

# Applications of hidden hybrid Markov/semi-Markov models: from stopover duration to breeding success dynamics

Rémi Choquet, Arnaud Béchet, Yann Guédon

► **To cite this version:**

Rémi Choquet, Arnaud Béchet, Yann Guédon. Applications of hidden hybrid Markov/semi-Markov models: from stopover duration to breeding success dynamics. *Ecology and Evolution*, Wiley Open Access, 2014, 4 (6), pp.817 - 826. <10.1002/ece3.962>. <hal-01084498>

**HAL Id: hal-01084498**

**<https://hal.inria.fr/hal-01084498>**

Submitted on 19 Nov 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Applications of hidden hybrid Markov/semi-Markov models: from stopover duration to breeding success dynamics

Rémi Choquet<sup>1</sup>, Arnaud Béchet<sup>2</sup> & Yann Guédon<sup>3</sup>

<sup>1</sup>Centre d'Ecologie Fonctionnelle et Evolutive, Campus CNRS, UMR 5175, 1919 Route de Mende, Montpellier Cedex 5, 34293, France

<sup>2</sup>Centre de recherche de la Tour du Valat, Le Sambuc, Arles 13200, France

<sup>3</sup>CIRAD, UMR AGAP and Inria, Virtual Plants, F-34095 Montpellier, France

### Keywords

Capture–recapture, incubation.

### Correspondence

Rémi Choquet, Centre d'Ecologie Fonctionnelle et Evolutive, Campus CNRS, UMR 5175, 1919 Route de Mende, 34293 Montpellier Cedex 5, France.  
Tel: +33 467613203; Fax: +33 467613336;  
E-mail: remi.choquet@cefe.cnrs.fr

### Funding Information

This research was supported by a BLOSTIC project 'Techniques markoviennes avancées en dynamique des populations' financed by the Montpellier sud de France University.

Received: 12 December 2013; Accepted: 4 January 2014

*Ecology and Evolution* 2014; 4(6): 817–826

doi: 10.1002/ece3.962

## Introduction

In the study of population dynamics, the generation (or life) time of species has long been used to classify life-history trait variations. Species may be placed along the r- and K-strategy gradient (Pianka 1970) depending on their reproductive and survival rates (Saether *et al.* 2004). High reproductive species with short life expectancy are at one end of the continuum (fast), and species that often produce a single offspring but are long lived are at the other (slow). This classification has, for instance, been used to predict how demographic stochasticity affects population dynamics (Saether *et al.* 2004) or to detect cost of reproduction (Hamel *et al.* 2010). The time spent by an individual in a particular biologic state (e.g., alive or breeding) is of critical importance for evolutionary ecologists. Beyond the time spent alive or breeding, remaining lifetime is spent in different phases (e.g., dormancy in plants) or in

## Abstract

Usually in capture–recapture, a model parameter is time or time since first capture dependent. However, the case where the probability of staying in one state depends on the time spent in that particular state is not rare. Hidden Markov models are not appropriate to manage these situations. A more convenient approach would be to consider models that incorporate semi-Markovian states which explicitly define the waiting time distribution and have been used in previous biologic studies as a convenient framework for modeling the time spent in a given physiological state. Here, we propose hidden Markovian models that combine several nonhomogeneous Markovian states with one semi-Markovian state and which (i) are well adapted to imperfect and variable detection and (ii) allow us to consider time, time since first capture, and time spent in one state effects. Implementation details depending on the number of semi-Markovian states are discussed. From a user's perspective, the present approach enhances the toolbox for analyzing capture–recapture data. We then show the potential of this framework by means of two ecological examples: (i) stopover duration and (ii) breeding success dynamics.

specific behavior (courtship display, resting, hibernating, molting in animals). In general, the probability of staying, and therefore exiting, in a given state depends on the time elapsed in this state. It may present little individual variation due to certain physiological constraints, or because it has been shaped by evolution (e.g., incubation or gestation). In contrast, the time spent in other biologic states may be very variable because of either intrinsic individual variations, or heterogeneous response to density dependence or environmental effects [e.g., courtship, refueling, molting; Dietz *et al.* (2013)].

Empirical knowledge of some state durations may help in estimating the causes of life-history trait variations. For example in Schmaltz *et al.* (2011), consideration of incubation duration led to an understanding of variations in greater flamingo breeding performance. After hatching, variations in adult provision rates also affect the duration of fasting periods in the juveniles of several bird species

and may help diagnose sex-specific costs of parental care (e.g., Booth *et al.* 2000; Amat *et al.* 2007). Tracking temporary and definitive changes in biologic stages is of critical importance in several species. Some species will respond to seasonal variations in environmental conditions by changing physiological state, for example, hibernation (Bryant and Page 2005) or by migrating, while others will permanently change their life-history stage: caterpillars becoming butterflies, tadpoles becoming frogs, etc. Such transitional or definitive stages constitute critical life phases, and their duration is central to the understanding of population dynamics.

A clear understanding of variation in life-history traits has numerous applications. Estimating of the stopover duration in migrating birds or time to extinction in population viability analyses is critical to conservation biologists. Similarly, an understanding of the dynamics of epidemic spread is also dependent on in-depth knowledge of the time spent by individuals in the infected state (Shaman and Karspeck 2012).

The state of an individual is determined from different clues (visual appearance, behavior, measurements, etc.). But, it may be observed incorrectly or remain unobserved. Even in the simplest situation, state duration may be difficult to estimate. For example, knowledge of life duration is conditional on knowledge of times of birth and death. In the wild, time of death is rarely known, but classical CR models can deal with this (Lebreton *et al.* 1992; Choquet *et al.* 2011) by estimating a capture probability. Thus, we can deal with the lack of information after the last capture of an individual by stating that after that date, an individual may be dead or missed at capture. But sometimes the age of an individual is imperfectly known, and time of birth should also be modeled (Colchero and Clark 2012; Matechou *et al.* 2013).

