Program E-SURGE: A Software Application for Fitting Multievent Models

Rémi Choquet, Lauriane Rouan, and Roger Pradel

Abstract Multievent models (Pradel 2005, 2008) handle state uncertainty, and they therefore cover a range of situations like hidden capture heterogeneity and sex determination from behaviour which cannot be treated in the multistate paradigm. We introduce a new software application called E-SURGE, built upon the concepts developed in program M-SURGE (Choquet et al. 2004) to encompass this new class of capture–recapture models. It also improves on M-SURGE by allowing the decomposition of transitions into several steps. We present the new concepts involved, notably the event and the multistep process, and how they are implemented in E-SURGE. We then illustrate the use of E-SURGE with three examples. One example deals with breeding propensity where the breeding state cannot always be ascertained; a further deals with emigration which is considered as a two-step process (Grosbois and Tavecchia 2003) and the last one with a version of a memory model where survival can be handled directly.

Keywords Capture–Recapture · Hidden Markov Chain

1 Introduction

Capture–recapture (CR) data, i.e. the histories of encounters of individually recognizable animals, have long been a main source of information on the dynamics of animal populations. Their use is regularly enlarged to address new questions like movements, trade-offs between reproduction and survival, transitions among life stages or the spread of a disease. The concept of state plays a key role in these new developments. It is flexible and can indifferently represent the breeding status, the developmental stage or the spatial location. However, it is not always possible to ascertain the state of an individual when it is encountered. For instance, in an epidemiological study, a diseased animal may not present any outer symptoms;

R. Choquet (✉)
Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, CNRS, 1919 Route de Mende, F34293 Montpellier cedex 5, France
e-mail: choquet@cefe.cnrs.fr

conversely, a medical test may be falsely positive. Yet, the multistate models currently used to analyze these data do not allow for uncertainty.

Recently, Pradel (2005) has developed a new approach to handle state uncertainty. This approach introduces a new concept, that of event, which is what is actually known. In our epidemiological study for instance, the event would be the test result or the observed presence or absence of the symptom. There is no one–one correspondance between event and state and a particular event may arise under different states. The new approach is not reducible to the multistate approach. It has logically been named multievent. This paper presents the first program for fitting multievent models to CR data, called E-SURGE (which stands for Multi-Event **SUR**vival **G**eneralized **E**stimation). E-SURGE extends M-SURGE, a previous program for fitting mutistate models developed by the same team.

In addition to handling the different structure required for multievent analysis, E-SURGE also extends M-SURGE in another respect: it allows the specification of transitions between states as a multistep process. For instance, in a study of movement between geographical locations, the transition toward a particular site can be decomposed as the probability to emigrate times the probability to settle on the arrival site given departure. With these two extensions, the new program allows to fit a much greater variety of models and to address in the same unified framework several problems which had been tackled individually:

- heterogeneity of capture (Pledger et al. 2003; Pradel 2005),
- sex uncertainty (Nichols et al. 2004),
- memory in movements (Hestbeck et al. 1991; Brownie et al. 1993; Rouan et al. 2008),
- mixture of live and dead encounters over several sites (Lebreton et al. 1999; Véran et al. 2007).

The motivation for developing program E-SURGE was to bring to biologists the multievent framework in a powerful and user-friendly environment. Several programs exist for CR analysis (Hines 1994; White and Burnham 1999; Choquet et al. 2005) but E-SURGE is the first general program for multievent models. It benefits from the experience gained in developing M-SURGE. M-SURGE introduced a powerful language (Choquet 2008) for describing the set of multistate CR models and used reduced statistics and advanced numerical algorithms to produce faster and more reliable estimates. E-SURGE has similar capabilities for the maximum likelihood optimization of complex age and time-dependent models with linear constraints among parameters in a generalized linear model (GLM) fashion. Its features include:

- A module called GEPAT where the steps making up the transitions (see above for an example) are specified. Similarly, in this same module, events can be declared to be linked to states through a multistep process. For instance, a female animal may be wrongly judged to be a male through three successive steps: it is encountered, its sex is assessed, an error occurs. This detailed decomposition of

