phi_{111}^t & \phi_{112}^t & \phi_{113}^t & \phi_{121}^t & \phi_{122}^t & \phi_{*23}^t & 0 & 0 & 0 & \phi_{11\dagger}^t \\
 & & & & & & & & & \phi_{12\dagger}^t \\
 & & & & & & & & & \phi_{13\dagger}^t \\
\phi_{211}^t & \phi_{212}^t & \phi_{213}^t & \phi_{221}^t & \phi_{222}^t & \phi_{223}^t & \phi_{131}^t & \phi_{132}^t & \phi_{133}^t & \phi_{21\dagger}^t \\
 & & & & & & & & & \phi_{22\dagger}^t \\
 & & & & & & \phi_{231}^t & \phi_{232}^t & \phi_{233}^t & \phi_{23\dagger}^t \\
\phi_{311}^t & \phi_{312}^t & \phi_{313}^t & \phi_{321}^t & \phi_{322}^t & \phi_{323}^t & \phi_{331}^t & \phi_{332}^t & \phi_{333}^t & \phi_{31\dagger}^t \\
 & & & & & & & & & \phi_{32\dagger}^t \\
 & & & & & & & & & \phi_{33\dagger}^t \\
 & & & & & & & & & 1
\end{pmatrix}$$

$$B_t^0 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}$$

$$B_t = \begin{pmatrix}
1-p_1^t & 1-p_2^t & 1-p_1^t & 1-p_3^t & 1-p_1^t \\
p_1^t & 0 & p_2^t & 0 & p_1^t \\
0 & p_2^t & 0 & p_3^t & 0 \\
1-p_3^t & 1-p_1^t & 1-p_2^t & 1-p_3^t & 1 \\
0 & p_1^t & 0 & p_2^t & 0 \\
p_3^t & 0 & p_3^t & 0 & 0
\end{pmatrix}$$

The parameters of the models are best handled using matrix notation and the ones that we will use are presented in Table 1.

Using these matrices, the probability of any encounter history h starting at time e can be written in compact form as

$$P(h) = \Pi_e \, \text{diag}(B_e^0(v_e, \cdot)) \left(\prod_{t=e+1}^{T} \Phi_{t-1} \, \text{diag}(B_t(v_t, \cdot)) \right) \mathbf{1}_N,$$

where v_t is the event observed at time t, $B(v_t, \cdot)$ is the row vector of B corresponding to event v_t, and $\mathbf{1}_N$ is the column vector of N ones. The likelihood is the product of the probabilities of all the encounter histories:

$$L = \prod_h \text{Pr}(h)^{\eta_h},$$

where η_h is the number of individuals whose capture history is h.

We note that Model B differs slightly from the memory model of Brownie et al. (1993) because, in multievent models, one conditions on the time of the first encounter, while Brownie et al. following the tradition of multistate models, conditioned on the time and state (see Table 1). For instance, the probability of the encounter history 121 is $\phi_{*12}^1 p_2^2 \phi_{121}^2 p_1^3$ for Brownie et al. (1993), but $\pi_1^1 \phi_{*12}^1 p_2^2 \phi_{121}^2 p_1^3$ in Model B. The difference lies in the presence or absence of the initial state probability. However, the multievent likelihood of Model B factorizes into two parts, one that is a function of the π's only and the other that is exactly the likelihood of Brownie et al. (1993) so that the estimates of the parameters common to the multistate and the multievent approaches are the same. Our Model B is thus a straightforward generalization of the model of Brownie et al. (1993). Under Model P, with two study sites, the probability of history 121 is more complicated as we have to go over the two different possible previous locations. It would be $(\pi_{11}^1 \phi_{112}^1 + \pi_{21}^1 \phi_{212}^1) p_2^2 \phi_{121}^2 p_1^3$.

In what follows, we will also use the following compound parameters:

- $\alpha_{ij}^t = \frac{\pi_{ij}^t}{\pi_j^t}$ with $\sum_{i=1}^{N} \alpha_{ij}^t = 1$: the conditional probability of having been in site i at $t-1$ for an animal seen for the first time at t in site j;

- $\beta_{ijk} = \phi_{ijk}^{T-1} p_k^T$: the probability for an animal in site i at $T-2$ and in site j at $T-1$ to be seen at T in site k. It has long been known in one-site, time-dependent CR models that only the product of the last survival by the last capture probability is estimable. This product has sometimes been denoted β (Burnham et al. 1987). The β_{ijk} parameter is the generalization of the β parameter to time-dependent, two-step memory models.

- $\beta_{*jk} = \phi_{*jk}^{T-1} p_k^T$: the probability for an animal seen for the first time at $T-1$ in site j to be seen at time T in site k.