A major drawback of hidden Markov models is the inflexible description of the time spent in a given state, as sojourn time (state occupancy) distributions are implicitly geometric. To overcome this limitation, a semi-Markovian framework may be considered where parametric sojourn time distributions are incorporated in the model, or where states are replaced by series-parallel networks of states with a common observation distribution; see Guédon (2005), Langrock and Zucchini (2011) and references therein. The sojourn time distributions of the macro-states defined in this way are built from the implicit geometric sojourn time distributions of the elementary Markovian states. These geometric distributions are combined either by convolution for states in series or by mixture for (series of) states in parallel. Guédon (2005) showed that hidden Markov models with macro-states are not a valid alternative to hidden semi-Markov models because of higher algorithm space complexity and strong constraints in the

definition of parametric macro-state sojourn time distributions. It should be noted that Markovian and semi-Markovian states can easily be combined in hidden hybrid models. In particular, Choquet *et al.* (2013b), unified the models proposed by Pledger *et al.* (2009), Pradel (2009) and Fewster and Patenaude (2009) and proposed a new model combining nonhomogeneous Markovian states with a semi-Markovian state and Markov chain observation models to represent trap effects.

In this article, we consider hidden Markovian models that combine several nonhomogeneous Markovian states with one semi-Markovian state. We show how the length of migratory birds stop overs can be easily written using this extended framework. In the second application, we incorporate accurately known duration of incubation to study the breeding dynamics of a colonial waterbird. The approach proposed is unusual in CR models. Finally, possible generalizations and applications of this work are discussed.

## Hidden Hybrid Markov/Semi-Markov Models

Assume that we have  $T$  capture occasions and  $N$  individuals captured at least once. Let the encounter history for individual  $i$  be  $h_i = (o_{i1}, \dots, o_{iT})$  where  $o_{it}$  denotes whether individual  $i$  is observed ( $o_{it} > 0$ ) or not ( $o_{it} = 0$ ) at occasion  $t$ .

### Model definition

Let  $S_t$  be a hybrid Markov/semi-Markov model with finite state space  $1, \dots, J$ ; see Kulkarni (1995) for a general description of Markov and semi-Markov models. This  $J$ -state hybrid Markov/semi-Markov model is defined by the following parameters:

- initial probabilities  $\pi_j = P(S_1 = j)$  with  $\sum_j \pi_j = 1$ ,
- transition probabilities
  - semi-Markovian state  $j$ : for each  $k \neq j$ ,  $\phi_{jk} = P(S_{t+1} = k | S_{t+1} \neq j, S_t = j)$  with  $\sum_{k \neq j} \phi_{jk} = 1$  and  $\phi_{jj} = 0$ ,
  - Markovian state  $j$ :  $\phi_{jk} = P(S_{t+1} = k | S_t = j)$  with  $\sum_k \phi_{jk} = 1$ .

An explicit sojourn time distribution is attached to each semi-Markovian state

$$d_j(u) = P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = 0, \dots, u-2 | S_{t+1} = j, S_t \neq j), \quad u = 1, M_j,$$

where  $M_j$  denotes the upper bound of the time spent in state  $j$ . Hence, we assume that sojourn time distributions are concentrated on finite sets of time points.

The output process  $\{O_t\}$ , here the encounter history  $h_i$  for individual  $i$ , is related to the hybrid Markov/semi-Markov chain  $\{S_t\}$  by the observation (or emission) distributions

$$b_j(y) = P(O_t = y | S_t = j) \text{ with } \sum_y b_j(y) = 1$$

Let  $\mathbf{b}(h) = (b_1(o_1), \dots, b_j(o_T))'$  denote the column vector of observation probabilities of history  $h$  and  $\pi$  denote the vector of initial probabilities. Let  $\Phi$  denote the matrix of transition probabilities. In the following, we assume that some observation distributions are a function of the index parameter  $t$  introducing some nonhomogeneity into the model.

### Estimation problem with one semi-Markovian state

By convention, the semi-Markovian state number is  $SM$ , the Markovian transient states that precede this semi-Markovian state are numbered from 1 to  $SM-1$  and the successive Markovian states that follow the semi-Markovian state are numbered from  $SM+1$  to  $J$ . We consider a model that is conditional on the first occasion. The probability of an encounter history  $h = (o_1, \dots, o_K)$  is

$$P(h) = \pi \text{diag}(\mathbf{b}^{(1)}(o_1)) \left( \prod_{t=1}^{K-1} \Phi^{(t)} \text{diag}(\mathbf{b}^{(t+1)}(o_{t+1})) \right) \mathbf{1}_{1:SM-1} + \pi \text{diag}(\mathbf{b}^{(1)}(o_1)) \sum_{ar=2}^{K-1} \left( \prod_{t=1}^{ar-1} \Phi^{(t)} \text{diag}(\mathbf{b}^{(t+1)}(o_{t+1})) \right) \times \sum_{dep=ar}^K \text{diag}(0, \dots, 0, d_{SM}(dep - ar) \prod_{t=ar}^{dep} b_{SM}^{(t)}(o_t), 0, \dots, 0) \times \prod_{t=dep}^{K-1} \Phi^{(t)} \text{diag}(\mathbf{b}^{(t+1)}(o_{t+1})) \mathbf{1}_{SM:J}, \tag{1}$$

where  $\text{diag}(x)$  is a diagonal matrix with  $i$ th diagonal element  $x_i$  and  $\mathbf{1}_{ij}$  is a  $J$  column vector of one's between  $i$  and  $j$  and 0 elsewhere. The first line of equation (1) corresponds to the probability of no occurrence of the semi-Markovian state. Rows 2 to 4 correspond to the sum of probabilities of occurrence of the semi-Markovian state between the time of arrival  $ar$  and the time of departure  $dep$ . The likelihood is conditional on the set of individuals seen at least once in the experiment. With the usual assumption of independence among individuals, the likelihood  $L$  is the product of the  $N$  individual contributions

$$L = \prod_{i=1}^N P(h_i) / (1 - P(\emptyset)) \tag{2}$$

where  $\emptyset$  is a history with no capture. In Choquet *et al.* (2013b), developed specifically for stopover duration, individuals enter the semi-Markovian state before the first observation. Equation (1) extends this work by allowing individuals to enter or not enter the semi-Markovian state at any occasion. We can also easily consider conditioning on the first capture occasion. To do this, we need to start

equation (1) from the first event and define the likelihood as  $L = \prod_{i=1}^N P(h_i)$ .

