



## Estimating demographic parameters using hidden process dynamic models

Olivier Gimenez, Jean-Dominique Lebreton, Jean-Michel Gaillard, Rémi Choquet, Roger Pradel

### ► To cite this version:

Olivier Gimenez, Jean-Dominique Lebreton, Jean-Michel Gaillard, Rémi Choquet, Roger Pradel. Estimating demographic parameters using hidden process dynamic models. *Theoretical Population Biology*, 2012, 82, pp.307 - 316. 10.1016/j.tpb.2012.02.001 . hal-03499089

**HAL Id: hal-03499089**

**<https://hal.science/hal-03499089>**

Submitted on 24 Dec 2021

**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.

**Estimating demographic parameters using hidden process dynamic models**

Olivier Gimenez<sup>1</sup>, Jean-Dominique Lebreton<sup>1</sup>, Jean-Michel Gaillard<sup>2</sup>, Rémi Choquet<sup>1</sup>, and  
Roger Pradel<sup>1</sup>

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

<sup>2</sup> Laboratoire de Biométrie et Biologie Evolutive, Unité Mixte de Recherche 5558, Université  
Claude Bernard Lyon I, 43 Bd du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Corresponding author:

Olivier Gimenez, [olivier.gimenez@cefe.cnrs.fr](mailto:olivier.gimenez@cefe.cnrs.fr)

UMR 5175 – Centre d'Ecologie Fonctionnelle et Evolutive  
1919, Route de Mende – 34293 Montpellier cedex 5, France

Tel: +33 (0)4 67 61 32 11

Fax: +33 (0)4 67 61 33 36

## Summary

Structured population models are widely used in plant and animal demographic studies to assess population dynamics. In matrix population models, populations are described with discrete classes of individuals (age, life history stage or size). To calibrate these models, longitudinal data are collected at the individual level to estimate demographic parameters. However, several sources of uncertainty can complicate parameter estimation, such as imperfect detection of individuals inherent to monitoring in the wild and uncertainty in assigning a state to an individual. Here, we show how recent statistical models can help overcome these issues. We focus on hidden process models that run two time series in parallel, one capturing the dynamic of the true states and the other consisting of observations arising from these underlying possibly unknown states. In a first case study, we illustrate hidden Markov models with an example of how to accommodate state uncertainty using Frequentist theory and maximum likelihood estimation. In a second case study, we illustrate state-space models with an example of how to estimate lifetime reproductive success despite imperfect detection, using a Bayesian framework and Markov Chain Monte Carlo simulation. Hidden process models are a promising tool as they allow population biologists to cope with process variation while simultaneously accounting for observation error.

**Keywords:** capture-recapture; hidden Markov models; lifetime reproductive success; multievent models; multistate models; state-space models; state uncertainty

## 1. INTRODUCTION

Structured population models (Tuljapurkar & Caswell 1997; Caswell 2001) allow a detailed description of population change over time by accounting for the number of individuals in discrete classes (grouped according to age, stage or state). These models are widely used in plant and animal demographic studies to assess population dynamics as they allow the projection of future population size, estimation of asymptotic population growth rate, age structure, net reproductive rate, generation time and life expectancy among other descriptive statistics. Matrix models have been used to assess population viability (e.g., Ghimire et al. 2008), evaluate the effects of management strategies on life history strategies (e.g., Servanty et al. 2011) and investigate the feedback between demography and evolution (van Tienderen 2000, see also Gamelon et al. [2011] for a recent application).

Traditionally, structured population models are calibrated using longitudinal data collected at the individual level. These data are analyzed with statistical models to estimate demographic parameters like survival, growth, dispersal and fecundity, which are then entered into structured population models. However, several sources of uncertainty are associated with parameter estimation that need to be properly accounted for. First, individuals are difficult to monitor in the wild (mobile animals in particular) because they may or may not be observed at a given sampling occasion. This raises the issue of imperfect detection, which, if ignored, can lead to biased estimates and flawed inference about demographic parameters (e.g., Nichols 1992, Gimenez et al. 2008). Second, there can be uncertainty related to measurement error, in particular, correctly assigning a state to an individual, even when the individual is observed or captured. This may be the gender of an individual whenever sex is determined through behavioral clues, or epidemiological status if serological tests generate false positives or negatives, or reproductive status whenever the number of offspring is visually determined by counting the number of young accompanying their parents.

To handle the complexity of the dynamics of free-ranging populations and accommodate the various sources of uncertainty associated with parameter estimation, there is a need to integrate structured population models with novel statistical tools. In this paper, we aim to review and demonstrate how recent statistical models can help bring theoretical models (mechanisms) and the analysis of empirical data (patterns) together. We will focus our attention on hidden process models (see Box 1) that have recently received much attention in the literature (Buckland et al. 2004; 2007; Newman et al. 2006). In hidden process models, two time series are run in parallel: one captures the dynamic of the true (latent) states via a stochastic process model; the other is static and describes the observations that are made from these underlying but possibly unknown states.

In the first section, we review standard multistate capture-recapture models (see Box 1; Lebreton et al. 2009) that are used to estimate key demographic parameters, which are then fed to structured population models. We pay particular attention to the imperfect detection of individuals in the wild for analyzing individual longitudinal data. In the second and third sections, we reformulate longitudinal multistate models as hidden process models, either as hidden Markov models (HMM; see Box 1, Pradel 2005) or state-space models (SSM; see Box 1, Gimenez et al. 2007). In both sections, case studies are used to illustrate these models. In the second section, we show how hidden-Markov models can be used to deal with uncertainty in the assignment of reproductive status and its importance when assessing life-history trade-offs. In the third section, we show how state-space models can be used to estimate lifetime reproductive success (sensu Clutton-Brock 1988) while accounting for a detection probability less than one. Finally, we discuss the limitations of hidden process models as well as several extensions of these models.

## **2. OVERVIEW OF MULTISTATE CAPTURE-RECAPTURE MODELS**

Capture-recapture (CR; see Box 1) models have become a central tool in population biology for estimating demographic parameters under imperfect detection of individuals (Lebreton et al. 1992). CR models were originally designed to estimate population size, with demographic parameters such as survival probabilities progressively incorporated in these models, first as nuisance parameters, and eventually as the main focus (for the early history see Lebreton et al. 2009 p. 93 ff). The standard Cormack-Jolly-Seber model (see Box 1; Cormack 1964, Jolly 1965, Seber 1965) is based on the embedding of two distinct processes: survival (with survival probabilities denoted as  $\phi$ ) and detection (with recapture probabilities denoted as  $p$ ) over discrete time occasions. These parameters are indexed appropriately depending on the type of variation of interest (e.g., time, age or sex; Lebreton et al. 1992).

A notable feature of the original development of CJS models is that the survival and the detection processes were viewed as intimately entangled. Adequately representing how they were imbricated was indeed at the core of early methodological developments to obtain a likelihood that could be used to properly estimate the parameters. For instance, under time-dependence, the probability that an individual is never seen again after occasion  $i$ , denoted as  $C_i$ , is obtained through a backward recurrence relationship (Cormack, 1964), accounting for the fact that an individual may die between occasion  $i$  and  $i+1$  or survive and not be detected:

$$C_i = 1 - \hat{f}_i + \hat{f}_i(1 - \hat{p}_{i+1})C_{i+1}$$

with  $C_{\text{last occasion}} = 1$ .

However, when comparing various types of detection, it clearly appeared that the survival-mortality process was a background process distinct from the detection process. For instance for the same time-dependent survival process, one must be able to accommodate several situations illustrated in Figure 1 (see also Lebreton et al. 1999): i) complete detection, as might occur in developed countries for the human population (with administrative declaration of

1 deaths), ii) incomplete detection of live individuals, i.e., live recaptures or resightings, to be  
2 analyzed with CJS models or iii) incomplete detection of deaths, as in ‘dead recoveries’ of  
3 banded birds (usually by the general public) to be analyzed with so called dead-recovery  
4 models (Brownie et al. 1985).

5  
6 [Figure 1 about here]  
7

8 In the survival process, death or survival between two sampling occasions is treated as strictly  
9 conditional on the fact that the individual is alive at the onset of the target interval: the system  
10 has no memory, i.e. the survival process is a Markov chain. Hence, CJS models are implicitly  
11 made up of a Markov chain on top of which is an imperfect observation process (e.g., Lebreton  
12 et al. 1999). In other words, CJS models are hidden process models where the demographic  
13 process is represented by a (first-order) Markov chain (see next sections) that is only partially  
14 observed due to imperfect detection. Conveniently, the explicit representation of the survival  
15 process in CJS models as a two-state Markov chain (alive / dead) readily provides the  
16 probability of individual histories, without having to use the complicated  $C$  parameters  
17 mentioned above (Caswell and Fujiwara 2004, pp. 477-478).

18 However, representation of the survival process as a Markov chain did not capture much  
19 attention for some time for a number of reasons. First, CR models were considered to be  
20 specific and exotic statistical tools, in particular because of the historical focus on population  
21 size estimation and the associated difficulties in producing an adequate fully stochastic model  
22 (Jolly 1965). The focus on the estimation of demographic flows developed only in the late 80s  
23 (Burnham et al. 1987; Clobert and Lebreton , 1987). Second, while the succession of events  
24 such as survival and death in CR models was clearly Markovian, historically, the statistical  
25 point of views on CR models and Markov chains were radically different: when considering  $n$

1 individuals moving over  $T$  time steps according to a Markov chain, the development of  
2 stochastic process theory for Markov chains was centered on asymptotic results when  $T$  tends  
3 to infinity, such as ergodicity for instance; on the contrary, CR models asymptotic results, such  
4 as the optimality of maximum likelihood estimators for example, are derived for  $n$  tending to  
5 infinity for a fixed value of  $T$ . Third, parameter estimation for Markov chains under partial  
6 observation was not fully developed at that time (e.g., Lawless and McLeish 1984) and could  
7 not offer solutions to the CR situation.

8 The initially slow development of the Markov chain approach to analyzing individual histories  
9 is illustrated in Figure 2. The first papers explicitly using this point of view were published in  
10 the early 1990s, and from there on, parallel growth in human and animal population studies  
11 took place at a common rate of  $\sim 9\%$  a year.

12  
13 [Figure 2 about here]  
14

15 Modeling survival as a hidden process has several strong advantages. First, the process model  
16 can easily be made more complex or detailed to address new biological questions. Existing CR  
17 extensions, such as multisite (Arnason 1972, 1973; Hestbeck et al. 1991) and recruitment  
18 (Clobert et al. 1994) models, entered this common framework; new generalizations could  
19 easily be produced, such as multisite recruitment models (Lebreton et al. 2003) and multistate  
20 CR models (reviewed in detail by Lebreton et al. (2009), and Williams et al. 2002 pp. 454-  
21 468). Second, it linked CR statistical models with parallel developments in population  
22 dynamics models and notably matrix population models (Caswell 2001 p. 37). These models  
23 are based on a reasonable representation of differences among individuals, through a finite  
24 number of so-called ‘i-state variables’ that summarize relevant information on the past life of  
25 an individual to determine its future in terms of, e.g., individual physiology, reproduction or  
26 death (Diekmann and Heesterbeek 2000). The same idea is clearly inherent in multistate CR



models. The empirical use of these models requires that individual data should translate into parameter estimates, i.e. estimates of probabilities of the transition from, say, life to death, rosette to flowering plant or non-breeder to breeder. Several papers have addressed estimation and the use of estimates in multistate dynamics models (e.g., Caswell and Fujiwara 2004). Third, the observation process can also be made more complex and tuned to specific situations. Dupuis (1995) was among the first to distinguish clearly between the two processes in CR models, considering the transitions between states as a Markov chain and imperfect detection as an independent process generating ‘missing data’. Soon, however, the detection process was considered to be potentially more complex, e.g., with the development and systematic exploration of the use of unobservable states for modeling temporary emigration (Kendall and Bjorkland 2001) and recruitment (Pradel and Lebreton 1999). Specific models for ‘mixtures of information’ (Burnham 1993) considered detection of individuals either as alive or at the time of death. They appear as specific multistate CR models based on the two states alive / just dead (Lebreton et al. 1999). The generalization of the observation process currently includes models that account for uncertainty in the assessment of states (Kendall et al. 2003). The observation process then considers a probabilistic relationship between the underlying states (such as breeder vs. non-breeder) and ‘events’ (such as seen at nest or seen out of a nest). Hence, an individual seen out of a nest may be in either state. Earlier approaches were specific to permanent states (such as sex, Nichols et al. 2004), but the more general models are multistate CR models that account for state uncertainty, so-called multievent models (Pradel 2005). The movement towards understanding and unifying CR models as hidden process models (Figure 3) has followed rapidly with the development of specific statistical tools for hidden process models, either hidden-Markov models or state-space models. We review these developments in the following next two sections.