3. PARAMETER REDUNDANCY

Identifiability is a current issue of time-dependent CR models for which formal tools have already been designed. In particular, Catchpole and Morgan (1997) developed a gen-

eral method for the study of parameter redundancy in models of the exponential family. This method, adapted to the multistate CR framework (Gimenez, Choquet, and Lebreton 2003), allows one successively to calculate the model rank, to identify the redundant parameters, and identifies estimable functions of the redundant parameters. However, as noted by Pradel (2005), its use for CR multievent models is restricted to the multinomial distribution of the individuals among the potential capture histories. With increasing numbers of time steps and sites, this multinomial has rapidly very many cells and only the first step of the method can be carried out symbolically in practice. This step consists of calculating the matrix $D(\theta)$ of the first derivatives of the capture history probabilities relative to the parameters.

$$D(\theta) = \left(\frac{\partial (\mu_i(\theta))}{\partial \theta_l} \right)_{\substack{1 \le i \le n_h \\ 1 \le l \le p}},$$

where $\mu_i(\theta)$ is the probability of the ith capture history, $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ is the vector of model parameters, and n_h is the number of possible capture histories.

The symbolic row rank of D is equal to the number of estimable parameter functions. Nonredundant parameters are detected by examining an eigenvector basis of the left null space of D: if parameter i is nonredundant, the ith elements of all the eigenvectors will be equal to zero.

As the rank and the eigenvectors could not be computed symbolically due to the large size of the matrix, we opted for a hybrid symbolic-numeric method (for examples of a hybrid method, see, for instance, Mourrain and Prieto 2000) in which the rank of D is calculated at n randomly selected points of the parameter range $\theta^{(k)}$. The formal rank r is such that

$$r \ge d = \max_{k \in \{1, 2, \ldots, n\}} \text{rank}\big(D(\theta^{(k)})\big).$$

In particular, if one of the numerical matrices is full-rank, there are no redundant parameters in the model.

Furthermore, for a full-rank model, the set of points for which $d < p$ is negligible (see Appendix A, page 352); the probability to draw n points (in our case $n = 10$) so that $d < p$ is thus zero. So, our reasoning was that, unless we are very unlucky, for a full-rank model, d will be equal to p.

In Catchpole and Morgan (1997), the determination of the estimable functions of parameters requires one to formally solve a system of partial differential equations. This step became impossible due to our transition from symbolic to numerical matrices. We, therefore, proceed differently. When we detect and identify redundant parameters, we examine the probability formulas to detect combinations of these redundant parameters. Using these combinations as new parameters, we reparameterize the model and apply again the procedure to ascertain that there are no further redundancy problems. We applied this procedure to Models B and P with 2 and 3 sites for successively 4, 5, and 6 time steps and identified ϕ_{*jk}^{T-1}, ϕ_{ijk}^{T-1}, and p_k^T as redundant parameters in Model B and ϕ_{ijk}^{1}, ϕ_{ijk}^{T-1}, p_k^T, π_{ij}^{1}, and π_{ij}^{T} as redundant parameters in Model P. We eventually obtained the following nonredundant

parameterizations:

for Model B:

$$\pi_i^t, \qquad i \in \{1, 2, \ldots, N-1\}, t \in \{1, 2, \ldots, T\};$$

$$\phi_{*jk}^t, \qquad j, k \in \{1, 2, \ldots, N\}, t \in \{1, 2, \ldots, T-2\};$$

$$\phi_{ijk}^t, \qquad i, j, k \in \{1, 2, \ldots, N\}, t \in \{2, 3, \ldots, T-2\};$$

$$p_i^t, \qquad i \in \{1, 2, \ldots, N\}, t \in \{2, \ldots, T-1\};$$

$$\beta_{ijk}, \qquad i, j, k \in \{1, 2, \ldots, N\};$$

$$\beta_{*ij}, \qquad i, j \in \{1, 2, \ldots, N\}.$$

for Model P:

$$\pi_i^t, \qquad i \in \{1, 2, \ldots, N-1\}, t \in \{1, 2, \ldots, T\};$$

$$\alpha_{ij}^t, \qquad i \in \{1, 2, \ldots, N-1\}, j \in \{1, 2, \ldots, N\}, t \in \{2, 3, \ldots, T-1\};$$

$$\phi_{*ij} = \phi_{*ij}^1, \qquad i, j \in \{1, 2, \ldots, N\};$$

$$\phi_{ijk}^t, \qquad i, j, k \in \{1, 2, \ldots, N\}, t \in \{2, 3, \ldots, T-2\};$$

$$p_i^t, \qquad i \in \{1, 2, \ldots, N\}, t \in \{2, \ldots, T-1\};$$

$$\beta_{ijk}, \qquad i, j, k \in \{1, 2, \ldots, N\}.$$

Although these results were established for $N = 2$ and 3 and $T = 4, 5$, and 6, we believe that they hold for any number of occasions and sites. We tried to establish this result using the expansion theorem of Catchpole and Morgan (2001), but have thus far obtained a formal demonstration only for Model B with three sites and any number of time steps (see Appendix B).