### Applications

We assume that the nonobservable Markov/semi-Markov model comprises several transient states and a final absorbing state, with all transient states being nonhomogeneous Markovian except one that is semi-Markovian.

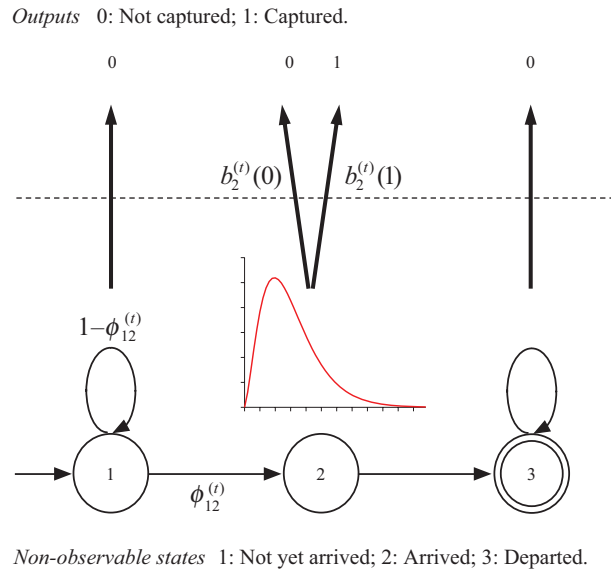
For the stopover duration application, the nonobservable Markov/semi-Markov model comprises three states: an initial transient Markovian state followed by a transient semi-Markovian state, and a final absorbing state. A state is said to be transient if the probability of definitively leaving this state without the possibility of returning to it is  $>0$ . The states are thus ordered, and each state can be visited exactly once ("left-right" model); see Fig. 1.

For the breeding attendance application, the nonobservable Markov/semi-Markov model comprises six states: two Markovian states forming an initial transient class

followed by a transient semi-Markovian state, then two successive transient Markovian states, and a final absorbing state. The initial transient class and the 4 following states are ordered; the states are thus partially ordered (no ordering between states 1 and 2 in the initial transient class), and only the first two states can be visited more than once; see Fig. 2.

The specific assumptions of the two applications are as follows:

- We assume that individuals enter in the experiment after the first sampling occasion. We assume that the last sampling occasion individual state is not the semi-Markovian state, that is, sampling covers all the time when individuals are present in the semi-Markovian state of interest (i.e., no right censoring of the time spent in this state).
- We assume that individuals are in a superpopulation and enter the experiment where captures are possible. Likelihood (2) is conditional on an animal being seen somewhere in the experiment, as in Link and Barker (2005); Pradel (1996).



**Figure 1.** Nonobservable hybrid Markov/semi-Markov chain for stopover duration: Each state is represented by a numbered vertex. Vertices representing transient states are edged by a single line while the vertex representing the final absorbing state is edged by a double line. The possible transitions between states are represented by arcs going left to right with the attached probabilities noted nearby when  $<1$ . The arc entering state 1 indicates that it is the only possible initial state. The sojourn time distribution of the semi-Markovian state 2 is shown above it. Observation process: In the case of degenerate observation distributions (states 1 and 3), the only possible output is shown above the corresponding state. In the other case, the possible outputs with associated observation probabilities are shown above the corresponding state.

### Estimating stopover duration

For birds that need to refuel during their migratory journey, the time spent at a stopover site may be identical whatever the date of arrival and may be determined by the number of days needed for feeding. Recently, Pledger *et al.* (2009); Pradel (2009); Fewster and Patenaude (2009) independently proposed a time elapsed as arrival-dependent modeling approaches for estimating stopover duration. To deal with residence time since arrival, we cannot condition on the past state as in the classical first-order Markov chain. We also need to know for how long an individual has been present. One convenient method for representing stopover duration is to introduce semi-Markovian states where the time spent in these states is explicitly modeled by appropriate sojourn time distributions.

To do this, we consider a three-state model made up of an initial "not yet arrived" transient state which is the state before arrival at the site, an intermediate "arrived" transient state which is the state of arrival at the site, and a final "departed" absorbing state which is the state of

leaving the site. The intermediate "arrived" state is semi-Markovian, while the other two "not yet arrived" and "departed" states are Markovian. We assume that "not yet arrived" state is the only possible initial state, that is,  $\pi_1=1$ . For this initial transient Markovian state, we assume that only the self-transition and the transition to the semi-Markovian "arrived" state are possible with a transition distribution that depends on the time index  $t$ . For this nonhomogeneous Markovian state, full time-dependent models are used for the transition distribution. Hence, the initial probability vector and the transition probability matrix are given by

$$\pi = (1, 0, 0),$$

$$\Phi^{(t)} = \begin{pmatrix} 1 - \phi_{12}^{(t)} & \phi_{12}^{(t)} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad (3)$$

with  $\phi_{12}^{(t)}$  the probability of an individual not present at the site actually arriving between occasions  $t$  and  $t+1$ . We assume that  $\phi_{12}^{(T-1)}=1$ , that is, all individuals are present at the site before occasion  $T$ . A parametric or nonparametric sojourn time distribution is attached to the semi-Markovian "arrived" state. In Choquet *et al.* (2013b), several non-geometric distributions were used and compared.