[Figure 3 about here]

### 3. HIDDEN MARKOV MODELS: DEALING WITH UNCERTAINTY IN STATE ASSIGNMENT

#### 3.1. *Hidden-Markov modeling*

##### 3.1.1. HMM for capture-recapture data

In the above multistate CR models, it is assumed that the state of an individual is ascertained upon encounter. This poses no difficulty if the state is a geographical location (e.g., wintering site or breeding colony e.g.), but is less obvious when the state is reproductive status (breeder or non-breeder) or a physiological state like diseased vs. healthy. Diagnosis is typically attempted based upon some criterion like the presence of a brood patch or nest-building activity (for a breeder) or antibody level (for a potentially diseased individual). The distinction between the criterion and the state leads to models that distinguish the visible layer (the criterion) from the hidden layer (the state). This distinction between what is observed and what is inferred is key to hidden process models. Because the hidden layer is a first order Markov chain, this naturally leads to the introduction of a HMM for analyzing CR data. The key step is to recognize that whether an individual is observed or not may itself be considered as belonging to the ‘observable’ layer. This conceptual leap led to the definition of multievent CR models (see Box 1; Pradel 2005), which both generalize multistate CR models to accommodate uncertainty in state assignment and firmly establish that CR models are a particular type of HMMs. However, unlike in HMMs, the recorded information in multievent models is named ‘events’ rather than ‘observations’, because the record can be a ‘non-observation’.

##### 3.1.2. Review of HMM applications to estimate demographic parameters

HMMs have a tremendous potential with applications in all domains of ecology (Pradel 2009), but they historically have been seldom used. An obvious application is in epidemiology (Conn & Cooch 2009). Another is sex identification in monomorphic or weakly polymorphic species where clues are mainly derived from behavior (Nichols et al. 2004; Pradel et al. 2010). A less obvious application is investigating the role of memory in movements among a discrete set of sites. When movements between the successive occasions are determined not only by the position occupied at a particular occasion but also by some previous positions, the previous positions must enter in the definition of the state. Then, when an individual is encountered, it may happen that because the previous relevant positions have not been observed, the exact state is unknown. This invalidates the use of multistate CR models but HMMs can be used (Rouan et al. 2009). The most elaborate application to date makes use of information from nest inspection to study the probability of skipping behavior and nest changing in a colony of Cory's shearwater (Sanz-Aguilar et al. 2011). A more methodological example is the incorporation of mark loss in mark-recapture models during a double-mark study (Juillet et al. 2011). Another interesting application copes with the issue of keeping track of the number of breeding episodes when detectability is less than one (Desprez et al. 2011; Pradel et al. 2011). It is also possible to account for tag loss and different causes of mortality (Tavecchia et al. 2012). However, by far, the greatest number of applications relates to class heterogeneity whereby a given individual belongs to a particular unknown class for the duration of the study. Cubaynes et al. (2010) emphasize the importance of explicitly modeling capture heterogeneity to avoid potentially large biases in the estimation of population size (see Ebert et al. 2010 for a review) and Crespin et al. (2008) point out that heterogeneity in capture rate may mirror social hierarchy. Véran et al. (2007) and Oliver et al. (2011) used capture heterogeneity to correct for unequal sampling over space. Péron et al. (2010a) went further by modeling class heterogeneity in survival, detection and movement simultaneously and assessed its influence on the study of senescence. Péron et al. (2010b) analyzed in further depth the significance of

movement class heterogeneity within a population. Given all these examples, we were surprised to not find studies of reproduction that account for uncertainty in the breeding status. Identification of reproductive status is not always obvious and failing to acknowledge this problem may lead to mistakes or at least to a loss of useful information. This issue is particularly important in studies of life-history trade-offs. Here we develop an example of how such data could be handled.

### 3.2. Case study 1: Dealing with uncertain states

#### 3.2.1. Data and models

Data were collected between 1940 and 1957 by Lance Richdale on Sooty shearwaters *Puffinus griseus* (titis hereafter) and were reanalyzed using multistate CR models by Scofield et al. (2001; see also Gimenez et al. 2005).

Following the way the data were collected, four states were originally considered: (1) Breeder (B); (2) Keeping company with another bird in a burrow; (3) Alone in a burrow; (4) On the surface. Because of numerical issues, Scofield et al. (2001) pooled 2-3-4 together in a Non-Breeder state (NB) that includes failed breeders (birds that had bred previously but were currently skipping reproduction or divorced) and pre-breeders (birds that had not yet bred). Note that because burrows were not checked before hatching, some birds in the category NB might have bred and failed before being checked. We therefore regard those birds in the B state as successful breeders, and those in the NB state as a mixture of non-breeders, pre-breeders and failed breeders.

We start by fitting a model with state-specific survival, detection and transition probabilities between states, with all parameters constant over time. The states were {non-breeder, breeder, dead} = {NB, B, D}, and the events were {non-detected, detected as not a successful breeder, detected as a successful breeder} = {0, 1, 2}. Upon its first encounter, an individual has a probability  $\pi^B$  to be a breeder and the complement probability to be a non-

breeder. Then, it moves among the states according to a first-order Markov chain that can be described by the state specific survival probabilities  $\phi^{NB}$  and  $\phi^B$  and the transition probabilities conditional on survival  $\psi^{NB \rightarrow B}$  and  $\psi^{B \rightarrow NB}$ . The state process can be represented by the transition matrix with departure states in rows and arrival states in columns:

$$\begin{bmatrix} f^{NB} \psi^{NB \rightarrow NB} & f^{NB} \psi^{NB \rightarrow B} & 1 - f^{NB} \\ f^B \psi^{B \rightarrow NB} & f^B \psi^{B \rightarrow B} & 1 - f^B \\ 0 & 0 & 1 \end{bmatrix}.$$

Equivalently, it can be represented as the product of a survival matrix by a conditional transition matrix:

$$\begin{bmatrix} f^{NB} & 0 & 1 - f^{NB} \\ 0 & f^B & 1 - f^B \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \psi^{NB \rightarrow NB} & \psi^{NB \rightarrow B} & 0 \\ \psi^{B \rightarrow NB} & \psi^{B \rightarrow B} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The event process conditional on underlying states is described via the detection probabilities  $p^{NB}$  and  $p^B$ . This event can be summarized by a row-stochastic matrix (i.e., probabilities in a row sum up to one) with states in rows and events in columns:

$$\begin{bmatrix} 1 - p^{NB} & p^{NB} & 0 \\ 1 - p^B & 0 & p^B \\ 1 & 0 & 0 \end{bmatrix}.$$

For instance, in row 1 column 2 is the probability that a non-breeder is detected as not a successful breeder  $p^{NB}$ .

To demonstrate how uncertain detection of reproductive status could be analyzed, we generated uncertainty in the assignment of Non-Breeder vs. Breeder as follows. For each individual at each detection occasion: a Non-Breeder was assumed to be judged a Non-Breeder with probability 0.2 (and assigned to event 1), and not judged with the complementary probability 0.8 (and assigned to the ambiguous event 3), while a Breeder was assumed to be judged a Breeder with probability 0.7 (and then assigned to event 2), and not judged with the complementary probability 0.3 (and assigned to event 3). This procedure was implemented in R (R Development Core Team 2011; see Appendix A2). To analyze these imperfect data, only the event process changes. In addition to the detection probabilities, we add state-specific probabilities of breeding state assignment  $\delta^{NB}$  and  $\delta^B$ . With the same presentation as before, the event process becomes:

$$\begin{array}{c} \begin{array}{ccccc} \begin{array}{c} \acute{e} \\ \grave{e} \\ \grave{e} \\ \grave{e} \\ \grave{e} \end{array} & \begin{array}{c} 1 - p^{NB} \\ 1 - p^B \\ 1 \\ 1 \\ 1 \end{array} & \begin{array}{c} p^{NB}d_{NB} \\ 0 \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} 0 \\ p^Bd_B \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} p^{NB}(1 - d_{NB}) \\ p^B(1 - d_B) \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} \grave{u} \\ \acute{u} \\ \acute{u} \\ \acute{u} \\ \acute{u} \end{array} \end{array} \end{array}.$$

Note that the new event ‘breeding state not ascertained’ (column 4) can arise for non-breeders (first row) as well as breeders (second row), though with potentially different probabilities. This matrix can also be written as a product of two matrices, highlighting the successive processes of detection and breeding state ascertainment:

$$\begin{array}{c} \begin{array}{ccccc} \begin{array}{c} \acute{e} \\ \grave{e} \\ \grave{e} \\ \grave{e} \\ \grave{e} \end{array} & \begin{array}{c} 1 - p^{NB} \\ 1 - p^B \\ 1 \\ 1 \\ 1 \end{array} & \begin{array}{c} p^{NB} \\ 0 \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} 0 \\ p^B \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} \grave{u} \\ \acute{u} \\ \acute{u} \\ \acute{u} \\ \acute{u} \end{array} \end{array} \begin{array}{c} \begin{array}{ccccc} \begin{array}{c} \acute{e} \\ \grave{e} \\ \acute{e} \\ \acute{e} \\ \acute{e} \end{array} & \begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} 0 \\ d_{NB} \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} 0 \\ 1 - d_{NB} \\ d_B \\ 1 - d_B \\ 1 - d_B \end{array} & \begin{array}{c} \acute{u} \\ \acute{u} \\ \acute{u} \\ \acute{u} \\ \acute{u} \end{array} \end{array} \end{array}.$$

Note that when the animal is first encountered, the capture process is not modeled because an

1 animal must be encountered at least once to enter the data set. But the reproductive state  
2 ascertainment remains valid. Models with or without imperfect assignment of reproductive  
3 state were implemented in the freely downloadable program E-SURGE (Choquet et al. 2009;  
4 see Appendix A1 for details).

5  
6 3.2.2. Results

7 The confidence intervals for the assignment probabilities include the true value, despite  
8 the considerable loss of information due to uncertain state, in particular for the non-breeders  
9 which are recognized as such only 20% of the time (Table 1). The estimates of other  
10 parameters remain similar to those of the original model. The most notable and logical change  
11 is a loss of precision. However, in a real study, there would probably be on the contrary a gain  
12 of precision because the data chosen for analysis with multistate models are those deemed most  
13 reliable. In contrast, the model with uncertainty would be run on the larger dataset that includes  
14 the less reliable data. For this reason, precision should be enhanced by explicitly dealing with  
15 breeding state uncertainty.

16  
17 [Table 1 about here]  
18

19 **4. STATE-SPACE MODELS: ESTIMATING LIFETIME REPRODUCTIVE SUCCESS**

20  
21 ***4.1. State-space modeling***

22 4.1.1. SSM for capture-recapture data

23 Parallel to the HMM formulation of CR models, there has been increasing use of state-  
24 space models (SSM) to handle data on marked individuals. State-space models are a  
25 generalization of hidden Markov models in that the states are not necessarily governed by a  
26 Markov process. In practice, parameters of SSMs have been estimated in a Bayesian

framework (using Markov Chain Monte Carlo [MCMC] simulation of the joint posterior distribution of parameters; see section 4.2.3) while HMM have been fit in a Frequentist framework using maximum likelihood methods. CR models were reformulated as SSMs by several authors (Gimenez et al., 2007; Royle, 2008; Schofield and Barker, 2008) to distinguish the underlying demographic process from the observation process.

#### 4.1.2. Use of SSMs to estimate demographic parameters data – a survey

The SSM formulation provides a great deal of flexibility and has led to substantial progress in modeling key demographic parameters. Link and Barker (2005) showed how to estimate correlations between demographic parameters such as survival and reproductive rates. Servanty et al. (2010) used this approach to estimate cause-specific mortalities and test for additivity in mortality. Other applications include comparisons among species (Papadatou et al., 2011), spatial structure (Péron et al., 2011) and the study of migration (Calvert et al., 2009; O’Hara et al., 2009). Demographic parameters can be estimated for species with complex life cycles using SSMs (Clark et al., 2005; Buoro et al., 2010). Like HMM, important developments included the incorporation of individual heterogeneity via random effects (Schofield and Barker, 2011; Royle, 2008; Marzolin et al., 2011; see Gimenez and Choquet 2010 for an implementation of individual random effects in HMM). In addition, Gimenez et al. (2006; 2009) used SSM to estimate selection gradients and visualize fitness surfaces based on capture-recapture data, while Papaïx et al. (2010) showed how to estimate components of phenotypic variance that can be attributed to genetic factors, environmental factors and other unknown factors, hence quantifying heritability of demographic parameters. All these applications have used SSM to estimate demographic parameters. However, SSM also allows the estimation of the unknown latent states. To our knowledge, this has rarely been attempted, although it could be useful in reconstructing individuals’ fates and population dynamics (Gimenez et al., in prep). We develop an example of this below.