It has long been known in time-dependent CR models that only the product of a last transition by the encounter probability at the site of arrival is estimable. All of the redundancy in Model B and part of that in Model P is of this nature. We treated it by introducing the parameters β's and β_*'s of Section 2.

For Model P, the rest of the redundancy involves the parameters "$\{\pi_{ij}\}$." Problems with the parameterization of the initial period are known for Markov models in general and have already been studied in other statistical frameworks (see, e.g., Ron, Singer, and Tishby 1996). Here, the partial sums of the π_{ij}'s, namely the π_j's of Section 2, are always estimable. Following a suggestion by Pradel (2005), we replaced the π_{ij}'s with the π_i's and the α's of Section 2. A careful examination of the likelihood then led to the identification of the quantities $\sum_{i=1}^N \alpha_{ij}^1 \phi_{ijk}^1$ as estimable functions of the parameters. One such quantity is the average over the previous sites ($i = 1, \ldots, N$) of the transition probabilities from j to k between time 1 and time 2, ϕ_{ijk}^1. This is more precisely a weighted average where the weights are the probabilities that the newly encountered individual comes from each particular site in turn (α_{ij}^1). It is thus the unconditional probability that a newly caught individual moves from j to k between time 1 and 2, that is, the ϕ_{*jk}^1 of Model B.

4. MODEL B IS MORE GENERAL THAN MODEL P: THE MULTIEVENT MEMORY MODEL WITH TRANSIENTS

An original feature of Model B is the presence of transition probabilities specific to newly encountered individuals. We see from the list of identifiable parameters that it is indeed possible to estimate transition probabilities separately for the newly caught and the previously caught animals at all times before the last interval, and then separate β parameters. On the other hand, the interesting feature of Model P lies in the ability to estimate the likely previous location of a newly encountered individual (through the α_{ij}^t parameters)—and consequently the possibility to apply the transition that takes this previous location into account, ϕ_{ijk}^t. Again, this is possible at all but the first and last occasions. At occasion 1, Model P is actually indistinguishable from Model B because, after reparameterization, we retrieved the parameter ϕ_{*jk}^1 of Model B. Hence, at occasion 1 it appears that estimating specific transitions for newly caught individuals under the assumption that their past location is irrelevant (Model B) is equivalent to estimating the probability of their previous location under the assumption that they move like the previously caught individuals (Model P). However, occasion 1 is peculiar in that there are no previously caught individuals available. Are Model B and Model P equivalent at all occasions? A tally of the parameters shows that Model B has $(T - 2)N$ more parameters than Model P:

- $(T - 2)N^2 \phi_{*jk}^t$'s plus $N^2 \beta_{*jk}$'s for Model B;

- $(T - 2)N(N - 1)\alpha_{ij}^t$'s plus $N^2 \phi_{*jk}^1$'s for Model P.

They are thus not equivalent and, given its lower number of parameters, Model P could be nested within Model B.

One way to generalize Model P is by allowing transients among the newly caught, that is, individuals that are passing through the study area, are caught by chance and will never be seen again. This generalization of Model P (hereafter Model P*) is interesting to compare to Model B because, when models with transients were first introduced for one-site studies (Pradel et al. 1997), they were shown to be equivalent to models with specific initial (survival) probabilities, just as happens in Model B. Now, Model P* differs from Model P through the introduction of location (j)- and time (t)-specific proportions of transients τ_j^t. The matrices of initial transitions are thus to be modified as indicated below:

$$
\begin{pmatrix}
(1-\tau_1^t)\phi_{11}^t & 0 & \cdots & 0 & (1-\tau_1^t)\phi_{11\dagger}^t + \tau_1^t \\
0 & (1-\tau_2^t)\phi_{12}^t & & 0 & (1-\tau_2^t)\phi_{12\dagger}^t + \tau_2^t \\
\vdots & & \ddots & \vdots & \vdots \\
0 & & & (1-\tau_N^t)\phi_{1N}^t & (1-\tau_N^t)\phi_{1N\dagger}^t + \tau_N^t \\
\vdots & & & \vdots & \vdots \\
(1-\tau_1^t)\phi_{N1}^t & 0 & \cdots & 0 & (1-\tau_1^t)\phi_{N1\dagger}^t + \tau_1^t \\
0 & (1-\tau_2^t)\phi_{N2}^t & & 0 & (1-\tau_2^t)\phi_{N2\dagger}^t + \tau_2^t \\
\vdots & & \ddots & \vdots & \vdots \\
0 & & & (1-\tau_N^t)\phi_{NN}^t & (1-\tau_N^t)\phi_{NN\dagger}^t + \tau_N^t \\
0 & \cdots & \cdots & 0 & 1
\end{pmatrix}
$$