the event generation conditional on the state allows for the incorporation in the same analysis of encounters where no judgement is made (Pradel et al. 2008). Multievent models also include initial state probabilities (see below) which can equally be decomposed at this stage.
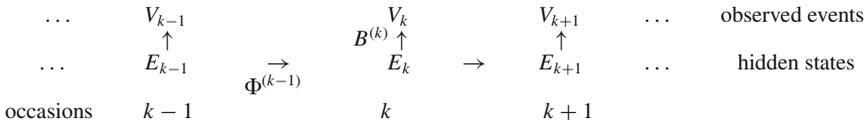
- A module called GEMACO, inherited from M-SURGE, where the mode of variation of the parameters defined in GEPAT is specified by means of a powerful model description language.
- Advanced convergence options. Convergence is a very sensitive issue in multievent as well as in multistate models. In E-SURGE, the user gains a greater control over convergence through a choice of starting options including the results of previous models, random initial values and multiple random initial values.
- Acute rank estimation. E-SURGE calculates, with an algorithm described in Rouan et al. (2008) very precise estimate of the rank of a model. This is one of the key steps for correct model selection using the AIC (Akaike 1987).
- The detection of redundant parameters. E-SURGE analyzes the likelihood in the neighborhood of the point of convergence and lists the parameters that are apparently redundant. Redundancy can then be double-checked by drawing profile likelihood curves.

In the next section, we present briefly the theory of multievent models; then in Sections 3 and 4, the way these models are implemented in E-SURGE in theory and in practice. Three examples are detailed in the application section. The first one concerns breeding status and shows an example of uncertainty in state assessment. The second is the classical Arnason–Schwarz multistate model but the movement between sites is decomposed into two steps. The last is an original version of the memory model where survival probabilities can be handled directly thanks to an appropriate decomposition of the movement.

2 Short Recalls on Multievent Models

Multievent models assume that individuals move independently among a finite set $\mathbb{E}$ of states over a finite number K of sampling occasions, and that the transitions among the successive states obey a Markov chain. The successive states occupied by an individual are not observed directly. Rather, at each occasion k, one member of a finite set Ω of events is observed. The event observed at occasion k is assumed to depend only on the unobserved underlying state of the individual at that occasion (see Fig. 1). Thus, these models belong to the class of Hidden Markov Models (HMM), see for example MacDonald and Zucchini (2000) and Cappé et al. (2005).

Multievent models are defined in terms of three kinds of parameters: initial state probabilities π, transition probabilities ϕ, and event probabilities b. Our presentation of multievent models will use the following general notation which generally follows Pradel (2005):

$\dots$	V_{k-1}		V_k		V_{k+1}	$\dots$	observed events
	$\uparrow$		$B^{(k)} \uparrow$		$\uparrow$		
$\dots$	E_{k-1}	$\xrightarrow{\Phi^{(k-1)}}$	E_k	$\rightarrow$	E_{k+1}	$\dots$	hidden states
occasions	$k-1$		k		$k+1$		

Fig. 1 Graph summarising the relations of dependence between successive states (transition matrix Φ) and observed events (event generation matrix B) in the multievent time-dependent model. V_k is the random variable of the event observed at occasion k, E_k that of the underlying state of the animal at the same occasion

N	the number of states.
U	the number of events.
K	the number of occasions.
K	the number of occasions.
A	the maximum age class.
NG	the number of groups.
LI, LT, LB	the number of step processes for initial state, transition and event.
$i = 1, \dots, N$	the index of the previous (or departure) state.
$j = 1, \dots, N$	the index of the current (or arrival) state.
$u = 1, \dots, U$	the index of the current event.
$k = 1, \dots, K$	the occasion index.
$a = 1, \dots, A$	the index of current age classes.
$ng = 1, \dots, NG$	the index of the current group.
$\ell = 1, \dots,$ LI, LT, or LB	the step process index
$\mathbb{E} = \{e_1, ..., e_N\}$	the set of states, where $e_N = $ '†' for the death.
$\Omega = \{v_1, ..., v_U\}$	the set of events, where the $v_1 = $ 'not seen'

Unlike traditional practice in CR but similar to Fujiwara and Caswell (2002), Pradel (2005) and consistent to Markov Chain practice MacDonald and Zucchini (2000), the dead (†) is explicity included in $\mathbb{E}$. All transition matrices are written with i as row index and j as column index, thus following the Markov chain convention in which transitions are from rows to columns. Event matrices use j (denoting the state) as row index and u (denoting an event) as column index. Transition and event matrices are row-stochastic.