## 4.2. Case study 2: Measuring Lifetime Reproductive Success in the wild

### 4.2.1. Estimating fitness in the wild can be difficult

Lifetime Reproductive Success (LRS), the number of young measured at some stage by an individual during its lifespan is a single-generation measure of the individual's contribution to future generations (Clutton-Brock, 1988; Newton, 1989) and thus provides a reliable measure of individual fitness (see Brommer et al. 2004 for an empirical demonstration). The estimation of LRS requires information on the number of offspring an individual has produced and successfully raised at some stage (e.g., born, weaned or fledged, first reproducing) at each reproductive attempt. In theory then, exhaustive monitoring of individuals over their whole reproductive life is needed to estimate LRS. In the field, however, estimating LRS is difficult, because individuals are typically observed (or recaptured) only a fraction of their lifetime. In years when a given individual is not observed, whether it gives birth or not and how many offspring it produces at birth or at the end of the parental care period if it does breed remain unknown. In the following section, we show that, by considering two states (breeder vs. non-breeder) with the associated number of offspring produced in a SSM, reliable estimates of LRS can be obtained while accounting for detection probability less than one. A similar approach was proposed by Rouan et al. (2009b) who used HMM to compute LRS.

### 4.2.2. Data and models

We illustrate our approach using the reproductive histories of 211 female roe deer that were monitored in the Trois Fontaines forest, North-East France, between 1976 and 2006 (see e.g. McLoughlin et al. 2007 for further details about this monitoring). We developed a CR model to analyze these live recaptures of female roe deer. In line with recent work by Gimenez

et al. (2007), Royle (2008) and Schofield and Barker (2008), we used a state-space formulation of CR models that explicitly separates the demographic process of interest (survival status in a particular reproductive state) from the observations (the data). We considered five states: alive non-breeder (NB), alive and weaning one fawn (B1), alive and weaning two fawns (B2), alive and weaning three fawns (B3) and dead (D). We denoted  $X_{i,t}$  the random state vector taking values (1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1) if, at time  $t$ , individual  $i$  is in state NB, B1, B2, B3 or D, respectively. Based on the roe deer life history (see e.g. Gaillard et al. 1998), we considered several possible observations that were generated from these underlying states: the animal is not seen (coded 0), the animal is seen without any fawn (coded 1), the animal is seen with one fawn (coded 2), the animal is seen with two fawns (coded 3) and the animal is seen with three fawns (coded 4) in the dataset. We denoted  $Y_{i,t}$  the random observation vector taking values (1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1) if, at time  $t$ , individual  $i$  is not seen or seen with zero, one, two or three fawns respectively.

The model has several parameters. We modeled the probability  $f_{i,t}^k$  that an individual  $i$  alive at time  $t$  in state  $k = \text{NB}$  (non-breeder) or  $\text{B}$  (breeder whatever the number of fawns weaned) survives to time  $t + 1$ . Age-dependence in  $f_{i,t}^k$  was modeled using three classes: yearling (1 to 2 years), prime-aged (2 to 5 years) and old adults (6 years and older) according to previous survival analyses (see e.g. Gaillard et al. 1993). We also modeled the probability  $y_{i,t}^{k \rightarrow k'}$  that an individual  $i$  makes a transition between states  $k$  and  $k'$  ( $k, k' = \text{NB}, \text{B1}, \text{B2}$  or  $\text{B3}$ ) between occasion  $t$  and  $t + 1$ . Finally, we considered the probability  $p_{i,t}^k$  that an individual  $i$  is recaptured at time  $t$  in state  $k = \text{NB}$  or  $\text{B}$ . Note that, although it is possible, we do not consider individual effects on these parameters, so that the index  $i$  is dropped hereafter.

The state-space model relies on a combination of two sets of equations: the state equations, which specify the state of individuals at time  $t + 1$  given their state at time  $t$ , and the

observation equations, which specify the observation of individuals at time  $t$  given their state at time  $t$ . The state-space formulation of the CR model (Gimenez et al. 2007) is given by:

$$X_{i,t+1} | X_{i,t} \sim \text{multinomial} \left( 1, X_{i,t} \begin{bmatrix} f_t^{NB} y_t^{NB \rightarrow NB} & f_t^{NB} y_t^{NB \rightarrow B1} & f_t^{NB} y_t^{NB \rightarrow B2} & f_t^{NB} y_t^{NB \rightarrow B3} & 1 - f_t^{NB} \\ f_t^B y_t^{B1 \rightarrow NB} & f_t^B y_t^{B1 \rightarrow B1} & f_t^B y_t^{B1 \rightarrow B2} & f_t^B y_t^{B1 \rightarrow B3} & 1 - f_t^B \\ f_t^B y_t^{B2 \rightarrow NB} & f_t^B y_t^{B2 \rightarrow B1} & f_t^B y_t^{B2 \rightarrow B2} & f_t^B y_t^{B2 \rightarrow B3} & 1 - f_t^B \\ f_t^B y_t^{B3 \rightarrow NB} & f_t^B y_t^{B3 \rightarrow B1} & f_t^B y_t^{B3 \rightarrow B2} & f_t^B y_t^{B3 \rightarrow B3} & 1 - f_t^B \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right)$$

and

$$Y_{i,t} | X_{i,t} \sim \text{multinomial} \left( 1, X_{i,t} \begin{bmatrix} 1 - p_t^{NB} & p_t^{NB} & 0 & 0 & 0 \\ 1 - p_t^B & p_t^B & 0 & 0 & 0 \\ 1 - p_t^B & 0 & p_t^B & 0 & 0 \\ 1 - p_t^B & 0 & 0 & p_t^B & 0 \\ 1 - p_t^B & 0 & 0 & 0 & p_t^B \end{bmatrix} \right)$$

where the two equations above are the state and the observation equations, respectively. In the state equation, the matrix gathers the conditional probabilities of being in a state at a particular time occasion (columns) given the state at the previous time occasion (rows). Given that an individual is alive in a given state, it can survive and move to another state, survive and remain in the same state or die. To ensure that these probabilities are within the interval  $[0, 1]$  and sum to 1, we used a generalized (or multinomial) logit link function for the transition probabilities (Choquet 2008). Consider for example the four transition probabilities from state NB to one of the four states NB, B1, B2 and B3. We considered the probability of remaining in the same state as a reference (one minus the sum of the three other transition probabilities) and used the

transformation  $\log \left( \frac{\psi_t^{NB \rightarrow k}}{\psi_t^{NB \rightarrow NB}} \right)$  for  $k = B1, B2$  or  $B3$  (see Appendix B for technical details). In

the observation equation, the matrix gathers the conditional probability of being observed or not at a particular time occasion (columns) given the state at this current occasion (rows).

Given that an individual is alive in state non-breeder or breeder, it can be recaptured or not in

state non-breeder or breeder respectively. Given that an individual is dead, it is missed with certainty. We had also to specify the probability of being in a given state when first captured. Because first reproduction starts at 2 years in roe deer (Gaillard et al., 1992) and females are marked in their first winter at about 8 months of age (hence as non-breeder, Gaillard et al. 1993), we considered the probability of being in the state non-breeder when first captured to be unity, and null for any other state.

#### 4.2.3. Estimation of LRS using MCMC methods in a Bayesian framework

We used MCMC simulations in a Bayesian context (McCarthy 2007; King et al. 2009; Kéry and Schaub 2011; Royle and Dorazio 2008; Link and Barker 2011) to fit the above SSM. Prior distributions need to be specified for the parameters of a given model. Bayes' theorem is then used to update the priors using the likelihood to obtain posterior probabilities of all model parameters. The MCMC algorithm then generates values from a Markov chain whose stationary distribution is the joint posterior distributions of all model parameters. We used empirical means and standard deviations to summarize these posterior parameter distributions.

A nice feature of MCMC algorithms is that reproductive states can be treated as parameters to be estimated, just like survival, transition or recapture probabilities (see Newman et al. 2008 and Toni et al. 2009 for alternatives to MCMC). We produced posterior distributions for breeding states of individuals, which were used to calculate LRS. Specifically, for each MCMC iteration, we calculated for each individual the number of fawns it weaned per time step and then computed an average over all (post burn-in) iterations for each of the females. Summarizing across individuals, we obtained the frequency distribution of LRS. We quantified uncertainty around the LRS estimate by computing a 'mean' standard deviation which was obtained as an average, for a particular value of reproductive output, of all the individual LRS standard deviation for individuals having this specific number of offspring (see Appendix B for more details).

We provided vague prior distributions for all model parameters to induce little prior knowledge. Specifically, we chose uniform distributions on [0, 1] for the recapture and survival probabilities, and normal distributions with mean zero and variances 0.01 for the logit of the transition probabilities. We conducted a sensitivity analysis by modifying these priors using beta distributions with informative moments and posterior estimates of parameters were not changed.

Two MCMC chains with over-dispersed initial values were generated using 50,000 iterations with 10,000 burn-in iterations. Convergence was assessed using the Gelman and Rubin statistic, which compares the within- to the between-variability of chains started at different and dispersed initial values (Gelman 1996). Mixing was checked visually by inspecting the chains. The simulations were performed using JAGS (Plummer 2003) and the R package rjags was used to call JAGS, export results in R and calculate LRS. The code is available in Appendix B.

#### 4.2.4. Results

Posterior mean estimates for all parameters of the model are provided in Table 2. There was strong heterogeneity in detection probability according to reproductive status, with breeder less prone to recapture than non-breeders. Survival decreased with increasing age, regardless of reproductive status.

[Table 2 about here]

Much of the density of the distribution of LRS falls at zero (Figure 4): there were many individuals that did not successfully wean any fawns over their entire lifetime. We also observed a long tail in the LRS distribution, suggesting that some females had very high reproductive output.

[Figure 4 about here]

## 5. DISCUSSION

Multistate CR models are a valuable tool for estimating demographic parameters in the wild when detection probability is less than one. Their extension to hidden process models allows improved estimation of demographic parameters by accounting for variation in the state process (e.g., demographic stochasticity) while dealing with observation error. Using the same data that one would analyze with multistate models, SSM allows the estimation of demographic parameters as well as unknown states with imperfect detection (a feature we used to estimate LRS in section 4.2). By optimizing the use of all available information including those data that could not be used in multistate analyses, HMM allow us to cope with uncertainty in state assignment (section 3.2).

Traditionally, in the CR literature, HMM are fit in a Frequentist framework while SSM are fitted in a Bayesian framework. This does not need to be the case (de Valpine 2009), and examples of Frequentist analyses of SSMs are becoming available (Lele et al. 2007; de Valpine 2011). We advocate pragmatism when making the decision of one framework over the other (Gimenez 2008). This decision can be made based on one's programming skills or familiarity with software. However, there is a risk of using a sledgehammer to crack a nut by fitting unnecessarily complex models that come with difficulties (Bolker 2009; Cressie et al. 2009) like non-identifiability for example (Gimenez et al. 2003, 2009; Luo et al. 2009; Rouan et al. 2009a). We encourage practitioners to think first of the biological question to be addressed, and resort to the Bayesian or Frequentist approach accordingly. For example, in the roe deer

example (section 4.2), we aimed at estimating LRS. We opted for SSM fit in a Bayesian framework using MCMC methods because it allowed the simultaneous estimation of unknown reproductive states and demographic parameters while accounting for parameter estimation error (Figure 2). We could have used Frequentist alternatives to compute LRS (as shown by Rouan et al. 2009b).

In the roe deer data analysis (section 4.2), we assumed that uncertainty in assessing the reproductive status was due to the issue of detection only: if a female was seen in the field, its reproductive status corresponded to the observation made. This was a fair assumption since fawns closely follow their mother during the period considered (Gaillard et al. 2000). However, if this assumption were to be violated, one could easily incorporate uncertainty in state assignment in the calculation of LRS by adapting the model used in the titis data analysis (section 3.2).

We reviewed two types of hidden process models, both having a discrete space of states. Continuous hidden process models have never been used to our knowledge but could be useful in fitting an alternative to matrix models known as integral projection models developed by Easterling et al. (2000) for plant populations. In contrast with matrix models, which require discretizing individuals' states into classes, demographic parameters are modeled as continuous functions of individuals' states in integral projection models (e.g., size in plants or body mass in animals). Continuous hidden process models could be used as a unifying framework to combine the analysis of individual data, the construction of integral projection models and the projection of population fate by considering relevant states as being continuous.

The extension of hidden process models to the analysis of other sources of information than individual data also holds promise. In particular, Besbeas et al. (2002) developed an integrated framework to jointly analyze individual data using CR models and count data collected at the population level. So-called integrated population models use hidden process models to build the likelihood of count data (e.g., Schaub et al. 2007), while standard product-

1 multinomials are used to form the likelihood of capture-recapture data. The formulation of  
2 integrated population models in a unifying framework using HMM or SSM for both count and  
3 individual data would certainly benefit from the flexibility of hidden process models.