The set of observations is  $O_t = \{0 \text{ for "not captured", } 1 \text{ for "captured"}\}$ . Concerning the observation models, we assume that only the "not captured" output can be observed in the initial "not yet arrived" state. The corresponding observation distribution is thus degenerate. For the intermediate "arrived" state, the observation distribution depends on the index parameter  $t$ . The observation probability matrix is thus given by

$$B^{(t)} = \begin{pmatrix} 1 & 0 \\ b_2^{(t)}(0) & b_2^{(t)}(1) \\ 1 & 0 \end{pmatrix} \quad (4)$$

with  $b_j^{(t)}(y) = (B^{(t)})_{j,y+1}$  the probability of an individual being in state  $j$  at time  $t$  to be observed as  $y$ . This model, represented in Fig. 1, can be easily generalized to deal with different sources of heterogeneity. For example, the output process can be adapted to the case where trap effects are present (Choquet *et al.* 2013b). Neglecting trap effects, sometimes due to the food used to catch the birds and stress caused by handling, can lead to marked bias in estimating stopover duration.

### Incorporating incubation period duration

Since 1974, greater flamingos have bred on an artificial island located in the commercial saltpans of Salin-de-Giraud (Camargue, southern France). Flamingos

generally start incubating in April. At age of 10 days, the chicks gather in small crèches on the breeding island before moving into the water to join a larger crèche which may contain up to several thousand individuals (Johnson and Cezilly 2007). Since 1977, 12% of the chicks fledged in the Camargue have been ringed each year with PVC rings engraved with alphanumeric codes that can be read from a distance of up to 300 m. We used the subsequent resightings of these ringed birds as breeding adults at the colony in 1991. In that year, an average of 10 hours per day was spent in the towel observing the colony from April 2 (first egg observed) to July 16, and 2 hours per day were spent at the crèche, from July 16 to September 8 (end of observation period). Ringed birds were resighted by means of a telescope located in a hide 70 m from the colony. Only flamingos observed at least once as breeders were considered. We divided the breeding season into 18 intervals, each of 10 days.

The set of observations is  $O_t = \{0 \text{ for "not captured", } 1 \text{ for "seen incubating an egg", } 2 \text{ for "feeding a chick on the breeding island", } 3 \text{ for "feeding a chick in the crèche"}\}$ .

We considered a six-state model made up of a transient class with two states "elsewhere/preincubating", and "incubation with failure", three successive transient states "incubation with success", "feeding on the island" and "feeding in the crèche", and a final absorbing state "departed after incubation with success". The intermediate transient state "incubation with success" is semi-Markovian, while the other states are Markovian; see Fig. 2 for a representation of this model. We used the fact that the length of the incubation period is fixed and is known to last 29 days. The sojourn time distribution of incubation becomes  $d_{SM}(3) = 1$  with SM= "incubation with success". An individual that has not succeeded in the incubation may move back to the initial state, "elsewhere/preincubating".

We assume here that "elsewhere/preincubating" is the only possible initial state, that is,  $\pi_1 = 1$ . Hence, the initial probability vector and the transition probability matrix are given by:

$$\pi = (1, 0, 0, 0, 0, 0),$$

$$\Phi^{(t)} = \begin{pmatrix} 1 - \phi_{12}^{(t)} - \phi_{13}^{(t)} & \phi_{12}^{(t)} & \phi_{13}^{(t)} & 0 & 0 & 0 \\ \phi_{21}^{(t)} & 1 - \phi_{21}^{(t)} - \phi_{23}^{(t)} & \phi_{23}^{(t)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 - \phi_{35}^{(t)} & \phi_{35}^{(t)} & 0 \\ 0 & 0 & 0 & 1 - \phi_{45}^{(t)} - \phi_{46}^{(t)} & \phi_{45}^{(t)} & \phi_{46}^{(t)} \\ 0 & 0 & 0 & 0 & 1 - \phi_{56}^{(t)} & \phi_{56}^{(t)} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5}$$

The observation probability matrix is:

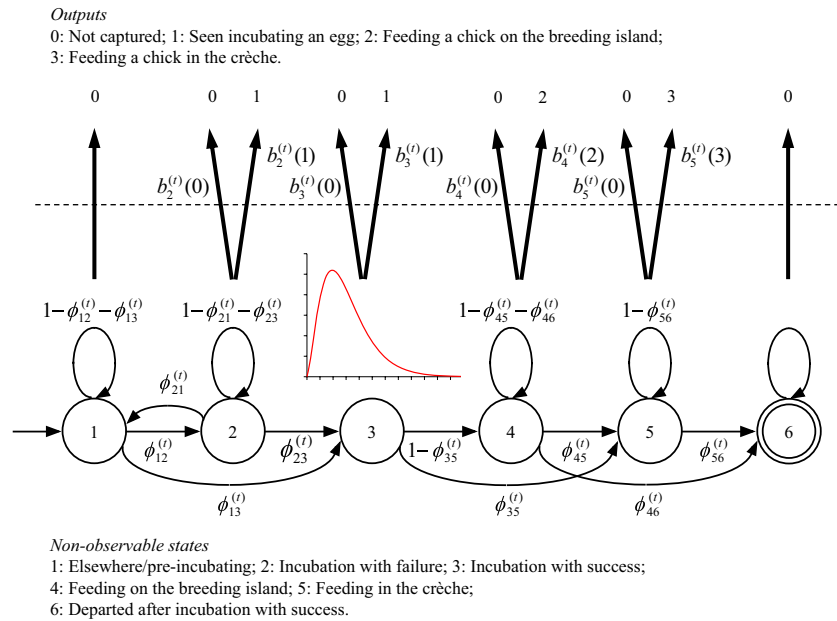
$$B^{(t)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ b_2^{(t)}(0) & b_2^{(t)}(1) & 0 & 0 \\ b_3^{(t)}(0) & b_3^{(t)}(1) & 0 & 0 \\ b_4^{(t)}(0) & 0 & b_4^{(t)}(2) & 0 \\ b_5^{(t)}(0) & 0 & 0 & b_5^{(t)}(3) \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{6}$$

Because the reproduction pattern changes with time, we need specific transition probabilities for the beginning, the middle, and the end of the season. Furthermore, Schmaltz *et al.* (2011) demonstrated that rainfall has an effect on reproduction at intervals 7 and 8 (i.e., between occasions 7 and 9). Thus, for states 1 to 3, we consider successive time periods corresponding to groupings of intervals 1 to 3, 4 to 6, 7, 8, 9 to 10, 11 to 15, and 16 to 18 for which transitions are set constant. Furthermore, we assume that  $\phi_{13}^{(t=15:18)} = 0$  and  $\phi_{23}^{(t=15:18)} = 0$ . For states 4 to 6, we consider three successive time periods, 4 to 6, 7 to 8, and 9 to 18.