For the time-dependent model, the elementary parameters are

- $\pi_i^{(k)}$ the probability of being in state e_i when first encountered at occasion k,
- $\phi_{ij}^{(k)}$ the probability of being in state e_j at time $k+1$ if in state e_i at occasion k,
- $b_{ju}^{(k,1)}$ the probability of event v_u for an animal in state e_j encountered for the first time at occasion k,
- $b_{ju}^{(k,2)}$ the probability of event v_u for an animal reencountered in state e_j at occasion k,
- $\mathbf{\Pi} = (\pi_i)$ denotes the $(1 \times N)$ vector of initial state probabilities,

- $\mathbf{\Phi} = (\phi_{ij})$ denotes the $(N \times N)$ matrix of unconditional transition probabilities, i.e. the matrix of probabilities that an individual moves from one state to another state over a time interval.
- $\mathbf{B} = (b_{ju})$ denotes the $(N \times U)$ matrix of event probabilities. There are potentially two such matrices, one for first encounters $(b_{ju}^{(k,1)})$ and one for reencounters $(b_{ju}^{(k,2)})$. They may as well be considered as instances of the same matrix corresponding to two different age-classes. This is what we will do thereafter.

Together, $\mathbf{\Pi}$, $\mathbf{\Phi}$ and $\mathbf{B}$ define the multievent model. A particular multievent model will be obtained by specifying the mode of variation (by time, age, group or state...) of each type of parameter.

The likelihood of a model is proportional to the probability of the data given that model. The basic unit of data in E-SURGE is the capture history. Let $h = (o_1, \ldots, o_K)$ be a capture history with first encounter at occasion e, event o_k, $k = e, \ldots, K$, has any value between 1 and U and β is the vector of all parameters. Then

$$P(h|\beta) = \mathbf{\Pi}^{(e)} D\left(\mathbf{B}^{(e,1)}(\cdot, o_e)\right) \left(\prod_{k=e+1}^{K} \mathbf{\Phi}^{(k-1)} D\left(\mathbf{B}^{(k,2)}(\cdot, o_k)\right) \right) \mathbf{1}_N \qquad (1)$$

where $\mathbf{B}^{(k,2)}(\cdot, o_k)$ is the o_kth column of the re-encounter matrix at occasion k, i.e. the instance of $\mathbf{B}$ corresponding to the second age-class at occasion k. $D(x)$ denotes the matrix with the elements of vector $\mathbf{x}$ on the diagonal and zeros elsewhere, and $\mathbf{1}_N$ is an N vector of ones.

Assuming that individuals are independent, the likelihood for the entire set of capture histories is obtained as the product of the likelihoods for each history,

$$L(\beta) = \prod_h P(h|\beta)^{n_h}$$

where n_h is the number of copies of capture history h in the data set.

3 Concepts in E-SURGE

3.1 Decomposition in Elementary Steps

In E-SURGE, all types of parameters (transition probabilities, initial state probabilities and event probabilities) may be decomposed in an indefinite number of steps as arising from a sequence of 'life processes'. The first task is to specify the actual structure that will be used. The familiar decomposition of the transition matrix into survival and transition conditional on survival (implemented in M-SURGE) is the classical example in CR, but in some cases more steps may be involved. For instance, if we want to model the movement between geographical locations with

probabilities of emigration and settlement on the new site, the transition probabilities between locations has to be decomposed into two steps. In the second step a migrant will not be allowed to settle on its departure site. Thus, in addition to specifying that there are three steps in total (one for survival and two for movement), we will also have to forbid certain transitions. These specifications are mainly done in module GEPAT.

In E-SURGE, the matrices $\mathbf{\Pi}$, $\mathbf{\Phi}$, $\boldsymbol{B}$ are called the full initial state vector, the full transition matrix, and the full event matrix, respectively. In an approach similar to that used for periodic matrix population models (Caswell 2001, Chapter 14) but between two states, the full matrices are written as products of elementary matrices.

$$\mathbf{\Pi} = \prod_{\ell=1}^{LI} \mathbf{\Pi}^{(\ell)}$$

$$\mathbf{\Phi} = \prod_{\ell=1}^{LT} \mathbf{\Phi}^{(\ell)} \qquad (2)$$

$$\boldsymbol{B} = \prod_{\ell=1}^{LB} \boldsymbol{B}^{(\ell)}$$

For Equation (2), we will say that we need a DES(LI,LT,LB) (Decomposition in Elementary Steps), where LI, LT and LB define the number of steps for the matrices $\mathbf{\Pi}$, $\mathbf{\Phi}$, $\boldsymbol{B}$, respectively.

3.2 Umbrella Model

After having specified the elementary steps, we have defined what we call the umbrella model (UM). The parameters of the UM are further constrained to compare different biological hypotheses of interest. In other words, the UM is the most general model that can be fitted, and the one within which all other models examined are nested.

There are six potential sources of variation in the parameters:

1. previous state,
2. current state,
3. current event,
4. time, i.e. occasions elapsed since the beginning of the capture session,
5. age, i.e., occasions elapsed since first capture,
6. groups, i.e. permanent categories of individuals, such as sexes or species, or discrete unconnected study sites.