## 6 **ACKNOWLEDGMENTS**

7 The authors warmly thank Mathieu Buoro for his help with Figure 3, Lauriane Rouan for  
8 discussion about LRS calculation using capture-recapture data, Paul Scofield and David  
9 Fletcher for sharing the titis data, Jean-Michel Marin and Christian Robert for a discussion  
10 about the distinction between HMM and SSM. We also thank the participants to the 2009 and  
11 2010 Montpellier workshops on the analysis of capture-recapture data for their feedback on the  
12 titis analyses. This research was supported by a grant (reference: ANR-08-JCJC-0028-01) from  
13 the ‘Jeunes Chercheuses et Jeunes Chercheurs’ program of the French ANR to work on  
14 capture-recapture mixed models.



## REFERENCES

- Arnason, A.N. (1972) Parameter estimates from mark-recapture experiments on two populations subject to migration and death. *Researches on Population Ecology* 13:97-113.
- Arnason, A.N. (1973) The estimation of population size, migration rates and survival in a stratified population. *Researches on Population Ecology* 15:1-8.
- Besbeas, P., S. N. Freeman, B. J. T. Morgan, and E. A. Catchpole (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* 58:540-547.
- Bolker, B. (2009). Learning hierarchical models: advice for the rest of us. *Ecological Applications* 19:588-592.
- Brommer, J. E., L. Gustafsson, H. Pietiäinen and J. Merilä (2004). Single-generation estimates of fitness as proxies for long-term genetic contribution. *American Naturalist* 163:505–517.
- Brownie, C., Anderson, D.R., Burnham, K.P. & Robson, D.S. (1985) *Statistical inference from band recovery data - A handbook*. US Fish and Wildlife Service Editions.
- Buckland, S.T., Newman, K.B., Thomas, L., and Koesters, N. (2004). State-space models for the dynamics of wild animal populations. *Ecological Modelling* 171:157–175.
- Buckland, S.T., Newman, K.B., Fernandez, C., Thomas, L., and Harwood, J. (2007). Embedding population dynamics models in inference. *Statistical Science* 22:44–58.
- Burnham, K.P. (1993) A theory for combined analysis of ring recovery and recapture data. *Marked individual in the study of bird population*. (eds J.D. Lebreton & P.M. North), pp. 199-213. Birkhäuser Verlag, Basel, Switzerland.
- Burnham, K.P., D.R. Anderson, G.C. White, C. Brownie, and K.H. Pollock (1987). Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society Monograph* 5. 437 pp.

- Buoro, M., Prévost, E. and O. Gimenez (2010). Investigating evolutionary trade-offs in wild populations of Atlantic salmon (*Salmo salar*): incorporating detection probabilities and individual heterogeneity. *Evolution* 64: 2629-2642
- Calvert, A., S. Bonner, I. Jonsen, J. Flemming, S. Walde, and P. Taylor. 2009. A hierarchical Bayesian approach to multi-state mark-recapture: simulations and applications. *Journal of Applied Ecology* 46:610–620.
- Caswell, H., *Matrix population models: Construction, analysis and interpretation. Second Edition*. 2001, Sunderland, Massachusetts: Sinauer Associates.
- Caswell, H. & Fujiwara, M. (2004) Beyond survival estimation: mark-recapture, matrix population models, and population dynamics. *Animal Biodiversity and Conservation* 27.1:471-488.
- Choquet, R. 2008. Automatic generation of multistate capture-recapture models. *The Canadian Journal of Statistics* 36:1–15.
- Choquet R., Rouan L. & Pradel R. (2009). Program E–SURGE: a software application for fitting multievent models. In: *Modeling Demographic Processes in Marked Populations* (eds. Thomson DL, Cooch EG & Conroy MJ). Springer Berlin, Germany, pp. 845-865.
- Clark, J. S., G. Ferraz, N. Oguge, H. Hays, and J. Dicostanzo. 2005. Hierarchical bayes for structured, variable populations: from recapture data to life-history prediction. *Ecology* 86:2232–2244.
- Clobert J. and Lebreton J.-D. (1987). About recent research in models for mark-recapture and mark-resighting data. A reply to C. Brownie. *Biometrics* 43:1019-1022.
- Clobert, J., Lebreton, J.-D., Allaine, D. & Gaillard, J.M. (1994) The estimation of age-specific breeding probabilities from recaptures or resightings of marked animals. II Longitudinal models. *Biometrics* 50:375-387.
- Clutton-Brock, T. H. (1988). *Reproductive Success. Studies of Individual Variation in Contrasting Breeding Systems*. University of Chicago Press, Chicago.

- 1 Conn, P.B. and E.G. Cooch. (2009) Multistate capture-recapture analysis under imperfect state  
2 observation: an application to disease models. *Journal of Applied Ecology* 46, 486-492.
- 3 Cormack, R.M. (1964) Estimates of survival from the sighting of marked animals. *Biometrika*  
4 51:429-438.
- 5 Crespin L., Choquet R., Lima M.A., Merritt J.F. & Pradel R. (2008) Is heterogeneity of  
6 catchability in capture-recapture studies a mere sampling artifact or a biologically relevant  
7 feature of the population? *Population Ecology* 50:247-256.
- 8 Cressie, N., C.A. Calder, J.S. Clark, J.M.V. Hoef and C.K. Wikle (2009). Accounting for  
9 uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical  
10 modelling. *Ecological Applications* 19:553-570
- 11 Cubaynes S., Pradel R., Choquet R., Duchamp C., Gaillard J.-M., Lebreton J.-D., Marboutin  
12 E., Miquel C., Reboulet A.-M., Poillot C., Taberlet P. & Gimenez O. (2010). Importance  
13 of accounting for detection heterogeneity when estimating abundance: The case of French  
14 wolves. *Conservation Biology* 24:621-626.
- 15 Desprez, M. R. Pradel, E. Cam, J.-Y. Monnat & O. Gimenez (2011). Now you see him, now  
16 you don't: Experience, not age, is related to reproduction in Kittiwakes. *Proceedings of the*  
17 *Royal Society B: Biological Sciences* 278:3060-3066
- 18 Diekmann, O. and Heesterbeek, J.A.P. (2000) *Mathematical Epidemiology of Infectious*  
19 *Diseases: Model Building, Analysis and Interpretation*. Wiley, New York.
- 20 Dupuis, J.A. (1995) Bayesian estimation of movement and survival probabilities from capture-  
21 recapture data. *Biometrika* 82:761-772.
- 22 Ebert C., Knauer F., Storch I. & Hohmann U. Individual heterogeneity as a pitfall in population  
23 estimates based on non-invasive genetic sampling: a review and recommendations.  
24 *Wildlife Biology* 16:225-240.
- 25 Easterling, M.R., S. P. Ellner, and P. M. Dixon. 2000. Size-specific sensitivity: applying a new  
26 structured population model. *Ecology* 81:694-708.

- 1 Gaillard, J. M., Sempéré, A. J., Boutin, J. M., Van Laere, G., and Boisaubert, B. (1992).  
2 Effects of age and body weight on the proportion of females breeding in a population of  
3 roe deer (*Capreolus capreolus*). *Canadian Journal of Zoology* 70:1541–1545.
- 4 Gaillard, J. M., Andersen, R., Delorme, D., and Linnell, J. (1998). Family effects on growth  
5 and survival of juvenile roe deer. *Ecology* 79:2878–2889.
- 6 Gaillard, J. M., Festa-Bianchet, M., Delorme, D., and Jorgenson, J. (2000). Body mass and  
7 individual fitness in female ungulates: bigger is not always better. *Proceedings of the*  
8 *Royal Society of London, Series B* 267:471–477.
- 9 Gamelon, M., A. Besnard, J.-M. Gaillard, S. Servanty, E. Baubet, S. Brandt and O. Gimenez  
10 (2011). High hunting pressure selects for earlier birth data: Wild boar as a case study.  
11 *Evolution* 65: 3100-3112.
- 12 Gelman, A. 1996. Inference and monitoring convergence. Pages 131–143 in W. R. Gilks, S.  
13 Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in practice*.  
14 Chapman and Hall, London, UK.
- 15 Ghimire, S., Gimenez, O., Pradel, R., McKey, D., and Y. Aumeeruddy-Thomas (2008).  
16 Demographic variation and population viability in a threatened Himalayan medicinal and  
17 aromatic herb (*Nardostachys grandiflora*): matrix modelling of harvesting effects in two  
18 contrasting habitats. *Journal of Applied Ecology* 45: 41-51.
- 19 Gimenez O., Choquet R., Amor L., Scofield P., Fletcher D., J.-D. Lebreton and R. Pradel  
20 (2005). Efficient Profile-Likelihood Confidence Intervals for Capture-Recapture Models.  
21 *Journal of Agricultural, Biological, and Environmental Statistics* 10: 184-196.
- 22 Gimenez, O., R. Covas, C. R. Brown, M. D. Anderson, M. Bomberger Brown and T.  
23 Lenormand. (2006). Nonparametric estimation of natural selection on a quantitative trait  
24 using mark-recapture data. *Evolution* 60: 460-466

- 1 Gimenez, O., V. Rossi, R. Choquet, C. Dehais, B. Doris, H. Varella, J.-P. Vila, and R. Pradel.  
2 (2007). State-space modelling of data on marked individuals. *Ecological Modelling*  
3 206:431–438.
- 4 Gimenez, O. (2008). Discussion: Towards a Bayesian analysis template? *The Canadian Journal*  
5 *of Statistics* 36: 21-24.
- 6 Gimenez, O., A. Viallefont, A. Charmantier, R. Pradel, E. Cam, C. R. Brown, M. D. Anderson,  
7 M. Bomberger Brown, R. Covas, J.-M. Gaillard (2008). The risk of flawed inference in  
8 evolutionary studies when detectability is less than one. *The American Naturalist* 172:441–  
9 448.
- 10 Gimenez, O., Morgan, B.J.T. and S.P. Brooks (2009). Weak Identifiability in Models for Mark-  
11 Recapture-Recovery Data. In *Modeling Demographic Processes in Marked Populations*.  
12 Thomson, D. L.; Cooch, E. G.; Conroy, M. J. (Eds.). Springer Series: Environmental and  
13 Ecological Statistics, Vol. 3, pages 1055-1067.
- 14 Gimenez, O. and R. Choquet (2010) Incorporating individual heterogeneity in studies on  
15 marked animals using numerical integration: capture-recapture mixed models. *Ecology* 91:  
16 951-957
- 17 Hestbeck, J.B., Nichols, J.D. and Malecki, R. (1991) Estimates of movement and site fidelity  
18 using mark-resight data of wintering canada geese. *Ecology* 72:523-533
- 19 Jolly, G.M. (1965) Explicit estimates from capture-recapture data with both death and  
20 immigration-stochastic model. *Biometrika* 52:225-247
- 21 Juillet C., Choquet R., Gauthier G. and Pradel R. (2011). A capture-recapture model with  
22 double-marking, live and dead encounters, and heterogeneity of reporting due to auxiliary  
23 mark loss. *Journal of Agricultural, Biological and Environmental Statistics* 16:88-104
- 24 Kendall, W.L. and Bjorkland, R. (2001) Using open robust design models to estimate  
25 temporary emigration from capture-recapture data. *Biometrics* 57:1113-1122

- 1 Kendall W.L., Hines J.E., Nichols J.D. (2003). Adjusting multistate capture-recapture models  
2 for misclassification bias: Manatee breeding proportions. *Ecology* 84:1058-1066
- 3 Kéry, M. and Schaub, M. 2011. *Bayesian population analysis using WinBUGS – a hierarchical*  
4 *perspective*. Academic Press.
- 5 King, R., Morgan, B.J.T., Gimenez, O. and S.P. Brooks. 2009. *Bayesian analysis for*  
6 *population ecology*. Chapman and Hall.
- 7 Lawless, J.F. and McLeish, D.L. (1984) The information in aggregate data from Markov  
8 chains. *Biometrika* 71:419-430.
- 9 Lebreton, J.-D., Burnham, K.P., Clobert, J. and Anderson, D.R. (1992) Modeling survival and  
10 testing biological hypotheses using marked animals: a unified approach with case studies.  
11 *Ecological Monographs* 62:67-118.
- 12 Lebreton, J.-D., Almeras, T. and Pradel, R. (1999) Competing events, mixtures of information  
13 and multistrata recapture models. *Bird Study* 46:39-46.
- 14 Lebreton, J.-D., Hines, J.E., Pradel, R., Nichols, J.D. and Spendelov, J.A. (2003) Estimation  
15 by capture-recapture of recruitment and dispersal over several sites. *Oikos* 101:253-264.
- 16 Lebreton, J. -D., J. D. Nichols, R. J. Barker, R. Pradel, and J. A. Spendelov. 2009. Modeling  
17 individual animal histories with multistate capture–recapture models. Pages 87–173 in H.  
18 Caswell, editor. *Advances in Ecological Research*, 41. Academic Press, Burlington, MA.
- 19 Lele, S. R., B. Dennis, and F. Lutscher (2007). Data cloning: easy maximum likelihood  
20 estimation for complex ecological models using Bayesian Markov chain Monte Carlo  
21 methods. *Ecology Letters* 10:551–563.
- 22 Link, W. A. and R. J. Barker. 2005. Modeling association among demographic parameters in  
23 open population capture–recapture data. *Biometrics* 61:46–54
- 24 Link, W. A. and Barker, R. J. (2010) *Bayesian inference with ecological applications*.  
25 Academic Press.