We consider also specific capture probabilities for successive time periods, occasions 4 to 7, 8, 9 and 10 to 19 for which capture probabilities are constant. Furthermore, as there is no capture for the first period,  $b_i^{t=1,\dots,3} = 0$ .

Models were built using the E-SURGE program (Choquet *et al.* 2009), and we computed maximum likelihood estimates (MLE) for each model using a quasi-Newton algorithm (Dennis and Schnabel 1983) with multiple starting points to avoid spurious local minima. MLE and confidence intervals are given in Tables 1 and 2. We did not attempt to find the most parsimonious model here using Akaike information criteria (Akaike 1987) as this was not the goal of this study. Guédon (1999) proposed a validation methodology relying on the fit of different types of characteristic distributions computed from model parameters to their empirical equivalents extracted from data. In particular, we used the fit of output distributions conditional on the observed sequence at each capture occasion; see Guédon (2005) for the recursive algorithms for computing state and output distributions conditional





**Figure 2.** Nonobservable hybrid Markov/semi-Markov chain for breeding success dynamics: The arc entering state 1 indicates that it is the only possible initial state. The sojourn time distribution of the semi-Markovian state 3 is shown above it. Observation process: In the case of degenerate observation distributions (states 1 and 6), the only possible output is shown above the corresponding state. In the other cases, the possible outputs with associated observation probabilities are shown above the corresponding state.

on the observed sequence for each successive time  $t$  in the case of hidden hybrid Markov/semi-Markov models. We compared the proportion of estimated and observed individuals for each possible output. Figure 3 shows that the model successfully reproduces the general pattern of the observations.

We show in Fig. 4 the state distributions at each capture occasion in order to illustrate the dynamics of incubation. The general pattern is shown in Fig. 4 corresponds to expectations regarding the successive steps of greater flamingo breeding dynamics. Birds incubated (successfully or not) from the start of the breeding period to occasion 14 at which point the island was deserted. At this occasion, no birds started to incubate as they would not have had enough time to complete the full breeding cycle before food and climatic conditions deteriorate in the fall. Hatching was first observed at occasion 6 (Fig. 3). From then on, the proportion of birds feeding a chick on the island then in the crèche increased gradually until the end of the observation period (Fig. 4). Regarding incubation, occasions 7 and 8 appeared to be pivotal given that the proportion of incubating birds that managed to hatch their egg (incubation with success) increased slowly until occasion 7 then peaked at occasions 8 and 9. In contrast after occasions 7 and 8, the proportion of unsuccessful incubating birds (incubation with failure) dropped almost to zero. Heavy

rains at occasions 7 and 8 [corresponding to occasions 4 and 5 in Schmaltz *et al.* (2011)] were shown to cause substantial nest desertion (flooding nests and eggs) and were likely the cause of the failure of early incubating birds. The peak seen in the proportion of successful incubating birds suggests that a second wave of incubating birds settled after the rainy period, with most of these birds being successful. As birds with failed incubation may make a second attempt, the second wave may have been partially made up of renesters. We also computed  $E(N_{21}|o_1o_2...o_T)$  the mean number of times that state 2: "incubation with failure" was followed by state 1: "elsewhere/preincubating". The estimated value for  $E(N_{21}|o_1o_2...o_T)$  was 1.1 [0;2.2] showing that an individual left the incubation area once on average. This result strengthens the proposed model which does not differentiate between individuals arriving for the first time or not in the incubation area.

### Extension to Models Incorporating Several Semi-Markovian States

The models in section 2 can be extended by incorporating several semi-Markovian states. In the following, we give several potential applications with more than one semi-Markovian state. If more than one semi-Markov state is incorporated, the approach used for building the

**Table 1.** Estimated transition probabilities in the incubation model.  $\phi_{ij}^{(t=k:l)}$  means transition between state  $i$  and state  $j$  for time periods grouping intervals  $k$  to  $l$ . MLE stands for maximum likelihood estimate, CI for confidence interval and SE for standard error. NA stands for nonavailable.