In the UM, parameters are allowed to vary freely over time and among groups. Only the number of states and number of age classes can be set to specific values.

Survival, transitions, and encounter probabilities may depend on age (i.e., time since first capture, not necessarily true chronological age). The user specifies an oldest relevant age class; all animals with this age or older are combined into a single age-class. Constraining the range of ages restricts the range of models that can be fitted but it can greatly save memory and reduce computation time. For transitions or survival, common choices for the maximum relevant age are 1 ($A = 1$), which implies no age effect, and 2 ($A = 2$), which creates a model in which the first age class is contrasted to animals older than one year (this is particularly useful when animals are marked as young). Setting $A = 2$ can also be used to model transience (Pradel et al. 1997).

Specifying age-dependence in encounters is slightly more complicated. In multistate (as opposed to multievent) applications, all calculations are conditional on the first encounter and hence the probability of first encounter is not estimated. In multievent applications, E-SURGE has the option of modelling the probability of the initial event $b_{ju}^{(k,1)}$. Therefore, E-SURGE always considers at least two age classes for encounters, allowing the first event probability (first age class) to be modelled or not (probability to be first captured fixed to 1). Thus if one chooses a maximum age $A = 1$, which implies no age effect, E-SURGE creates nonetheless 2 age classes for events. If one sets $A = 2$, E-SURGE creates 3 age classes for events (see Table 1).

3.3 Constrained Models

Model-building in E-SURGE (as in M-SURGE) proceeds by imposing linear constraints on the parameters of the umbrella model in the spirit of GLM (Lebreton et al. 1992). The vector θ of 'biological parameters' (parameters of direct interest to the biologist e.g., $\theta = (\Pi, \Phi, B)$, organized as a vector) is expressed as a linear transformation of a vector β of 'mathematical parameters'. To keep the biological parameters, which are probabilities, in their permissible range (0,1), a link function f is generally applied:

$$f(\theta) = X\beta$$

The matrix X is a 'matrix of constraints' or genuine design matrix in the case of a designed experiment. In general, it expresses hypotheses about the dependence of the parameters on state (of departure or arrival), age, time, group, covariates, and

Table 1 Variations considered in the parameters of the umbrella models of E-SURGE. The type of variation is represented by upper indices for time(k), age(a), group(g) and step(l)

$A \neq 1$	$\Pi^{(k,g,(l))}$
	$\Phi^{(k,a \leq A,g,(l))}$
	$B^{(k,a \leq A+1,g,(l))}$
$A = 1$	$\Pi^{(k,g,(l))}$
	$\Phi^{(k,a=1,g,(l))}$
	$B^{(k,a \leq 2,g,(l))}$

so on. The design matrix is built by the module GEMACO (**GE**nerator of **MA**trices of **CO**nstraints), using the model definition language described in Choquet et al. (2004) and Choquet (2008). Often X will contain both discrete indicator 0/1 variables (for equality constraints) and continuous covariates (e.g., effort or weather covariates). An overview of linear constraints in CR models with a single state is given by Lebreton et al. (1992), linear constraints in multistate models are considered in Choquet (2008).

An important difference in the application of GEMACO in E-SURGE, as compared to M-SURGE, is that the GEMACO keywords ('from', 'to', etc.) refer to the elementary matrices in E-SURGE. The rows (from) and columns (to) in these matrices do not necessarily correspond to the states in the model (e.g., in the encounter matrix, the columns refer to events, not states), whereas in M-SURGE they always correspond to states.

3.4 Maximum Likelihood Estimation

Reduced-form data descriptions like the *m*-array are not available for multievent models. Thus the likelihood calculation depends on the application of the transition and event probabilities to individual capture histories. The maximum likelihood algorithm is as follows

1. Calculate the number of identical histories.
2. Select initial values for the vector β of mathematical parameters.
3. Calculate the vector of biological parameters $\boldsymbol{\theta} = f^{-1}(X\beta)$, with f the generalized (or multinomial) logit link or the identity link.
4. Calculate the elementary matrices, and (as the product of the elementary matrices) each of the full matrices $\boldsymbol{\Pi}$, $\boldsymbol{\Phi}$, and $\boldsymbol{B}$ according to Equation (2).
5. Use the full matrices $\boldsymbol{\Pi}$, $\boldsymbol{\Phi}$, and $\boldsymbol{B}$ to calculate the individual probability $P(h|\beta)$ of each capture history according to equation (1).
6. Calculate the deviance

$$\mathrm{Dev}(\beta) = -2\log L(\beta) = -2\sum_h n_h \log P(h|\beta)$$

7. Iterate steps 3–6 in a Quasi-Newton minimization method updating the vector of mathematical parameters to decrease the deviance, until convergence.
8. Obtain the MLE's and the deviance and various by-products of Maximum Likelihood estimation like AIC (Akaike 1987) and derivate quantities (QAIC, QAICc).