(for $i, j \in \{1, \ldots, N\}$, ϕ_{ij}^t is the row-vector ($\phi_{ij1}^t \quad \cdots \quad \phi_{ijN}^t$), and $\phi_{ij\dagger}^t = 1 - \sum_{k=1}^{N} \phi_{ijk}^t$).

The τ parameters are estimable at all occasions from $t = 2$ to $T - 1$. There are thus exactly $(T - 2)N$ identifiable τ's, that is, the difference in number of parameters between Model P and Model B. It can actually be shown that Model B and Model P* are indeed two different parameterizations of the same model with the following correspondence between their parameters:

$$\phi_{*jk}^t = (1 - \tau_j^t)\left(\sum_{i=1}^{N} \alpha_{ij}^t \phi_{ijk}^t\right), \qquad 2 \le t \le T - 2;$$

$$\beta_{*jk} = (1 - \tau_j^{T-1})\left(\sum_{i=1}^{N} \alpha_{ij}^{T-1} \beta_{ijk}\right).$$

In these formulas, the quantity between the second parentheses is the weighted average over all previous sites of the transition probabilities from j to k, the weight being the probabilities of coming from the different sites. Just as at the end of Section 3, we recognize here the unconditional probability of transition from j to k for a resident individual, one that was previously on one of the N monitored sites. This unconditional transition is then multiplied by the proportion of residents $(1 - \tau_j^t)$ to get the initial transition of Model B. The formula thus states that, in presence of transients, the initial "apparent" transition from j to k is equal to the unconditional transition from j to k of residents times the proportion of residents, a direct generalization of the formula for the apparent initial survival in the single-site transient model (Pradel et al. 1997). Because all the matrices $(\phi_{ijk}^t)_{i,k}$ for fixed j and t are formally invertible, there is a unique solution for the α_{ij}^t's and the τ_j^t's as functions of the ϕ_{*jk}^t's.

5. EXAMPLE: WINTERING SITE FIDELITY OF CANADA GEESE

We illustrate our study with a reanalysis of the movements of Canada geese *Branta canadensis* between three wintering regions, Mid-Atlantic (*Ma*), Chesapeake (*Ch*), and Carolinas (*Ca*), between 1984 and 1989 (Hestbeck, Nichols, and Malecki 1991). As described in the introduction, previous studies had established the need for a memory model to describe these movements as well as the likely presence of transients on at least the two peripheral sites, *Ma* and *Ca* (Pradel, Gimenez, and Lebreton 2005). Open questions were whether the restricted memory modeling approach of Hestbeck, Nichols, and Malecki (1991) was sufficient and whether transients could be considered as absent from the central region.

Therefore, we considered three levels of transience (no transience, transience in the peripheral regions only, transience everywhere) and four ways of modelling memory:

1. "no memory": the transition to any wintering site is independent of the current location (this is the particular case of the AS model in which transition probabilities are constrained in this way: $\phi_{ij}^t = \phi_j^t$, $i, j \in \{1, 2, \ldots, N\}$);

Table 2. Notation for the different fitted models: they differ in the treatment of memory (full, restricted, or absent) and of transience (full, partial or absent). Some, as signaled between parentheses, are among the models studied in the previous sections. Others ($M\tau$ and $m\tau$) are equivalent to studied models (resp. Models B and H) when the proportions of transients, τ's, are estimated in the range [0, 1].

	Transience everywhere (τ)	Transience in peripheral regions (τ_p)	No transience (0)
Full two-step memory (M)	$M\tau$ (Model P*)	$M\tau_p$	$M0$ (Model P)
Restricted two-step memory (m)	$m\tau$	$m\tau_p$	$m0$
One-step memory (AS)	$AS\tau$	$AS\tau_p$	$AS0$
No memory (NoM)	$NoM\tau$	$NoM\tau_p$	$NoM0$

    2. "one-step memory" as in the AS model;

    3. "restricted two-step memory" as in Hestbeck, Nichols, and Malecki (1991);

    4. "full two-step memory" as in Brownie et al. (1993).