Parameters	MLE	CI	SE
$\phi_{1,1}^{(t=1:3)}$	0.720	(0.637, 0.790)	0.039
$\phi_{1,3}^{(t=1:3)}$	0.012	(0.004, 0.036)	0.007
$\phi_{2,1}^{(t=1:3)}$	0.886	(0.727, 0.958)	0.055
$\phi_{2,3}^{(t=1:3)}$	0.112	(0.041, 0.271)	0.055
$\phi_{1,1}^{(t=4:6)}$	0.840	(0.813, 0.864)	0.013
$\phi_{1,3}^{(t=4:6)}$	0.079	(0.062, 0.100)	0.010
$\phi_{2,1}^{(t=4:6)}$	0.312	(0.221, 0.419)	0.051
$\phi_{2,3}^{(t=4:6)}$	0.099	(0.072, 0.135)	0.016
$\phi_{3,5}^{(t=5:6)}$	0.651	(0.317, 0.883)	0.161
$\phi_{4,5}^{(t=6)}$	0.00	NA	NA
$\phi_{4,6}^{(t=6)}$	0.308	(0.027, 0.875)	0.300
$\phi_{5,6}^{(t=6)}$	0.00	NA	NA
$\phi_{1,1}^{(t=7)}$	0.686	(0.613, 0.750)	0.035
$\phi_{1,3}^{(t=7)}$	0.300	(0.237, 0.372)	0.034
$\phi_{2,1}^{(t=7)}$	0.681	(0.554, 0.785)	0.059
$\phi_{2,3}^{(t=7)}$	0.300	(0.198, 0.426)	0.059
$\phi_{3,5}^{(t=7:8)}$	0.864	(0.677, 0.951)	0.066
$\phi_{4,5}^{(t=7:8)}$	0.00	NA	NA
$\phi_{4,6}^{(t=7:8)}$	.594	( 0.373, 0.783)	0.111
$\phi_{5,6}^{(t=7:8)}$	1.00	NA	NA
$\phi_{1,1}^{(t=8)}$	0.722	(0.643, 0.790)	0.037
$\phi_{1,3}^{(t=8)}$	0.261	(0.196, 0.338)	0.036
$\phi_{2,1}^{(t=8)}$	0.00	NA	NA
$\phi_{2,3}^{(t=8)}$	0.00	NA	NA
$\phi_{1,1}^{(t=9:10)}$	0.835	(0.754, 0.894)	0.035
$\phi_{1,3}^{(t=9:10)}$	0.013	(0.005, 0.033)	0.006
$\phi_{2,1}^{(t=9:10)}$	0.873	(0.754, 0.940)	0.046
$\phi_{2,3}^{(t=9:10)}$	0.126	(0.060, 0.246)	0.046
$\phi_{3,5}^{(t=9:18)}$	0.327	(0.278, 0.381)	0.026
$\phi_{4,5}^{(t=9:18)}$	0.567	(0.530, 0.603)	0.019
$\phi_{4,6}^{(t=9:18)}$	0.00	NA	NA
$\phi_{5,6}^{(t=9:18)}$	0.049	(0.023, 0.098)	0.018
$\phi_{1,1}^{(t=11:15)}$	1.00	NA	NA
$\phi_{1,3}^{(t=11:15)}$	0.00	NA	NA
$\phi_{2,1}^{(t=11:15)}$	0.597	(0.477, 0.706)	0.059
$\phi_{2,3}^{(t=11:15)}$	0.00	NA	NA
$\phi_{1,1}^{t=16}$	1.000	(1.000, 1.000)	0.000
$\phi_{2,1}^{t=16}$	1.000	(1.000, 1.000)	0.000

likelihood 1 becomes intractable. Guédon (2005) developed a specific expectation-maximization (EM) algorithm for models with semi-Markovian states. In the following, we give several potential applications with more than one transient semi-Markovian state. We distinguish two cases:

- semi-Markovian state visited at most once (this semi-Markovian state constitutes a transient class);
- semi-Markovian state visited more than once (this semi-Markovian state belongs to a transient class made up of more than one state).

**Table 2.** Estimated capture probabilities in the incubation model.  $p_i^{(t=k:l)}(j)$  means capture probability for observation  $j$  conditional on being in state  $i$  for time periods grouping occasions  $k$  to  $l$ . MLE stands for maximum likelihood estimate, CI for confidence interval and SE for standard error. NA stands for non-available.

Parameters	MLE	CI	SE
$b_2^{(t=4:7)}(1)$	0.660	(0.521, 0.774)	0.066
$b_3^{(t=4:7)}(1)$	0.182	(0.145, 0.225)	0.020
$b_4^{(t=4:7)}(2)$	0.744	(0.290, 0.953)	0.190
$b_5^{(t=4:7)}(3)$	0.00	NA	NA
$b_2^{(t=8)}(1)$	1.00	NA	NA
$b_3^{(t=8)}(1)$	0.124	(0.097, 0.156)	0.015
$b_4^{(t=8)}(2)$	0.287	(0.097, 0.603)	0.139
$b_5^{(t=8)}(3)$	0.00	NA	NA
$b_2^{(t=9)}(1)$	0.761	(0.227, 0.972)	0.221
$b_3^{(t=9)}(1)$	0.267	(0.233, 0.304)	0.018
$b_4^{(t=9)}(2)$	0.360	(0.124, 0.690)	0.162
$b_5^{(t=9)}(3)$	0.00	NA	NA
$b_2^{(t=10:20)}(1)$	0.729	(0.546, 0.857)	0.081
$b_3^{(t=10:20)}(1)$	0.710	(0.680, 0.739)	0.015
$b_4^{(t=10:20)}(2)$	0.875	(0.816, 0.917)	0.025
$b_5^{(t=10:20)}(3)$	0.088	(0.076, 0.102)	0.006

### Potential applications with semi-Markovian states visited at most once

Lorentzen *et al.* (2012) estimated hatching success from photo series data in a cliff-nesting seabird. The model comprised two states: "egg" and "chick" and made the assumption that the incubation process started the first time an egg was seen. This is obviously prone to errors because an egg viewed for the first time may have been missed before. In this particular case, we could use one Markovian state "no egg" and two semi-Markovian states: "egg" and "chick". Any information on the time spent in each "egg" and/or "chick" state could also be incorporated.

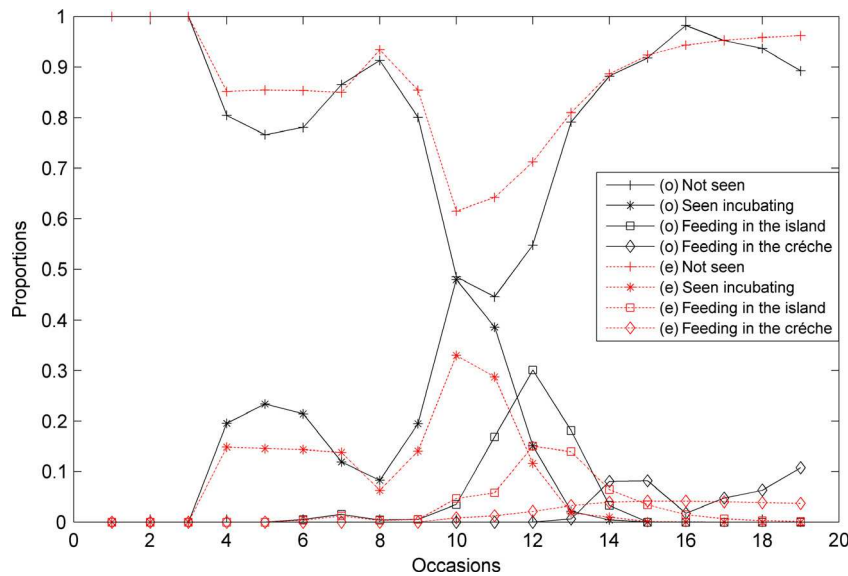
As heterogeneity is very common in ecological data, it might be relevant to consider two or more classes, each with a specific duration. In this case, the generalization of the likelihood (1) is quite straightforward.