Note 1 The choice of initial values is critical in multievent models as the problem of local minima is worse than for multistate (Lebreton and Pradel 2002). For that purpose, advanced features were developed in E-SURGE. Users can choose between several options for initial values, among them are random initial values, starting

values estimated from a previous model and starting values given in files. In the context of Markov chains, the well known Expectation Maximization (EM) algorithm has also poor properties regarding local minima, see Celeux et al. (1995).

Note 2 AIC or derivate quantities (Burnham and Anderson 2002) permit model selection inside an umbrella model. A likelihood ratio test between two models under two different umbrella models is not currently available and would require bootstrapping. We emphasize the difficulty to do bootstrapping in the presence of local minima. For a more complete discussion about model selection, see Burnham and Anderson (2002) and Pledger et al. (2003).

4 E-SURGE in Practice

In this section we present three examples illustrating the potential of E-SURGE. The first example with breeding propensity is a good example to demonstrate its application when uncertain states occur; the second one considers the decomposition of the survival-transition probabilities into several steps and the last one combines the two previous issues by showing the implementation of a particular case of memory model.

4.1 First Example: A Simple Model for Studying Breeding Propensity

The study of transitions between the states breeder (Br) and non-breeder (NBr) is a topic of major interest to biologists. Nevertheless, the assessment of the breeding status can be difficult; indeed animals are not always observed in breeding activity and their status can remain unknown which motivates the use of multievent model. Let us define formally the underlying multievent model.

To be consistent with CR practice, events are originally coded in the encounter histories as:

   0 for $v_1 =$ 'not seen';
   1 for $v_2 =$ 'seen breeding';
   2 for $v_3 =$ 'seen but status unknown'.

Inside E-SURGE, each event code is increment by 1 to be consistent with Equation (1) but this is transparent for the user.

According to the available information, the 'hidden states' chosen to define the model are:

   Br for 'breeder';
   NBr for 'non breeder';
   † for 'dead'.

We have first to specify the parameterization of the model (i.e $\mathbf{\Pi}$, $\mathbf{\Phi}$, and $\boldsymbol{B}$):

- The initial state probabilities: at the first encounter, the individuals can be 'breeder' or 'non breeder' but cannot be 'dead'.

$$\mathbf{\Pi} = \begin{pmatrix} \text{Br} & \text{NBr} & \dagger \\ \pi_B & 1 - \pi_B & 0 \end{pmatrix},$$

- The survival-transition probabilities: all the transitions between states Br and NBr are possible. Transitions between Br and NBr to † are constraint to sum to one, and transitions between † and Br and between † and NBr are not possible:

$$\mathbf{\Phi} = \begin{array}{c} \text{Br} \\ \text{NBr} \\ \dagger \end{array} \begin{pmatrix} \text{Br} & \text{NBr} & \dagger \\ \phi_{\text{Br,Br}} & \phi_{\text{Br,NBr}} & 1 - \phi_{\text{Br,Br}} - \phi_{\text{Br,NBr}} \\ \phi_{\text{NBr,Br}} & \phi_{\text{NBr,NBr}} & 1 - \phi_{\text{NBr,Br}} - \phi_{\text{NBr,NBr}} \\ 0 & 0 & 1 \end{pmatrix}$$

- The event probabilities: in this particular context, a 'breeder' can be 'not seen', 'seen breeding' or 'seen but status unknown'; a 'non breeder' is either 'not seen' or 'seen but status unknown' and a 'dead' individual can only be 'not seen'.

$$\mathbf{B}^{(a=2)} = \begin{array}{c} \text{Br} \\ \text{NBr} \\ \dagger \end{array} \begin{pmatrix} 0 & 1 & 2 \\ 1 - p_{1|\text{Br}} - p_{2|\text{Br}} & p_{1|\text{Br}} & p_{2|\text{Br}} \\ 1 - p_{\text{NBr}} & 0 & p_{\text{NBr}} \\ 1 & 0 & 0 \end{pmatrix} \quad (3)$$

Since the first encounter of the individuals is modelled explicitly here, one needs to pay attention to the initial-event probabilities. At first encounter, all individuals alive are obligatorily observed which leads to fix to 0 the probabilities of the event 'not seen'.