By crossing the two factors, memory and transience, we obtain the different models of Table 2. $M\tau$ and $M0$ are respectively Models P* and P of the previous sections. Also, if all proportions of transients are estimated within the range [0, 1], Models $M\tau$ and $m\tau$ are respectively Model B and Model H (see Section 1).

We estimated the parameters of the different models using the software E-SURGE (Choquet, Rouan, and Pradel 2008).

Model selection is based on the Akaike Information Criterion modified for overdispersion (QAIC). Given the large size of the dataset, there was no need for a small-sample correction. The number of estimable parameters is deduced from our study of the parameter redundancy (see Section 3). We calculated the variance inflation factor from the goodness-of-fit test (GOF) for multistate data implemented in program U-CARE (Choquet et al. 2005). After discarding the components corresponding to the effects incorporated in the model (memory: component WBWA; transience: component 3G.SR), we reached a variance inflation factor of 2.376 corresponding to the residual lack of fit. This value is typical of large datasets with several thousands of individuals where individual differences inherent in any animal population are detected. More details about the goodness-of-fit procedure for the goose data can be found in Pradel, Gimenez, and Lebreton (2005). The best model is Model $m0$ with restricted memory and no transients. The second best one is Model $m\tau_p$ with restricted memory and transients only in the Mid-Atlantic and Carolinas regions. The total absence of transients in the best model seems to contradict the results of the goodness-of-fit tests. The details of the GOF component aimed at transience (test 3G.SR) indeed show a systematic trend for the presence of transients at all but one combination of occasions and sites (11 out of 12 contingency tables) and the directional test of the presence of transients is highly significant even after correcting for overdispersion ($z = 5.46$, $P < 10^{-7}$). However, transients may well come in bursts and not be present in significant numbers in all years and sites making them difficult to detect with

Table 3. $\Delta QAIC$ values (for each model relative to the best one) for the models fitted to the Canada goose data.

	Transience everywhere (τ)	Transience in peripheral regions (τ_p)	No transience (0)
Full two-step memory (M)	66.02	58.28	49.97
Restricted two-step memory (m)	8.84	2.28	0.00
One-step memory (AS)	156.70	149.98	181.53
No memory (NoM)	3,157.88	3,157.12	3,161.21

the general Model $m\tau$. It would be interesting to try to model when transients are present, for example, in relation to climatic conditions, but this goes beyond the purpose of this article. Among the set of models fitted, we decided to opt for the model with transients in the peripheral regions, $m\tau_p$ over Model $m0$, as a more biologically realistic alternative. Models with constant parameters were tried but had worse QAIC values.

The estimates of proportions of transients and transitions from Model $m\tau_p$ are given in Tables 4 and 5 respectively. In agreement with the transient detection tests (Pradel, Gimenez, and Lebreton 2005), transients are found in the peripheral regions except in 1987. (The boundary parameters of this year prevent the (α, τ) parameterization we used from being strictly equivalent to the $\phi_\star$ parameterization, but the two approaches remain very close.)

The estimates of annual transition probabilities (Table 5) show clearly the phenomenon of memory for these animals: a site is more likely to be visited at $t + 1$ if it was the site occupied at $t - 1$. For example, we can see in Table 5 that an individual present at time $t = 3$ in the Mid-Atlantic region will be more likely to move to the Chesapeake region at time $t = 4$ if it was in Chesapeake at time $t = 2$ ($\phi^3_{ChMaCh} = 0.375$), than if it was not ($\phi^3_{oMaCh} = 0.077$). However, the probability to move to the Chesapeake region does not differ significantly between a bird that was in the Mid-Atlantic and one that was in the Carolinas at time 2. The restricted memory approach of Hestbeck et al. is sufficient for the description of the geese movements. Transience is more likely limited to the peripheral sites ($m\tau$ vs. $m\tau_p$: $\Delta QAIC = 6.56$). However, transients, as we have seen, may not be present in all years. No transience is detected in 1987 (Table 5).

Table 4. Maximum likelihood estimates of the proportions of transients for Model $m\tau_p$ (estimated standard errors in parentheses).

	1985	1986	1987	1988
Mid-Atlantic	0.053 (0.102)	0.156 (0.048)	0.000 (0.118)	0.269 (0.081)
Carolinas	0.213 (0.044)	0.305 (0.226)	0.000 (0.195)	0.206 (0.184)

Table 5. Maximum-likelihood estimates of survival-transitions probabilities between sites Ma, Ch, and Ca for Model $m\tau_p$ (estimated standard errors between parentheses). The sign ○ indicates that the site occupied at $t-1$ is different from the location at $t+1$. Since the study of parameter redundancy has shown that the ϕ^5_{ijk} parameters are not separately estimable we have not mentioned them.