- 1 Luo, Y.Q., E.S. Wend, X.W. Wu, C. Gao, X.H. Zhou and L. Zhang (2009). Parameter
- 2 identifiability, constraint, and equifinality in data assimilation with ecosystem models
- 3 Ecological Applications 19:571-574
- 4 Marzolin, G., A. Charmantier and O. Gimenez (2011). Frailty in state-space models:
- 5 application to actuarial senescence in the dipper. Ecology 92:562-567
- 6 McCarthy, M. A. (2007) *Bayesian methods for ecology*. Cambridge University Press,
- 7 Cambridge, UK.
- 8 McGraw, J. B. and Caswell, H. (1996). Estimation of individual fitness from life-history data.
- 9 American Naturalist 147:47–64.
- 10 McLoughlin, P. D., J.-M. Gaillard, M. S. Boyce, C. Bonenfant, F. Messier, P. Duncan, D.
- 11 Delorme, B. Van Moorter, S. Said and F. Klein (2007). Lifetime reproductive success and
- 12 composition of the home range in a large herbivore. Ecology 88:3192-3201.
- 13 Newman, K.B., Buckland, S.T., Lindley, S.T., Thomas, L., and Fernandez, C. (2006). Hidden
- 14 process models for animal population dynamics. Ecological Applications 16:74–86.
- 15 Newton, I., ed. (1989) *Lifetime Reproduction in Birds*. Academic Press, London.
- 16 Nichols, J.D. (1992). Capture-recapture models. Bioscience 42:94-102.
- 17 Nichols J.D., Kendall W.L., Hines J.E. & Spendelov J.A. (2004). Estimation of sex-specific
- 18 survival from capture-recapture data when sex is not always known. Ecology 85:3192-
- 19 3201.
- 20 O’Hara, R., S. Lampila, and M. Orell. 2009. Estimation of rates of births, deaths, and
- 21 immigration from mark-recapture data. Biometrics 65:275–281
- 22 Oli, M.K., M. Venkataraman, P. A. Klein, L. D. Wendland and M. B. Brown (2006).
- 23 Population dynamics of infectious diseases: A discrete time model. Ecological Modelling
- 24 198: 183–194.
- 25 Oliver L.J., Morgan B.J.T., Durant S.M. & Pettorelli N. Individual heterogeneity in recapture
- 26 probability and survival estimates in cheetah. Ecological Modelling 222:776-784.

- Papadatou, E., R. Pradel, M. Schaub, D. Dolch, H. Geiger, C. Ibañez, G. Kerth, A. Popa-Lisseanu, W. Schorcht, J. Teubner and O. Gimenez (2011). Comparing survival among species with imperfect detection using multilevel analysis of mark-recapture data: a case study on bats. *Ecography* [doi: 10.1111/j.1600-0587.2011.07084.x]
- Papaïx, J., S. Cubaynes, M. Buoro, A. Charmantier, P. Perret and O. Gimenez (2010). Combining capture–recapture data and pedigree information to assess heritability of demographic parameters in the wild. *Journal of Evolutionary Biology* 23: 2176-2184
- Péron, G., Y. Ferrand, F. Gossmann, C. Bastat, M. Guenezan and O. Gimenez (2011). Nonparametric spatial regression of survival probability: visualization of population sinks in Eurasian Woodcock. *Ecology* [doi:10.1890/10-2224.1]
- Péron G., Crochet P.A., Choquet R., Pradel R., Lebreton J.-D. and Gimenez O. (2010). Capture-recapture models with heterogeneity to study survival senescence in the wild. *Oikos* 119:524-532.
- Péron G., Lebreton J.D. & Crochet P.A (2010). Breeding dispersal in black-headed gull: the value of familiarity in a contrasted environment. *Journal of Animal Ecology* 79:317-326.
- Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X.
- Pradel, R. and Lebreton, J.-D. (1999) Comparison of different approaches to study the local recruitment of breeders. *Bird Study* 46:74-81.
- Pradel R. (2005). Multievent: an extension of multistate Capture-Recapture Models to uncertain states. *Biometrics* 61:442-447.
- Pradel R. (2009). The stakes of capture-recapture models with state uncertainty. In: *Modeling Demographic Processes in Marked Populations* (eds. Thomson DL, Cooch EG & Conroy MJ). Springer Berlin, Germany, pp. 781-795.



- 1 Pradel R., Choquet R., Lima M.A., Merritt J.F. & Crespin L. (2010). Estimating population  
2 growth rate from capture-recapture data in presence of capture heterogeneity. *Journal of*  
3 *Agricultural, Biological and Environmental Statistics* 15:248-258.
- 4 Pradel R., Choquet R., & Béchet A. (2011). Breeding experience might be a major determinant  
5 of breeding probability in long-lived species: the case of the greater flamingo.  
6 oai:hal.archives-ouvertes.fr:hal-00655507
- 7 R Development Core Team (2011). R: A language and environment for statistical computing.  
8 R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL  
9 <http://www.R-project.org/>.
- 10 Rouan L., Choquet R. & Pradel R. (2009a). A general framework for modeling memory in  
11 capture-recapture data. *Journal of Agricultural, Biological and Environmental Statistics*  
12 14:338-355.
- 13 Rouan L., Gaillard J.-M., Y. Guédon & Pradel R. (2009b). Estimation of Lifetime  
14 Reproductive Success When Reproductive Status Cannot Always Be Assessed. In:  
15 *Modeling Demographic Processes in Marked Populations* (eds. Thomson DL, Cooch EG  
16 & Conroy MJ). Springer Berlin, Germany, pp. 867-879.
- 17 Royle, J. A. and Dorazio, R. M. 2008. *Hierarchical modelling and inference in ecology*.  
18 Academic Press.
- 19 Royle, J. A. 2008. Modeling individual effects in the Cormack-Jolly-Seber model: a state-  
20 space formulation. *Biometrics* 64:364-370.
- 21 Sanz-Aguilar A., Tavecchia G., Genovart M., Igual J.M., Oro D., Rouan L. & Pradel R. (2011).  
22 Studying the reproductive skipping behavior in long-lived birds by adding nest-inspection  
23 to individual-based data. *Ecological Applications* 21:555-564.
- 24 Schaub, M., Gimenez, O., Sierro, A. and R. Arlettaz (2007). Use of Integrated Modeling to  
25 Enhance Estimates of Population Dynamics Obtained from Limited Data. *Conservation*  
26 *Biology* 21: 945–955.

- 1 Schofield, M. R. and R. J. Barker. 2008. A Unified Capture-Recapture Framework. *Journal of*  
2 *Agricultural Biological and Environmental Statistics* 13:458-477.
- 3 Schofield, M. R. and R. J. Barker. 2011. Full open population capture-recapture models with  
4 individual covariates. *Journal of Agricultural, Biological and Environmental Statistics*  
5 16:253-268.
- 6 Scofield, P., Fletcher, D. and C. Robertson (2001). Titi (Sooty Shearwaters) on Whero Island:  
7 An analysis of historic banding data using modern capture-recapture techniques. *Journal of*  
8 *Agricultural, Biological and Environmental Statistics* 6:268-280.
- 9 Seber, G.A.F. (1965) A note on the multiple-recapture census. *Biometrika*, 52:249-259.
- 10 Servanty, S., J.-M. Gaillard, Ronchi, F., Focardi, S., Baubet E. & O. Gimenez (2011).  
11 Influence of harvesting pressure on demographic tactics: implications for wildlife  
12 management. *Journal of Applied Ecology* 48: 835–843
- 13 Servanty, S., R. Choquet, E. Baubet, S. Brandt, J.-M. Gaillard, M. Schaub, C. Toïgo, J.-D.  
14 Lebreton, M. Buoro and O. Gimenez (2010). Assessing whether mortality is additive using  
15 marked animals: a Bayesian state-space modeling approach. *Ecology* 91: 1916–1923
- 16 Tavecchia, G., J. Adrover, A. Muñoz Navarro, and R. Pradel (2012). Modelling mortality  
17 causes in longitudinal data in the presence of tag loss: application to raptor poisoning and  
18 electrocution. *Journal of Applied Ecology* 49:297-305
- 19 Toni, T. D. Welch, N. Strelkova, A. Ipsen and M. P.H Stumpf (2009). Approximate Bayesian  
20 computation scheme for parameter inference and model selection in dynamical systems. *J.*  
21 *R. Soc. Interface* 6:187-202
- 22 Tuljapurkar, S. and H. Caswell, *Structured-population models in marine, terrestrial, and*  
23 *freshwater systems*. 1997, New York: Chapman and Hall.
- 24 de Valpine (2009). Shared challenges and common ground for Bayesian and classical analysis  
25 of hierarchical statistical models. *Ecological Applications* 19:584-588

- 1 de Valpine, P. 2011. Frequentist analysis of hierarchical models for population dynamics and  
2 demographic data. *Journal of Ornithology* doi: 10.1007/s10336-010-0642-5.
- 3 Van Tienderen, P. H. 2000. Elasticities and the link between demographic and evolutionary  
4 dynamics. *Ecology* 81: 666-679.
- 5 Véran S., Gimenez O., Flint E., Kendall W.L., Doherty P.F. & Lebreton J.-D. (2007).  
6 Quantifying the impact of longline fisheries on adult survival in the black-footed albatross.  
7 *Journal of Applied Ecology* 44:942-952.
- 8 Williams, B.K., Nichols, J.D. and Conroy, M.J. (2002) *Analysis and management of animal*  
9 *populations*. Academic Press, San Diego.

10

## FIGURE LEGENDS

**Figure 1:** Three types of registration of a same discrete time survival process. Top: follow-up data, bottom left: CJS models, bottom right: dead recovery models (see text for details). The complete registration is a particular case of each of the two other ones.

**Figure 2:** Number of papers about Markov models (keywords ‘Markov’ or ‘Markovian’) in human and animal population biology (keywords ‘man or human’ and ‘animal or mammals or birds’ respectively) extracted from the Web of Science (search completed on the 9<sup>th</sup> of August, 2011). Predicted trends were obtained from a Poisson regression in which both the year and the type of population biology (human vs. animal) effects were significant on an additive scale.

**Figure 3:** Schematic representation of a hidden process model for an individual  $i$  between two sampling occasions  $t - 1$  and  $t$ . The first layer is a succession of hidden states (solid circles) or latent states of individual  $i$  at time  $t$  ( $X_{i,t}$ ) depending on its states at time  $t - 1$  ( $X_{i,t-1}$ ). The dynamic of the states is driven by transition probabilities, e.g., survival probability if states are alive or dead. The second layer corresponds to observation or not of individual  $i$  at time  $t$  ( $Y_{i,t}$ ) conditional on individual  $i$  being in state  $X_{i,t}$ . These events are driven by some probabilities, e.g., detection probabilities if observations are being captured or not. The observation process is the visible part of the state process. Note that this representation holds for both hidden-Markov models and state-space models.

**Figure 4:** Frequency distribution of lifetime reproductive success in roe deer. Error bars are  $\pm 1$  ‘mean’ standard deviation where ‘mean’ standard deviation was obtained as an average, for a particular value of reproductive output, of all the individual LRS standard deviation for individuals having this specific number of weaned offspring (see Appendix B for more details).