$$\mathbf{B}^{(a=1)} = \begin{array}{c} \text{Br} \\ \text{NBr} \\ \dagger \end{array} \begin{pmatrix} 0 & 1 & 2 \\ 0 & p^0_{1|\text{Br}} & 1 - p^0_{1|\text{Br}} \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

To fit this model in E-SURGE we have to execute five main stages. In the first stage, we implement the structure of the model (see Fig. 2) with:

1. the definition of the number of events and states;
2. the definition of the general structure of the matrices of parameters using the GEPAT interface. The different matrices are represented symbolically using

(a)
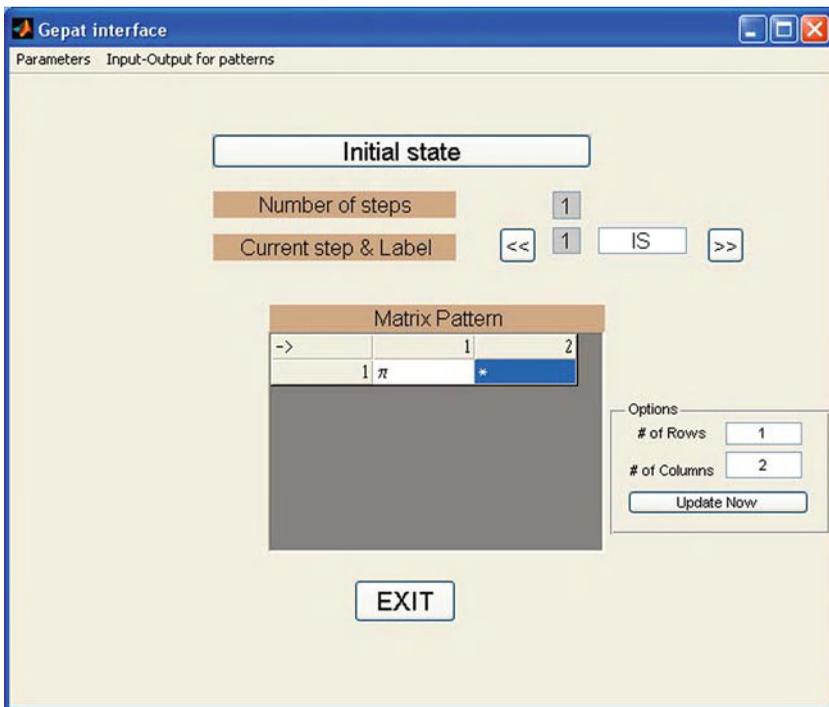
Fig. 2 In the window (1), we specified the number of states ($N = 3$), events ($U = 3$), age classes ($A = 1$) and groups ($G = 1$). The GEPAT interface is then opened. Patterns of $\mathbf{\Pi}$, $\mathbf{\Phi}$ and $\mathbf{B}$ are defined in windows 2a, 2b, 2c, respectively using the following rules: the parameters of interest are indicated using any greek letter, the parameters constraint to zero are indicated using '-' and the complementary parameters are indicated using '*'

an 'excel-like' interface; this step consists in filling different cells using the following rules:

- a greek letter indicates a parameter of interest i.e. one that will be estimated or fixed;
- '*' indicates the complementary parameter (there is one and only one '*' by row due to the row-stochasticity of the matrices);
- '–' indicates parameters constrained to zero.