### Potential applications with semi-Markovian state visited more than once

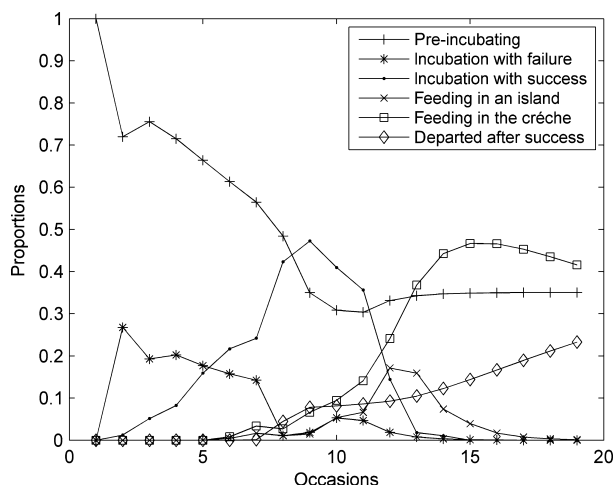
A hidden hybrid Markov/semi-Markov model has been used by Guédon (2005) to analyze branching and flowering patterns in plants. In this case, the states corresponded to the successive homogeneous branching and axillary flowering zones along a shoot, and the sojourn time distributions to the lengths of each successive zone.

In plant dynamics, the probability of finding species  $i$  in a given quadrat at time  $t$  may depend not only on time





**Figure 3.** Observed (in black) vs estimated (in red) proportions of individuals for each possible output at each capture occasion (from 1 to 19).



**Figure 4.** Estimated proportions of individuals in each state at each capture occasion (from 1 to 19).

$t-1$  but also on times before (Van Hulst 1979). We can wonder whether knowledge about the past is required and how much time the species  $i$  will remain in that quadrat. This question may be better considered using semi-Markovian states.

Biologic cycles are made up of different stages which may have different durations. This is the case for ovarian cycle in cattle (O’Connell *et al.* 2011). We should also mention here that measurements were used in O’Connell *et al.* (2011) as observations given that progesterone concentrations were the measurements from which the state of an individual was deduced. This may also be the case in eco-epidemiology (Choquet *et al.* 2013b) where the

estimated duration of an illness may be time since being ill dependent. As pointed out by Zipkin *et al.* (2010), care should be taken regarding the time scale over which the disease occurs. As an important remark, we note that we might condition on the first capture for eco-epidemiology CR data. As an important remark, we note that we might condition on the first capture for eco-epidemiology CR data.

### Conclusion

In this article, we have developed a new framework for capture–recapture data that can be used to explicitly model the time spent in one state. This formulation is well suited to dealing with imperfect and variable detection. These new models provide a generalization of hidden Markov models which are increasingly used to analyze capture–recapture data. We believe that this framework has many potential applications (see the previous section).

However, work remains to be performed on this kind of model in order to render it more accessible to ecologists. Concerning model estimation, Guédon (2003, 2005) proposed an efficient algorithm for hidden hybrid Markov/semi-Markov models potentially incorporating several successive semi-Markovian states. Chaubert-Pereira *et al.* (2010) extended these algorithms to models incorporating fixed and random effects in the observation process. O’Connell and Hojsgaard (2011) developed an R-package for hidden semi-Markov models. And we implemented the likelihood (1) in E-SURGE (Choquet *et al.* 2009). However, in a broader framework, we need not only to

adapt these algorithms to the specificity of CR data, but also design a software package.

## Acknowledgments

We greatly thank Lucie Schmalz and Christophe Germain for providing us qualitative information on the flamingo database. We thank the associate editor, Elena Machetou, and one anonymous reviewer for very helpful comments on an earlier version of the article. This research was supported by a BIOSTIC project ‘Techniques markoviennes avancées en dynamique des populations’ financed by the Montpellier sud de France University.

## Conflict of Interest

None declared.