**TABLE LEGENDS**

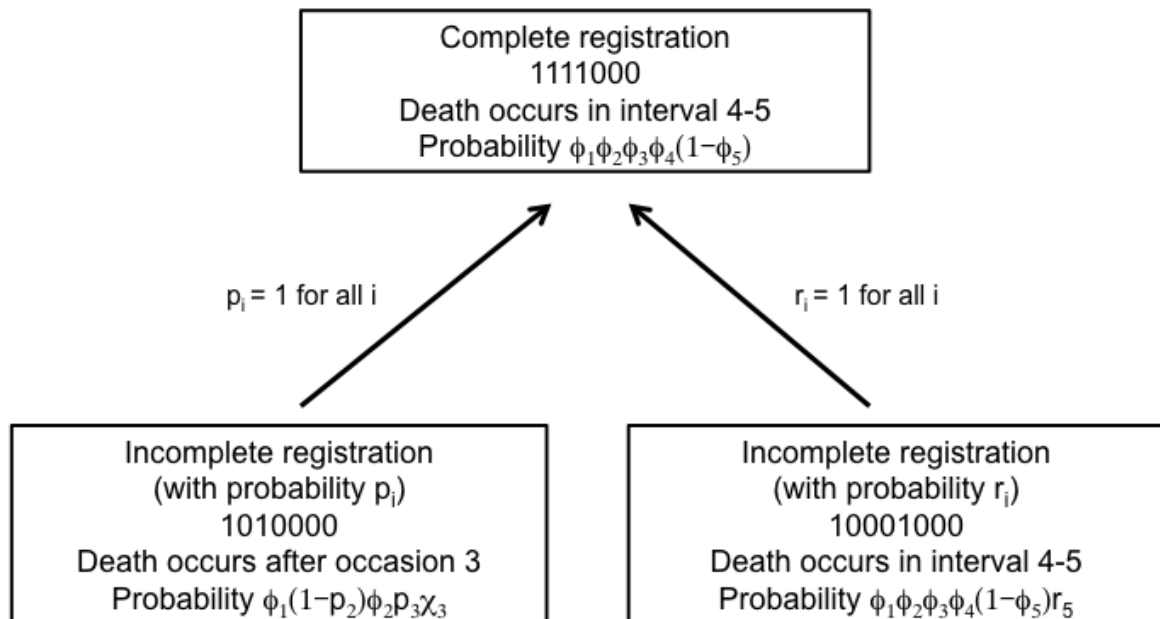
Box 1: Glossary (alphabetical order)

Table 1: Parameter maximum likelihood estimates (MLE) with standard errors (SE) for the hidden-Markov model fitted to the titis data, with and without uncertainty in state assignment. All parameters are probabilities. Note that the probability of remaining non-breeder and the probability of making the transition from state breeder to state non-breeder are not displayed but can be obtained as a complement.

Table 2: Monte Carlo parameter numerical summaries (posterior means with standard deviations) from the state-space model fitted to the roe deer data. All parameters are probabilities. Note that the probability of remaining in the same reproductive state is not displayed but can be obtained as a complement.

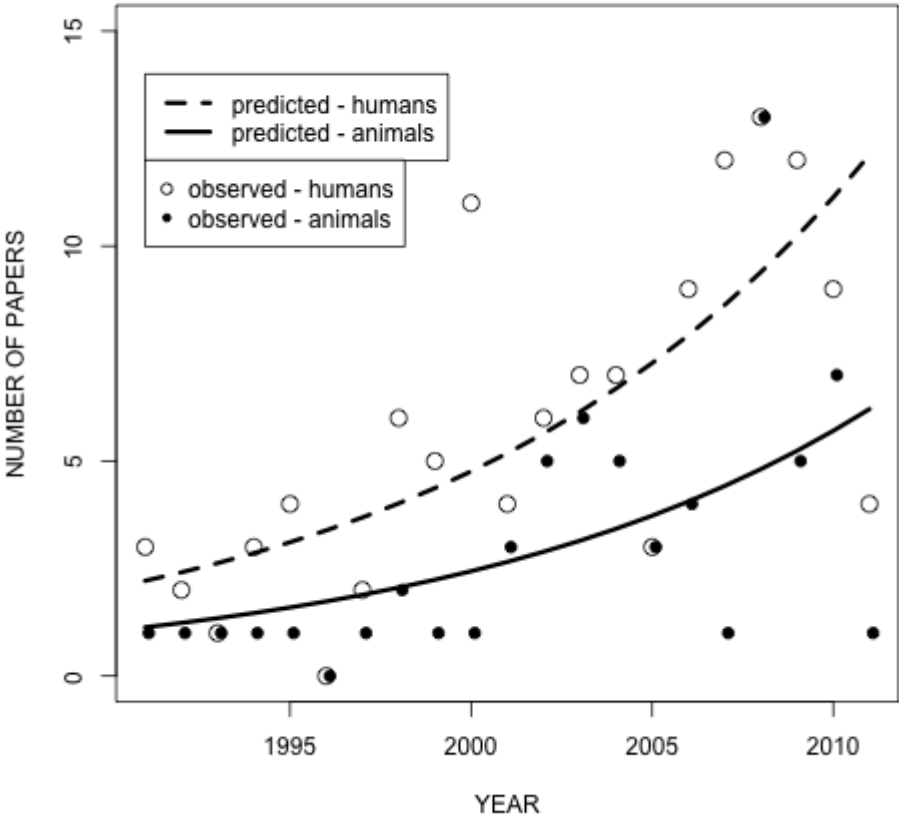
1 Figure 1:

2



3

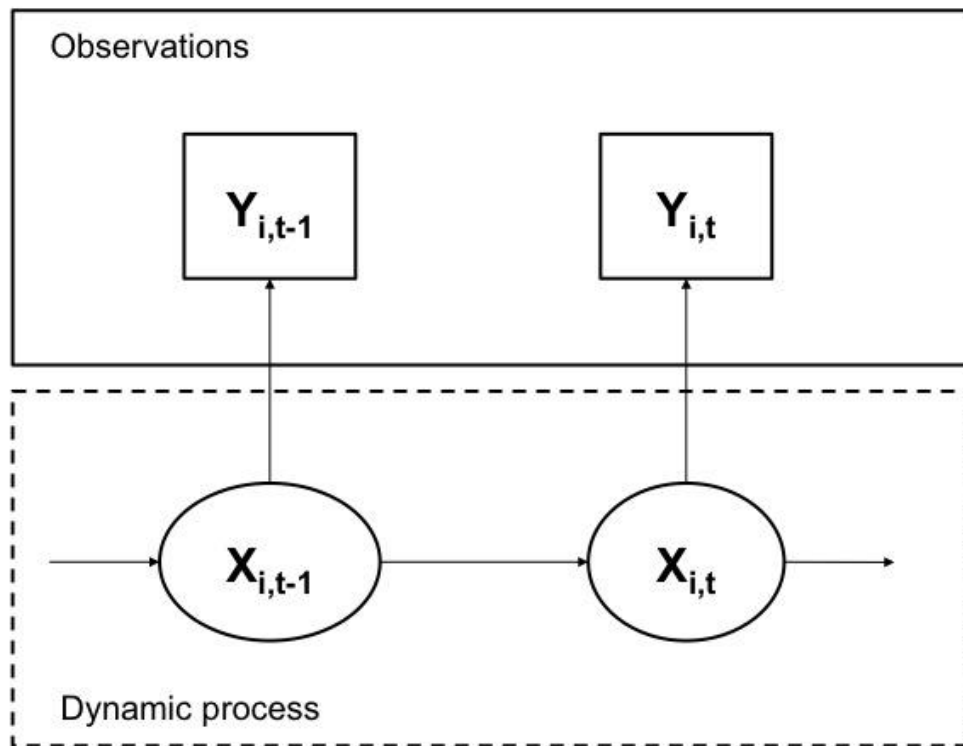
1    Figure 2:



2

3

1 Figure 3:



2

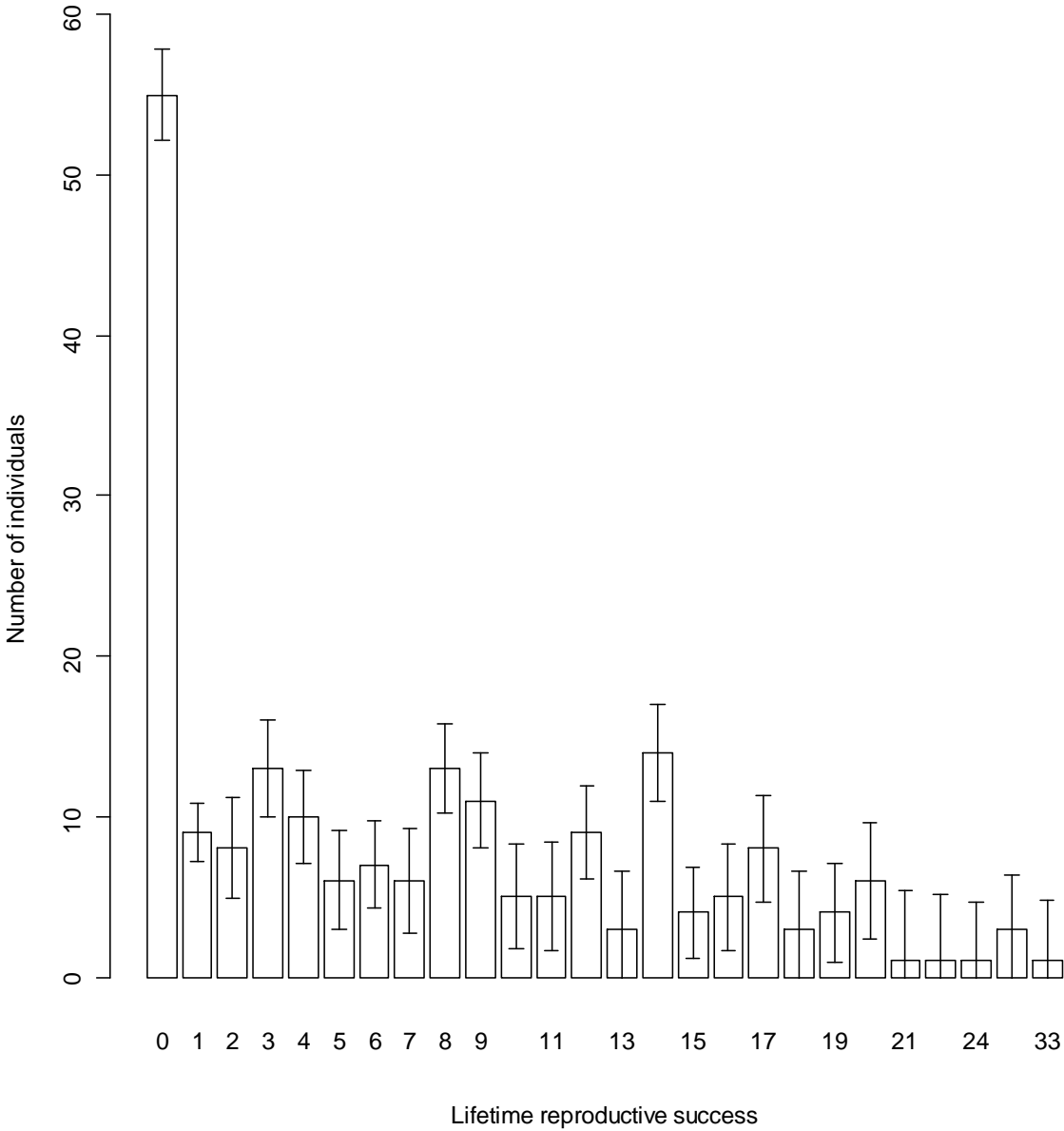
3

4

5



1 Figure 4:



2

Box 1:

**CAPTURE-RECAPTURE (CR):** A generic term embedding several protocols for monitoring (mainly) plants and animals in the wild using marks to uniquely identify individuals. The analysis of these data allows the estimation of, e.g., survival, dispersal and abundance while explicitly accommodating for detection probability less than one.

**CORMACK-JOLLY-SEBER MODEL (CJS):** A capture-recapture model that is used for open populations to estimate time-dependent survival and detection probabilities. The Cormack-Jolly-Seber model is easily formulated as a hidden process model by considering two states (alive or dead) and two observations or events (detected or not). The dynamic of states is governed by a Markov model parameterized with survival probabilities while the observations are made from these states according to detection probabilities.

**HIDDEN MARKOV MODEL (HMM):** A particular case of state-space model in which the states are Markovian, i.e. the next state depends only on the current state and not on the sequence of states that occurred before.

**HIDDEN PROCESS MODEL:** A generic term referring to either a state-space model or a hidden Markov model.

**MULTIEVENT MODEL:** A particular type of hidden Markov model. Multievent models correspond to extensions of multistate models that allow accommodating uncertainty in state assignment.

**MULTISTATE MODEL:** An extension of the Cormack-Jolly-Seber model that allows estimation of transition between states. Examples of such states are geographical sites, behavioral or physiological conditions like, e.g., breeding vs. non-breeding or healthy vs. diseased.

**STATE-SPACE MODEL (SSM):** A model that runs two time series in parallel, one captures the dynamic of the true states (latent) and the other consists of observations that are made from these underlying but possibly unknown states.