Because the pattern must be the same for the two kinds of event matrices $\mathbf{B}^{(a=1)}$ and $\mathbf{B}^{(a=2)}$, which are considered by the program as two instances of the same matrix corresponding to two different age-classes. A particular attention must be payed for this example, to the matrices of the event probabilities. For $\mathbf{B}^{(a=1)}$, we need to fix to 0 the probabilities of the event 'not seen' conditioning respectively on the states Br and NBr. This constraint leads us to consider this cell as a parameter of interest rather than a complementary parameter. As a consequence,
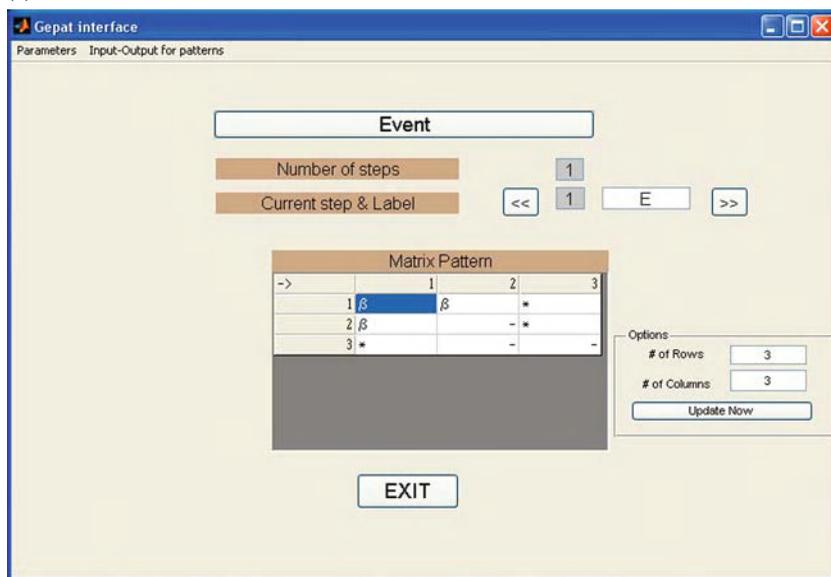
(b)
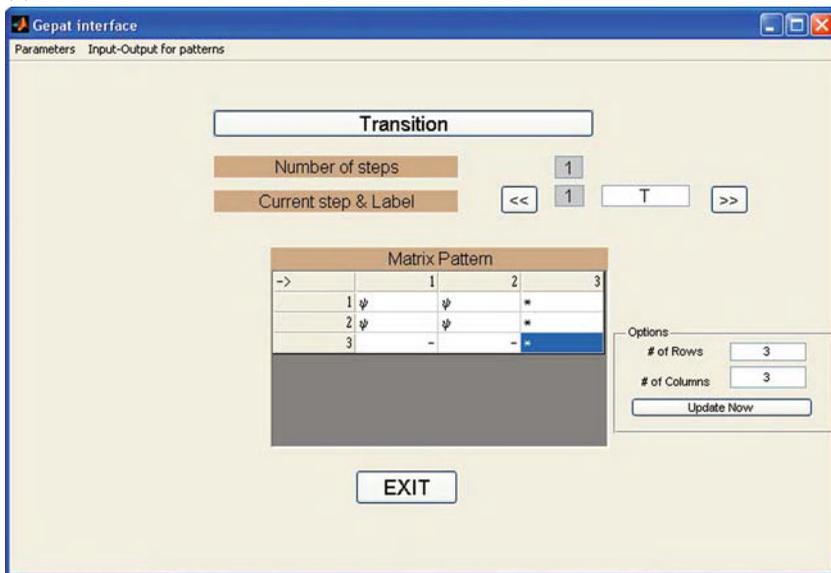
(c)

Fig. 2 (continued)

the pattern used in E-SURGE is not a direct transcription of the previous matrix $\boldsymbol{B}^{(a=2)}$ as it was the case for the vector $\boldsymbol{\Pi}$ or the matrix $\boldsymbol{\Phi}$.

In the second stage, we define the constraints of the model using the GEMACO interface (Choquet 2008) as shown in Fig. 3. Once again, particular attention must be payed to constraints applied to the event probabilities: the first encounter corresponds to the first age-class of b (denoted 'a(1)' or 'firste' in the GEMACO syntax)