Succesion of sites	$t=2$ (1985–1986)	$t=3$ (1986–1987)	$t=4$ (1987–1988)
Ma Ma Ma	0.581 (0.049)	0.581 (0.034)	0.527 (0.036)
○ Ma Ma	0.492 (0.065)	0.313 (0.044)	0.272 (0.049)
Ch Ma Ch	0.218 (0.059)	0.375 (0.051)	0.268 (0.046)
○ Ma Ch	0.065 (0.027)	0.077 (0.017)	0.148 (0.020)
Ca Ma Ca	0.000 (0.019)	0.102 (0.106)	0.000 (0.066)
○ Ma Ca	0.000 (0.005)	0.006 (0.004)	0.006 (0.003)
Ma Ch Ma	0.293 (0.069)	0.125 (0.038)	0.211 (0.036)
○ Ch Ma	0.065 (0.011)	0.045 (0.007)	0.045 (0.007)
Ch Ch Ch	0.639 (0.031)	0.676 (0.029)	0.569 (0.023)
○ Ch Ch	0.367 (0.061)	0.512 (0.039)	0.544 (0.044)
Ca Ch Ca	0.044 (0.027)	0.144 (0.042)	0.309 (0.085)
○ Ch Ca	0.003 (0.003)	0.008 (0.003)	0.020 (0.004)
Ma Ca Ma	0.000 (0.068)	0.396 (0.004)	0.474 (0.227)
○ Ca Ma	0.048 (0.025)	0.011 (0.010)	0.019 (0.016)
Ch Ca Ch	0.514 (0.307)	0.216 (0.113)	0.242 (0.086)
○ Ca Ch	0.316 (0.061)	0.160 (0.036)	0.072 (0.026)
Ca Ca Ca	0.428 (0.065)	0.629 (0.062)	0.651 (0.062)
○ Ca Ca	0.143 (0.131)	0.444 (0.151)	0.218 (0.079)

6. DISCUSSION

Setting all previous capture–recapture memory models in the multievent framework has allowed comparison of these models. All of them have a common feature: they take memory into account by considering that, of the animals present the same year at the same site, those that occupied different sites the previous year may behave differently in their future movements, that is, have different probabilities of occupying the different sites the next year. A difficulty arises the first time an individual is encountered: its previous site is unknown. It appeared that the previous models (Brownie et al. and Pradel) correspond to two very different options. One option consists of trying to estimate this previous location (Pradel 2005); the other simply allows the newly encountered individuals to have an entirely different behavior than the previously encountered individuals (Brownie et al. 1993). We have seen that this second option is more general. In essence, this means that the newly encountered individuals cannot always be reduced to being a mixture of the different types of previously encountered individuals. This may arise for many different reasons: they may be younger animals on average; they may suffer from being handled when marked; some may also happen to be truly new to the study area. We have studied this last situation by considering the possible presence of transients among the newly encountered and we have seen that, with this allowance, there is no longer a difference between the two options under two weak conditions: that the relative proportions of the newly encountered individuals moving to the different sites can be reproduced by mixing in an appropriate way the differ-

ent categories of previously encountered individuals (this is possible in general); that the resulting mixture of previously encountered individuals has a yearly survival higher than that of the newly encountered individuals. When the second condition is met, the ratio of the two survivals estimates the proportion of residents, and, therefore, its complement to one estimates the proportion of transients (see the last paragraph of Section 4). When the last condition is not verified, as happened once in our example, one may get a negative estimate of the proportion of transients. To avoid this, because the discrepancy was not large and could easily have arisen by chance, we chose to constrain the proportion of transients to remain in the range $[0, 1]$.

Using a symbolic-numeric version of the Catchpole–Morgan method (1997), we have also been able to identify the estimable quantities in the models. The last interval is as always problematic in CR and the ability to estimate survivals and transitions thus supposes at least three time steps. Additionally, the probability of the previous location cannot be estimated at the first date.

For more readability, this article is focused on the two time step models, but the multievent formulation of the memory models also provides a frame of reference for further generalizations. Thus, it may be reasonable to consider the locations occupied more than one time step before. It may also be reasonable to consider models where the probability to move to a site increases with the number of times this site has been visited. All of these kinds of models can be fitted as multievent models. Still within the multievent framework, it is possible to work at achieving more parsimony. The standard errors of many parameters are large in our best model and it is likely overparameterized (see Table 5). A possibility would be to decompose the transitions as in the multistate Arnason–Schwarz model in order to manipulate survival separately from conditional movement. This is readily doable with program E-SURGE (Choquet, Rouan, and Pradel 2008). In this way, it would be possible to relate survival to some covariate of interest like weight or a measure of overall size. Differences in body conditions could, for example, explain why, of the geese present in the Chesapeake region in Year 2, those which were already in the Chesapeake region in Year 1 have a much higher survival probability (0.707 s.d. 0.061) than those which were in the Carolinas (0.476 s.d. 0.027). More flexibility is clearly needed and we believe that further developments should aim at studying models where survival and proportion of transients are explicit parameters. This would permit the comparison of these parameters among different categories of birds and probably a reduction of the total number of parameters.