1 Table 1:

parameter	Reproductive state	Reproductive state
	uncertainty	uncertainty
	ignored	modeled
being a breeder upon first encounter $\pi^B$	0.704 (0.014)	0.704 (0.028)
survival of non-breeders $\phi^{NB}$	0.801 (0.016)	0.814 (0.016)
survival of breeders $\phi^B$	0.849 (0.016)	0.837 (0.019)
transition non-breeder to breeder $\psi^{NB \rightarrow B}$	0.249 (0.021)	0.219 (0.025)
transition breeder to breeder $\psi^{B \rightarrow B}$	0.761 (0.021)	0.774 (0.049)
detection of non-breeders $p^{NB}$	0.597 (0.028)	0.565 (0.028)
detection of breeders $p^B$	0.563 (0.027)	0.598 (0.032)
non-breeder state assignment $\delta^{NB}$	NA	0.188 (0.014)
breeder state assignment $\delta^B$	NA	0.738 (0.055)

2

1 Table 2:

parameter	posterior mean estimate (standard deviation)
transition non-breeder to weaning 1 fawn $\psi^{NB \rightarrow B1}$	0.160 (0.033)
transition non-breeder to weaning 2 fawns $\psi^{NB \rightarrow B2}$	0.149 (0.031)
transition non-breeder to weaning 3 fawns $\psi^{NB \rightarrow B3}$	0.208 (0.036)
transition weaning 1 fawn to non-breeder $\psi^{B1 \rightarrow NB}$	0.274 (0.067)
transition weaning 1 to 2 fawns $\psi^{B1 \rightarrow B2}$	0.163 (0.068)
transition weaning 1 to 3 fawns $\psi^{B1 \rightarrow B3}$	0.316 (0.086)
transition weaning 2 fawns to non-breeder $\psi^{B2 \rightarrow NB}$	0.229 (0.059)
transition weaning 2 to 1 fawns $\psi^{B2 \rightarrow B1}$	0.159 (0.059)
transition weaning 2 to 3 fawns $\psi^{B2 \rightarrow B3}$	0.302 (0.082)
transition weaning 3 fawns to non-breeder $\psi^{B3 \rightarrow NB}$	0.261 (0.048)
transition weaning 3 to 1 fawns $\psi^{B3 \rightarrow B1}$	0.156 (0.045)
transition weaning 3 to 2 fawns $\psi^{B3 \rightarrow B2}$	0.224 (0.052)
detection of breeders $p^B$	0.341 (0.032)
detection of non-breeders $p^{NB}$	0.821 (0.110)
survival of yearling non-breeders $\phi^{NB}(y)$	0.941 (0.027)
survival of prime-age non-breeders $\phi^{NB}(pa)$	0.843 (0.031)
survival of old non-breeders $\phi^{NB}(o)$	0.648 (0.057)
survival of prime-age non-breeders $\phi^B(pa)$	0.934 (0.019)
survival of old non-breeders $\phi^{NB}(o)$	0.831 (0.030)

3

## 1 **Appendix A1. Instructions for fitting the HMM to titis data in program E-SURGE**

2 To fit the model without uncertainty in E-SURGE, three states are considered {alive non-  
3 breeder, alive breeder, dead} = {NB, B, D}, and the events are three as well {non-detected,  
4 seen non-successful breeder, seen successful breeder} = {0,1,2}.

5 In GEPAT, declare two steps for the transitions (survival then conditional transitions).

6 Initial state:  $\Pi = [\pi \quad * \quad -]$

7

8 Transition 1 (survival):  $S = \begin{bmatrix} \phi & - & * \\ - & \phi & * \\ - & - & * \end{bmatrix}$

9 Transition 2 (conditional transition):  $T = \begin{bmatrix} * & \psi & - \\ \psi & * & - \\ - & - & * \end{bmatrix}$

10

11 Event:  $C = \begin{bmatrix} * & \beta & - \\ * & - & \beta \\ * & - & - \end{bmatrix}$

12

13

14 In GEMACO, the syntax is:

15

16 For Initial State: IS = i

17 For Transition:

18 S (i.e. step 1) = f

19 T (i.e. step 2) = f

20 For Event: C = firste+nexte.f

21

22 To fit the model with uncertainty in E-SURGE, the same three states are required {alive non-  
23 breeder, alive breeder, dead} = {NB, B, D}, but the events are four {non-detected, seen and  
24 ascertained as non-breeder, seen and ascertained as breeder, not ascertained} = {0, 1, 2, 3}.

25 In GEPAT, declare two steps for the transitions (survival then conditional transitions) and two  
26 steps for the events (capture then ascertainment).

27 Initial state:  $\Pi = [\pi \quad * \quad -]$

28 Transition 1 (survival):  $S = \begin{bmatrix} \phi & - & * \\ - & \phi & * \\ - & - & * \end{bmatrix}$

1 Transition 2 (conditional transition):  $T = \begin{bmatrix} * & \psi & - \\ \psi & * & - \\ - & - & * \end{bmatrix}$

2

3 Event 1 (capture):  $C = \begin{bmatrix} * & \beta & - \\ * & - & \beta \\ * & - & - \end{bmatrix}$

4 Event 2 (ascertainment):  $A = \begin{bmatrix} * & - & - & - \\ - & \delta & - & * \\ - & - & \delta & * \end{bmatrix}$

5

6 In GEMACO, the syntax is:

7

8 For Initial State: IS = i

9 For Transition:

10 S (i.e. step 1) = f

11 T (i.e. step 2) = f

12 For Event:

13 C (i.e. step 1) = firste + nexte.f

14 A (i.e. step 2) = f

15

16

## 1    **Appendix A2. R Code for generating uncertainty in titis data**

2    To artificially generate uncertainty on both the states non-breeder and breeder, we used the R  
3    script below to alter the raw capture-recapture data

```
4  
5    # original data are stored in R matrix titi with individuals in rows and years in columns  
6    # 1 seen as non-breeder  
7    # 2 seen as breeder  
8    # 0 not seen  
9  
10    # nb of capture occasions  
11    ny <- ncol(titi)  
12    # nb of individuals  
13    nind <- nrow(titi)  
14  
15    # the amended data with uncertainty are stored in R matrix titi2  
16    titi2 <- titi  
17    for (i in 1:nind)  
18    {  
19    for (j in 1:ny){  
20    # 1 seen and ascertained Non-Breeder (with probability .2)  
21    # 2 seen and ascertained Breeder (with probability .7)  
22    # 3 seen but not ascertained (Non-Breeders with probability .8 + Breeders with probability .3)  
23    # 0 not seen  
24  
25    # Non-Breeders are ascertained with probability .2  
26    if (titi[i,j] == 1)  
27    {  
28    temp <- rbinom(1,size=1,prob=.2)  
29    if (temp == 1) titi2[i,j] <- 1 # if ascertained NB, event = 1  
30    if (temp == 0) titi2[i,j] <- 3 # if not ascertained, event = 3  
31    }  
32  
33    # Breeders are ascertained with probability .7 (event = 1),  
34    # or not ascertained with probability .3 (event = 2)  
35    if (titi[i,j] == 2)  
36    {  
37    temp <- rbinom(1,size=1,prob=.7)  
38    if (temp == 1) titi2[i,j] <- 2 # if ascertained B, event = 2  
39    if (temp == 0) titi2[i,j] <- 3 # if not ascertained, event = 3  
40    }  
41  
42    }  
43    }  
44  
45
```

46

## Appendix B. BUGS and R codes to estimate LRS using SSM fitted to deer data

```
#####  
# 0 - READ IN DATA (subsample of 50 individuals from original dataset used in the paper)  
#####  
mydata <- matrix(  
c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,1,0,0,1,0,1,0,1,0,1,0,1,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,1,0,1,0,0,  
0,0,0,0,0,0,1,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,1,1,1,1,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,1,0,0,0,0,0,1,0,0,0,  
0,0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,0,1,0,0,1,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,1,1,1,1,0,0,1,1,0,1,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,1,0,1,  
0,0,0,0,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,  
0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,1,1,1,0,0,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,1,0,0,1,0,0,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,  
0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,0,1,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,0,0,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,  
0,0,0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
```



```
0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),byrow=T,nrow=50)
```

1	
2	0,0,0,0,0,0,0,1,3,3,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
3	
4	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,0,2,2,0,0,0,0,
5	
6	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,3,3,2,2,0,0,0,
7	
8	0,
9	
10	0,1,3,0,3,2,1,0,1,0,
11	
12	0,1,0,0,0,2,0,0,2,0,0,0,
13	
14	0,2,3,1,0,
15	
16	0,0,0,0,0,3,0,1,0,
17	
18	0,
19	
20	0,
21	
22	0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
23	
24	0,
25	
26	0,0,0,0,0,0,0,0,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
27	
28	0,0,0,3,1,0,3,3,3,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
29	
30	0,0,0,0,3,3,3,2,2,0,
31	
32	0,0,0,0,0,2,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
33	
34	0,1,1,3,0,2,2,0,
35	
36	0,
37	
38	0,0,0,0,0,0,0,0,2,3,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
39	
40	0,0,0,0,0,0,0,0,1,2,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
41	
42	0,0,0,0,0,2,1,0,
43	
44	0,2,1,2,0,0,
45	
46	0,
47	
48	0,
49	
50	0,
51	
52	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,3,0,2,0,1,2,3,2,2,0,0,0,0,
53	
54	0,0,0,0,0,0,0,0,0,0,0,0,0,0,3,3,2,0,0,3,1,2,0,0,0,0,0,0,0,
55	
56	0,0,0,0,0,0,3,3,1,3,3,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
57	
58	0,0,0,0,3,0,
59	
60	0,2,0,0,0,0,0,0,
61	



```

1  {
2  lgr = c(lgr,length(e[i]:K))
3  }
4  lgr = 1:sum(lgr)
5
6  index = matrix(NA,nrow=n,ncol=K)
7  ind=1
8  for (i in 1:n)
9  {
10 for (j in e[i]:K)
11 {
12 index[i,j] = lgr[ind]
13 ind=ind+1
14 }
15 }
16
17 # compute age
18 age <- matrix(NA,nrow=nrow(mydata),ncol=ncol(mydata))
19 for (i in 1:nrow(age)){
20   for (j in e[i]:ncol(age)){
21     if (j == e[i]) age[i,j] <- 1 # yearling
22     if ((j > e[i]) & (j <= e[i]+5)) age[i,j] <- 2 # prime-age
23     if (j > e[i]+5) age[i,j] <- 3 # senescent
24   }
25 }
26
27 #####
28 # 2 - BUGS MODEL
29 #####
30
31 sink("multistatect.bug")
32 cat("
33
34 # BUGS code of the state-space formulation
35 # of capture-recapture models (O. Gimenez, May, 2011)
36
37 # notation used
38
39 # OBSERVATIONS
40 # 0 = female not observed;
41 # 1 = female observed with no fawn;
42 # 2 = female observed with 1 fawn;
43 # 3 = female observed with 2 fawns;
44 # 4 = female observed with 3 fawns.
45
46 # STATES
47 # NB = non-breeding female;
48 # B1 = female breeding with 1 fawn;
49 # B2 = female breeding with 2 fawns;
50 # B3 = female breeding with 3 fawns;
51 # B4 = dead.
52
53 # parameters
54 # phiNB survival prob. of non-breeding individuals / by age
55 # phiB survival prob. of breeding individuals / by age
56 # pNB detection prob. of NB individuals
57 # pB detection prob. of B individuals
58 # psiNBB1 transition prob. from NB to B1
59 # psiB1B2 transition prob. from B1 to B2
60 # psiB2B3 transition prob. from B2 to B3
61 # psiB1B3 transition prob. from B1 to B3

```

```

1 # piNB prob. of being in initial state NB
2
3 model
4 {
5
6 #####
7 # LIKELIHOOD
8 #####
9
10 # probabilities for each initial state
11     px0[1] <- 1 # prob. of being in initial state NB
12     px0[2] <- 0 # prob. of being in initial state B1
13     px0[3] <- 0 # prob. of being in initial state B2
14     px0[4] <- 0 # prob. of being in initial state B3
15     px0[5] <- 0 # because of the conditioning on first capture, prob. of being in initial state dead is 0
16
17 # probabilities of observations at a given occasion given states at this occasion
18     po[1,1] <- 1-pNB
19     po[1,2] <- pNB
20     po[1,3] <- 0
21     po[1,4] <- 0
22     po[1,5] <- 0
23
24     po[2,1] <- 1-pB
25     po[2,2] <- 0
26     po[2,3] <- pB
27     po[2,4] <- 0
28     po[2,5] <- 0
29
30     po[3,1] <- 1-pB
31     po[3,2] <- 0
32     po[3,3] <- 0
33     po[3,4] <- pB
34     po[3,5] <- 0
35
36     po[4,1] <- 1-pB
37     po[4,2] <- 0
38     po[4,3] <- 0
39     po[4,4] <- 0
40     po[4,5] <- pB
41
42     po[5,1] <- 1
43     po[5,2] <- 0
44     po[5,3] <- 0
45     po[5,4] <- 0
46     po[5,5] <- 0
47
48 for (i in 1:N) # for each female
49 {
50
51 # estimated probabilities of initial states are the proportions in each state at first capture occasion
52 alive[i,First[i]] ~ dcat(px0[1:5])
53
54     for (j in (First[i]+1):Years) # loop over time
55     {
56
57 # define age-dependent survival probabilities
58 # if ageij = 1, then phi = phi-young
59 # if ageij = 2, then phi = phi-subadult
60 # if ageij = 3, then phi = phi-adult
61 phiNB[i,j-1] <- phiNBy*equals(age[i,j-1],1)+phiNBsa*equals(age[i,j-1],2)+phiNBa*equals(age[i,j-1],3)