## References

- Akaike, H. 1987. Factor analysis and AIC. *Psychometrika* 52: 317–332.
- Amat, J. A., F. Hortas, G. M. Arroyo, M. A. Rendón, J. M. Ramirez, and M. Rendon-Martos. 2007. Interannual variations in feeding frequencies and food quality of greater flamingo chicks (*phoenicopterus roseus*): evidence from plasma chemistry and effects on body condition. *Comp. Biochem. Physiol. Physiol.* 147:569–576.
- Booth, A. M., E. O. Minot, R. A. Fordham, and M. J. Imber. 2000. Co-ordinated food provisioning in the little shearwater *Puffinus assimilis haurakiensis*: a previously undescribed foraging strategy in the procellariidae. *Ibis* 142: 144–146.
- Bryant, A. A., and R. E. Page. 2005. Timing and causes of mortality in the endangered vancouver island marmot (*Marmota vancouverensis*). *Can. J. Zool.* 83: 674–682.
- Chaubert-Pereira, F., Y. Guédon, C. Lavergne, and C. Trottier. 2010. Markov and semi-Markov switching linear mixed models used to identify forest tree growth components. *Biometrics* 66: 753–762.
- Choquet, R., L. Rouan, and R. Pradel. 2009. Program E-SURGE: A software application for fitting multievent models. Pp. 845–865. *in* D. L. Thomson, E. G. Cooch and M. J. Conroy, eds. *Modeling demographic processes in marked populations*, volume 3 of Springer series: environmental and ecological statistics environmental and ecological statistics. Springer, New York.
- Choquet, R., A. Viallefont, L. Rouan, K. Gaanoun, and J. M. Gaillard. 2011. A semi-Markov model to assess reliably survival patterns from birth to death in free-ranging populations. *Methods Ecol. Evol.* 2: 383–389.
- Choquet, R., C. Carrié, T. Chambert, and T. Boulinier. 2013a. Estimating transitions between states using measurements with imperfect detection: application to serological data. *Ecology* 94: 2160–2165.
- Choquet, R., Y. Guédon, A. Besnard, M. Guillemain, and R. Pradel. 2013b. Estimating stop over duration in presence of trap-effects. *Ecol. Modell.* 250: 111–118.
- Colchero, F., and J. S. Clark. 2012. Bayesian inference on age-specific survival for censored and truncated data. *J. Anim. Ecol.* 81: 139–149.
- Dennis, J., and R. Schnabel. 1983. Numerical methods for unconstrained optimization and nonlinear equations. *Classics in applied mathematics*. SIAM, Englewood Cliffs.
- Dietz, M. W., K. G. Rogers, and T. Piersma. 2013. When the seasons don't fit: Speedy molt as a routine carry-over cost of reproduction. *PLoS ONE* 8: e53890.
- Fewster, R., and N. Patenaude. 2009. Cubic splines for estimating the distribution of residence time using individual resightings data. Pp. 393–415. *in* D. L. Thomson, E. G. Cooch and M. J. Conroy, eds. *Modeling demographic processes in marked populations*, Vol. 3. Springer, New York.
- Guédon, Y. 1999. Computational methods for discrete hidden semi-Markov chains. *Appl. Stoch. Model. Bus.* 15: 195–224.
- Guédon, Y. 2003. Estimating hidden semi-Markov chains from discrete sequences. *J. Comput. Graph. Stat.* 12: 604–639.
- Guédon, Y. 2005. Hidden hybrid Markov/semi-Markov chains. *Comput. Stat. Data Anal.* 49: 663–688.
- Hamel, S., J. M. Gaillard, N. G. Yoccoz, A. Loison, C. Bonenfant, and S. Descamps. 2010. Fitness costs of reproduction depend on life speed: empirical evidence from mammalian populations. *Ecol. Lett.* 13: 915–935.
- Johnson, J., and F. Cezilly. 2007. *The Greater Flamingo*. T. and A. D. Poyser, London.
- Kulkarni, V. 1995. *Modeling and analysis of stochastic systems*. Chapman and Hall, London.
- Langrock, R., and W. Zucchini. 2011. Hidden Markov models with arbitrary state dwell-time distributions. *Comput. Stat. Data Anal.* 55: 715–724.
- Lebreton, J. D., K. Burnham, J. Clobert, and D. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecol. Monogr.* 62: 67–118.
- Link, W. A., and R. J. Barker. 2005. Modeling association among demographic parameters in analysis of open population capture–recapture data. *Biometrics* 61: 46–54.
- Lorentzen, E., R. Choquet, and H. Steen. 2012. Modelling state uncertainty with photo series data for the estimation of breeding success in a cliff-nesting seabird. *J. Ornithol.* 152: 477–483.
- Matechou, E., S. Pledger, M. Efford, B. J. T. Morgan, and D. L. Thomson. 2013. Estimating age-specific survival when age is unknown: open population capture–recapture models with age structure and heterogeneity. *Methods Ecol. Evol.* 4: 654–664.
- O'Connell, J., and S. Hojsgaard. 2011. Hidden semi-Markov models for multiple observation sequences: The mhsmm package for R. *J. Stat. Softw.* 39: 1–22.

- O'Connell, J., F. Torgersen, N. Friggens, P. Lovendahl, and S. Hojsgaard. 2011. Combining cattle activity and progesterone measurements using hidden semi-Markov models. *J. Agric. Biol. Environ. Stat.* 16: 1–16.
- Pianka, E. R. 1970. R-selection and k-selection. *Am. Nat.* 104: 592–597.
- Pledger, S.M. G., Efford, K. H. Pollock, J. Collazo, and J. Lyons. 2009. Stopover duration analysis with departure probability dependent on unknown time since arrival. Pp. 349–363 *in* D. L. Thomson, E. G. Cooch and M. J. Conroy, eds. *Modeling demographic processes in marked populations*, volume 3 of Springer series: environmental and ecological statistics. Springer, New York.
- Pradel, R. 1996. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics* 52: 703–709.
- Pradel, R. 2009. The stakes of capture–recapture models with state uncertainty. Pp. 781–795 *in* D. L. Thomson, E. G. Cooch and M. J. Conroy, eds. *Modeling demographic processes in marked populations*, volume 3 of Springer series: environmental and ecological statistics. Springer, New York.
- Saether, B. E., S. Engen, A. P. Moller, H. Weimerskirch, M. E. Visser, and W. Fiedler. 2004. Life-history variation predicts the effects of demographic stochasticity on avian population dynamics. *Am. Nat.* 164: 793–802.
- Schmaltz, L., F. Cézilly, and A. Béchet. 2011. Using multistate recapture modelling to assess age-specific bottlenecks in breeding success: a case study in the greater flamingo *Phoenicopterus roseus*. *J. Avian Biol.* 42: 1–9.
- Shaman, J. and A. Karspeck. 2012. Forecasting seasonal outbreaks of influenza. *Proc. Natl Acad. Sci. USA* 109: 20425–20430.
- Van Hulst, R. 1979. On the dynamics of vegetation: Markov chains as models of succession. *Vegetatio* 40: 3–14.
- Zipkin, E. F., C. S. Jennelle, and E. G. Cooch. 2010. A primer on the application of Markov chains to the study of wildlife disease dynamics. *Methods Ecol. Evol.* 1: 192–198.