APPENDIX A: NEGLIGIBLE SET OF SINGULAR POINTS

Suppose that the formal rank of the derivative matrix D is full and equal to p. We show in this appendix that the set of vector-parameters θ such that $\text{rank}(D(\theta)) < p$ is negligible—that is, that the Lebesque measure of the set $E = \{\theta \in \Theta, \text{rank}(D(\theta)) < p\}$ is null with $\Theta = \mathbb{R}^p$, using the logit link, or $\Theta =]0, 1[^p$, using the identity link.

The derivative matrix D is supposed full rank so, there exists one $[p \times p]$ submatrix denoted D^* and $\theta_* \in \Theta$ such that $\det(D^*(\theta_*)) \neq 0$.

We reason by contradiction. Suppose that the Lebesgue measure of E is nonnull. Then, there exists $\theta_0 \in E$ and $\varepsilon > 0$ and an open ball $B(\theta_0, \varepsilon)$ such that for $\forall \theta \in B(\theta_0, \varepsilon)$ and $\forall A$ submatrix of $D(\theta)$ of size p,

$$\det(A(\theta)) = 0$$

(see, e.g., Choquet 1973).

In particular for $A = D^*$, $f(\theta) = \det(D^*(\theta)) = 0, \forall \theta \in B(\theta_0, \varepsilon)$.

For the identity link f is a polynomial so an analytic function on $]0, 1[^p$ null on $B(\theta_0, \varepsilon)$. We can easily check that for the logit link, f is still an analytic function on $\mathbb{R}^p$ null on $B(\theta_0, \varepsilon)$. So, according to Cartan (p. 124), f is null on $\mathbb{R}^p$ (resp. $]0, 1[^p$) which is not possible since $f(\theta_*) \neq 0$. Then the Lebesgue measure of E is null.

APPENDIX B: EXPANSION RESULTS FOR MODEL B WITH THREE SITES

Let us assume that we have established, for T capture occasions, that the model is full rank; in other words, its derivative matrix (see Section 3) is full rank:

$$\operatorname{rank}(D_{H_T}(\theta_T)) = \operatorname{rank}\left(\left(\frac{\partial \mu_i(\theta_T)}{\partial \theta_{T,l}}\right)_{\substack{1 \leq i \leq n_h^T \\ 1 \leq l \leq N_T}}\right)$$
$$= N_T,$$

where θ_T is the vector containing the N_T parameters defining the model and n_h^T is the number of possible capture histories with T occasions (H_T denote the set of these histories).

If we consider, now, $T + 1$ capture occasions, the set H_{T+1} of capture histories can be divided into two subsets:

- H_{T+1}^0: the encounter histories associated with the individuals first captured at time 2 or later. This subset is made of the capture histories of H_T with a leading 0. It is of size n_h^T.

- H_{T+1}^1: the encounter histories associated with the individuals first captured at time 1. This subset has size $(N + 1) \times n_h^T$.

Thus, H_{T+1} can be considered as the extension of H_T by addition of an extra occasion at the beginning.

We denote θ_* the vector of additional parameters relative to the extra occasion at the beginning needed to define the model with $T + 1$ capture occasions; the vector of parameters can be written $\theta_{T+1} = (\theta_*, \theta_T)$. The derivative matrix $D_{H_{T+1}}(\theta_{T+1})$ can be partitioned

into blocks:

$$D_{H_{T+1}}(\theta_{T+1}) = \begin{pmatrix} H_{T+1}^0 \\ H_{T+1}^1 \end{pmatrix} \begin{pmatrix} \theta_* & \theta_T \\ 0 & D_{H_{T+1}^0}(\theta_T) \\ \hline D_{H_{T+1}^1}(\theta_*) & D_{H_{T+1}^1}(\theta_T) \end{pmatrix}.$$

Given that we assumed that $D_{H_{T+1}^0}(\theta_T)$ is full rank, $D_{H_{T+1}}(\theta_{T+1})$ is full rank if and only if $D_{H_{T+1}^1}(\theta_*)$ is full rank.