```

```

1  phiB[i,j-1] <- 0 * equals(age[i,j-1],1) + phiBsa * equals(age[i,j-1],2) + phiBa * equals(age[i,j-1],3)
2
3  # probabilities of states at a given occasion given states at previous occasion
4  # uses multinomial logit for the transition probabilities between breeding states
5  px[1,i,j-1,1] <- phiNB[i,j-1] * 1/(1+exp(alpha[1,1])+exp(alpha[1,2])+exp(alpha[1,3]))
6  px[1,i,j-1,2] <- phiNB[i,j-1] * exp(alpha[1,1])/(1+exp(alpha[1,1])+exp(alpha[1,2])+exp(alpha[1,3]))
7  px[1,i,j-1,3] <- phiNB[i,j-1] * exp(alpha[1,2])/(1+exp(alpha[1,1])+exp(alpha[1,2])+exp(alpha[1,3]))
8  px[1,i,j-1,4] <- phiNB[i,j-1] * exp(alpha[1,3])/(1+exp(alpha[1,1])+exp(alpha[1,2])+exp(alpha[1,3]))
9  px[1,i,j-1,5] <- 1-phiNB[i,j-1]
10
11  px[2,i,j-1,1] <- phiB[i,j-1] * exp(alpha[2,1])/(1+exp(alpha[2,1])+exp(alpha[2,2])+exp(alpha[2,3]))
12  px[2,i,j-1,2] <- phiB[i,j-1] * 1/(1+exp(alpha[2,1])+exp(alpha[2,2])+exp(alpha[2,3])) #
13  px[2,i,j-1,3] <- phiB[i,j-1] * exp(alpha[2,2])/(1+exp(alpha[2,1])+exp(alpha[2,2])+exp(alpha[2,3]))
14  px[2,i,j-1,4] <- phiB[i,j-1] * exp(alpha[2,3])/(1+exp(alpha[2,1])+exp(alpha[2,2])+exp(alpha[2,3]))
15  px[2,i,j-1,5] <- 1-phiB[i,j-1]
16
17  px[3,i,j-1,1] <- phiB[i,j-1] * exp(alpha[3,1])/(1+exp(alpha[3,1])+exp(alpha[3,2])+exp(alpha[3,3]))
18  px[3,i,j-1,2] <- phiB[i,j-1] * exp(alpha[3,2])/(1+exp(alpha[3,1])+exp(alpha[3,2])+exp(alpha[3,3]))
19  px[3,i,j-1,3] <- phiB[i,j-1] * 1/(1+exp(alpha[3,1])+exp(alpha[3,2])+exp(alpha[3,3])) #
20  px[3,i,j-1,4] <- phiB[i,j-1] * exp(alpha[3,3])/(1+exp(alpha[3,1])+exp(alpha[3,2])+exp(alpha[3,3]))
21  px[3,i,j-1,5] <- 1-phiB[i,j-1]
22
23  px[4,i,j-1,1] <- phiB[i,j-1] * exp(alpha[4,1])/(1+exp(alpha[4,1])+exp(alpha[4,2])+exp(alpha[4,3]))
24  px[4,i,j-1,2] <- phiB[i,j-1] * exp(alpha[4,2])/(1+exp(alpha[4,1])+exp(alpha[4,2])+exp(alpha[4,3]))
25  px[4,i,j-1,3] <- phiB[i,j-1] * exp(alpha[4,3])/(1+exp(alpha[4,1])+exp(alpha[4,2])+exp(alpha[4,3]))
26  px[4,i,j-1,4] <- phiB[i,j-1] * 1/(1+exp(alpha[4,1])+exp(alpha[4,2])+exp(alpha[4,3]))
27  px[4,i,j-1,5] <- 1-phiB[i,j-1] #
28
29  px[5,i,j-1,1] <- 0
30  px[5,i,j-1,2] <- 0
31  px[5,i,j-1,3] <- 0
32  px[5,i,j-1,4] <- 0
33  px[5,i,j-1,5] <- 1
34
35  ## STATE EQUATIONS ##
36  # draw states at j given states at j-1
37  alive[i,j] ~ dcat(px[alive[i,j-1],i,j-1,1:5])
38
39  ## OBSERVATION EQUATIONS ##
40  # draw observations at j given states at j
41  mydata[i,j] ~ dcat(po[alive[i,j],1:5])
42
43  }
44
45  }
46
47  #####
48  # PRIORS
49  #####
50
51  pNB ~ dunif(0, 1) # non-breeder detectability
52  pB ~ dunif(0, 1) # breeder detectability
53
54  phiBsa ~ dunif(0, 1) # breeder survival sub-adult
55  phiBa ~ dunif(0, 1) # breeder survival adult
56
57  phiNBs ~ dunif(0, 1) # non-breeder survival young
58  phiNBsa ~ dunif(0, 1) # non-breeder survival sub-adult
59  phiNBa ~ dunif(0, 1) # non-breeder survival adult
60
61  # transition probabilities - multinomial logit

```

```

1  for (i in 1:4){
2    for (j in 1:3){
3      alpha[i,j] ~ dnorm(0,0.1)
4    }
5  }
6
7  # for each female at each occasion, store its state
8  for (i in 1:N){
9    for (j in First[i]:Years){
10     lrs[indicateur[i,j]] <- alive[i,j]
11    }
12  }
13
14
15 }
16
17 ",fill=TRUE)
18 sink()
19
20 #####
21 # 3 - BAYESIAN COMPUTATION WITH JAGS
22 #####
23
24 # data
25 mydatax <- list(N=n, Years=K, mydata=as.matrix(mydata+1), First=e, indicateur=index, age=as.matrix(age))
26
27 # first list of inits
28 alive = mydata
29 for (i in 1:n) {
30   for (j in 1:K) {
31     if (j > e[i] & mydata[i,j]==0) {alive[i,j] = which(rmultinom(1, 1, c(1/4,1/4,1/4,1/4))==1)}
32     if (j < e[i]) {alive[i,j] = NA}
33   }
34 }
35 alive <- as.matrix(alive)
36
37 init1 <- list(pB=0.5, phiNBy=0.3, alive=alive)
38 # second list of inits
39 init2 <- list(pB=0.5, phiNBy=0.6, alive=alive)
40 # concatenate list of initial values
41 inits <- list(init1, init2)
42
43 # specify the parameters to be monitored
44 parameters <- c("phiBsa", "phiBa", "phiNBy", "phiNBsa", "phiNBa", "pB", "pNB", "alpha", "lrs")
45
46 # load R package to call JAGS from R
47 library(rjags)
48
49 # run JAGS
50 start<-as.POSIXlt(Sys.time())
51 jmodel <- jags.model("multistatect.bug", mydatax, inits, n.chains = 2, n.adapt = 50000)
52 jsample <- coda.samples(jmodel, parameters, n.iter=10000, thin = 1)
53 end <-as.POSIXlt(Sys.time())
54 duration = end-start
55
56 # save results
57 save(jsample, jmodel, duration, file='agedeer30.Rdata')
58
59 #####
60 # 4 - RESULTS
61 #####

```

```

1
2 # check convergence
3 plot(jsample, trace = TRUE, density = FALSE, ask = dev.interactive())
4 gelman.diag(jsample)
5
6 # numerical summaries and posterior distributions
7 summary(jsample)
8 plot(jsample, trace = FALSE, density = TRUE, ask = dev.interactive())
9
10 #-- display transition probabilities (back-transformation using inverse multinomial logit)
11
12 # psiNBBNB psiNBB1 psiNBB2 psiNBB3
13 a12 <- c(jsample[[1]][, 'alpha[1,1]'], jsample[[2]][, 'alpha[1,1]'])
14 a13 <- c(jsample[[1]][, 'alpha[1,2]'], jsample[[2]][, 'alpha[1,2]'])
15 a14 <- c(jsample[[1]][, 'alpha[1,3]'], jsample[[2]][, 'alpha[1,3]'])
16 a11 <- rep(0, length(a12)) ## ref
17 a1 <- cbind(a11, a12, a13, a14)
18 apply(exp(a1)/apply(exp(a1), 1, sum), 2, mean)
19 apply(exp(a1)/apply(exp(a1), 1, sum), 2, sd)
20
21 # psiB1NB psiB1B1 psiB1B2 psiB1B3
22 a21 <- c(jsample[[1]][, 'alpha[2,1]'], jsample[[2]][, 'alpha[2,1]'])
23 a23 <- c(jsample[[1]][, 'alpha[2,2]'], jsample[[2]][, 'alpha[2,2]'])
24 a24 <- c(jsample[[1]][, 'alpha[2,3]'], jsample[[2]][, 'alpha[2,3]'])
25 a22 <- rep(0, length(a21)) ## ref
26 a2 <- cbind(a21, a22, a23, a24)
27 apply(exp(a2)/apply(exp(a2), 1, sum), 2, mean)
28 apply(exp(a2)/apply(exp(a2), 1, sum), 2, sd)
29
30 # psiB2NB psiB2B1 psiB2B2 psiB2B3
31 a31 <- c(jsample[[1]][, 'alpha[3,1]'], jsample[[2]][, 'alpha[3,1]'])
32 a32 <- c(jsample[[1]][, 'alpha[3,2]'], jsample[[2]][, 'alpha[3,2]'])
33 a34 <- c(jsample[[1]][, 'alpha[3,3]'], jsample[[2]][, 'alpha[3,3]'])
34 a33 <- rep(0, length(a31)) ## ref
35 a3 <- cbind(a31, a32, a33, a34)
36 apply(exp(a3)/apply(exp(a3), 1, sum), 2, mean)
37 apply(exp(a3)/apply(exp(a3), 1, sum), 2, sd)
38
39 # psiB3NB psiB3B1 psiB3B2 psiB3B3
40 a41 <- c(jsample[[1]][, 'alpha[4,1]'], jsample[[2]][, 'alpha[4,1]'])
41 a42 <- c(jsample[[1]][, 'alpha[4,2]'], jsample[[2]][, 'alpha[4,2]'])
42 a43 <- c(jsample[[1]][, 'alpha[4,3]'], jsample[[2]][, 'alpha[4,3]'])
43 a44 <- rep(0, length(a41)) ## ref
44 a4 <- cbind(a41, a42, a43, a44)
45 apply(exp(a4)/apply(exp(a4), 1, sum), 2, mean)
46 apply(exp(a4)/apply(exp(a4), 1, sum), 2, sd)
47
48 #----- compute LRS
49
50 # merge two MCMC chains
51 res <- rbind(jsample[[1]], jsample[[2]])
52 dim(res) # 20000 x 962
53 lrs <- res[, 13:955]
54 dim(lrs)
55
56 # number of simulations
57 nrowarray = dim(lrs)[1]
58
59 # matrix of estimated states
60 lrsind = array(NA, c(nrowarray, n, K))
61 for (k in 1:nrowarray)

```



```

1  {
2  ind=1
3  for (i in 1:n)
4  {
5  for (j in e[i]:K)
6  {
7  lrsind[k,i,j] = lrs[k,index[i,j]]
8  ind=ind+1
9  }
10 }
11 }
12 dim(lrsind)
13
14 lrstot <- matrix(0,nrow=n,ncol=nrowarray)
15 for (kk in 1:nrowarray){
16   for (i in 1:n){
17     for (j in e[i]:K){
18       if (lrsind[kk,i,j] == 1) lrstot[i,kk] <- lrstot[i,kk] + 0
19       if (lrsind[kk,i,j] == 2) lrstot[i,kk] <- lrstot[i,kk] + 1
20       if (lrsind[kk,i,j] == 3) lrstot[i,kk] <- lrstot[i,kk] + 2
21       if (lrsind[kk,i,j] == 4) lrstot[i,kk] <- lrstot[i,kk] + 3
22       if (lrsind[kk,i,j] == 5) lrstot[i,kk] <- lrstot[i,kk] + 0
23     }
24   }
25 }
26
27 # compute LRS, credible intervals and stadard deviation
28 lrsfinal <- apply(lrstot,1,median)
29 lrs.lb <- apply(lrstot,1,quantile,probs=2.5/100)
30 lrs.ub <- apply(lrstot,1,quantile,probs=97.5/100)
31 lrs.sd <- apply(lrstot,1,sd)
32
33 # frequency distribution of LRS
34 lrsfinal.mod <- as.factor(lrsfinal)
35 res <- NULL
36 for(i in levels(lrsfinal.mod))
37 {
38   mask=(lrsfinal.mod==i)
39   res <- c(res,mean(lrs.sd[mask]))
40 }
41 barx <- barplot(table(lrsfinal),col="white",ylim=c(0,60),xlab="Lifetime reproductive success", ylab="Number of
42 individuals")
43 arrows(barx,table(lrsfinal)+res, barx, table(lrsfinal)-res, angle=90, code=3,length = 0.05)
44
45
46

```