$D_{H_{T+1}^1}(\theta_*)$ is a $[N^3 + (N+1)^T] \times [N^3 + N^2 + 2N - 1]$ matrix. For $T > 2$, $N^3 + (N+1)^T > N^3 + N^2 + 2N - 1$ so that for all $T > 2$ the maximum rank of $D_{H_{T+1}^1}(\theta_*)$ is equal to $N^3 + N^2 + 2N - 1$ that is the number of new parameters. Now, if we decompose the set of capture histories H_{T+1}^1 into two subsets:

- $H_{T+1}^{1,0}$: the $N(N+1)^3$ capture histories with $T - 3$ trailing 0's;

- $H_{T+1}^{1,1}$: the $N(N+1)^3 N^{T-3}$ other capture histories.

We show, using the symbolic-numeric method, that the rank of $D_{H_{T+1}^{1,0}}(\theta_*)$ is equal to $N^3 + N^2 + 2N - 1$. Thus, $D_{H_{T+1}^1}(\theta_*)$ is full rank and using the previous result, $D_{H_{T+1}}(\theta_{T+1})$ is full rank.

SUPPLEMENTAL MATERIALS

The supplemental material illustrates the step-by-step procedures in fitting these models using E-SURGE.

ACKNOWLEDGMENTS

REFERENCES

Arnason, A. N. (1973), "The Estimation of Population Size, Migration Rates and Survival in a Stratified Population," *Researches on Population Ecology*, 15, 1–8.

Brownie, C., Hines, J. E., Nichols, J. D., Pollock, K. H., and Hestbeck, J. (1993), "Capture–Recapture Studies for Multiple Strata Including Non-Markovian Transitions," *Biometrics*, 49, 1173–1187.

Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., and Pollock, K. H. (1987), *Design and Analysis Methods for Fish Survival Experiments Based on Release-Recapture*, Bethesda, MD: American Fisheries Society.

Catchpole, E. A., and Morgan, B. J. T. (1997), "Detecting Parameter Redundancy," *Biometrika*, 84, 187–196.

——— (2001), "Deficiency of Parameter-Redundant Models," *Biometrika*, 88, 593–598.

Choquet, G. (1973), *Cours d'analyse, Tome 2, Topologie*, Paris: Masson.

Choquet, R., Reboulet, A.-M., Lebreton, J.-D., Gimenez, O., and Pradel, R. (2005), "U-CARE 2.2 User's Manual," CEFE, Montpellier, France.

Choquet, R., Rouan, L., and Pradel, R. (2008), "Program E-SURGE: A Software Application for Fitting Multi-event Models," in *Modeling Demographic Processes in Marked Populations. Environmental and Ecological Statistics*, Vol. 3, eds. D. L. Thomson, E. G. Cooch, and M. J. Conroy, New York: Springer, pp. 845–865.

Gimenez, O., Choquet, R., and Lebreton, J.-D. (2003), "Parameter Redundancy in Multistate Capture–Recapture Models," *Biometrical Journal*, 45, 704–722.

Hestbeck, J. B., Nichols, J. D., and Malecki, R. A. (1991), "Estimates of Movement and Site Fidelity Using Mark-Resight Data of Wintering Canada Geese," *Ecology*, 72, 523–533.

Mourrain, B., and Prieto, H. (2000), "A Framework for Symbolic and Numeric Computations," Rapport de Recherche 4013, INRIA.

Pradel, R. (2005), "Multievent: An Extension of Multistate Capture–Recapture Models to Uncertain States," *Biometrics*, 61, 442–447.

Pradel, R., Gimenez, O., and Lebreton, J.-D. (2005), "Principles and Interest of Gof Tests for Multistate Capture–Recapture Models," *Animal Biodiversity and Conservation*, 28, 189–204.

Pradel, R., Hines, J. E., Lebreton, J.-D., and Nichols, J. D. (1997), "Capture–Recapture Survival Models Taking Account of Transients," *Biometrics*, 53, 60–72.

Pradel, R., Wintrebert, C., and Gimenez, O. (2003), "A Proposal for a Goodness-of-Fit Test to the Arnason–Schwarz Multisite Capture–Recapture Model," *Biometrics*, 59, 43–53.

Ridout, M. S. (1999), "Memory in Coal Tits: An Alternative Model," *Biometrics*, 55, 660–662.

Ron, D., Singer, Y., and Tishby, N. (1996), "The Power of Amnesia: Learning Probabilistic Automata With Variable Memory Length," *Machine Learning*, 25, 117–149.

Schwarz, C., Schweigert, J., and Arnason, A. (1993), "Estimating Migration Rates Using Tag-Recovery Data," *Biometrics*, 49, 177–193.