(a)
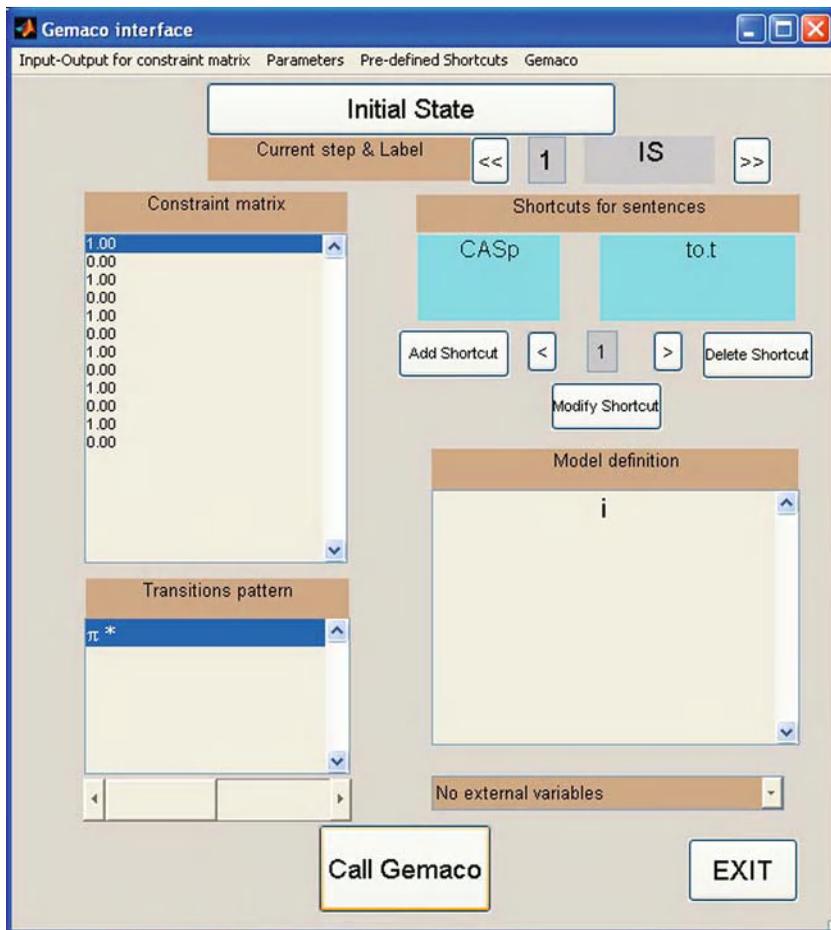
Fig. 3 The GEMACO interface defines constraints for the UM. In window (**a**), the user has entered the phrase 'i' to define a constant initial state probability. In window (**b**), the user has entered the phrase 'from.to.t' to define a state and time dependent effect for survival-transition. In window (**c**), the user has entered the phrase 'firste+a(2).from.to'. The first mathematical parameter corresponding to 'firste' is fixed to 0 in the next stage. The parameters $p_{1|Br}$, $p_{2|Br}$, p_{NB} in equation (3) are set to be different and constant across time with the phrase 'a(2).from.to'
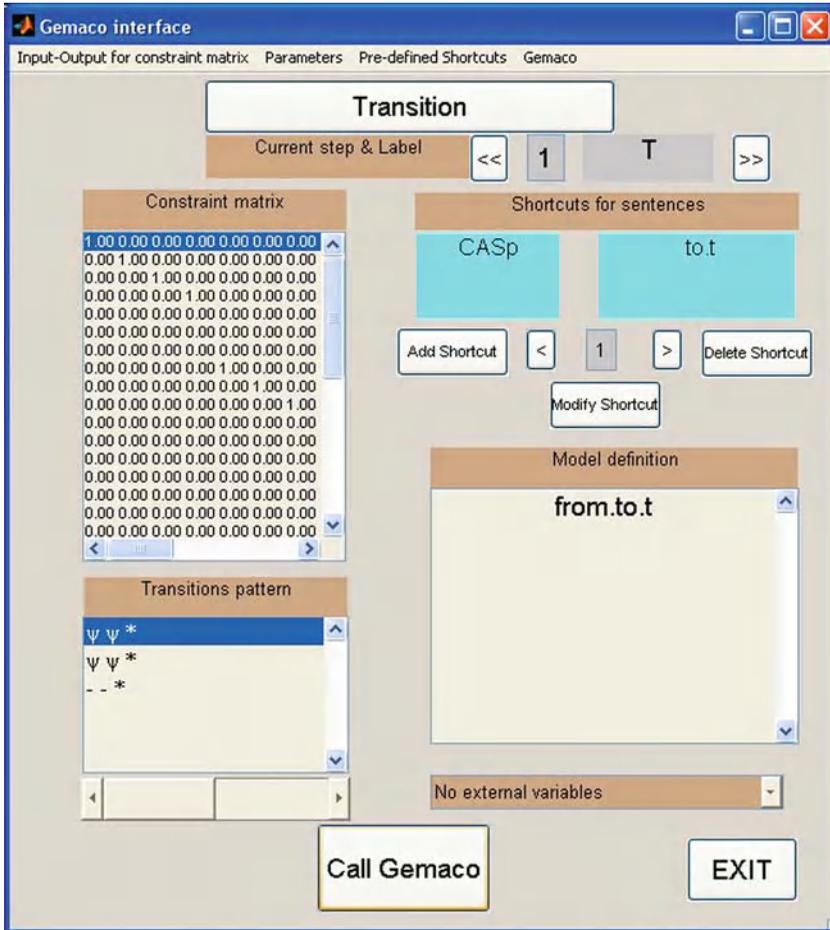
(b)

Fig. 3 (continued)

and the next encounters corresponds to the second age-class of *b* (denoted 'a(2)' or 'nexte' in the GEMACO syntax).

In the next three stages, we proceed like in M-SURGE (Choquet et al. 2004):

- we fix the parameter of **B** corresponding to the probability to be 'not seen' when first encountered to 0 (in the probabilities scale) and change the initial values if needed using the IVFV interface;
- we run the model;
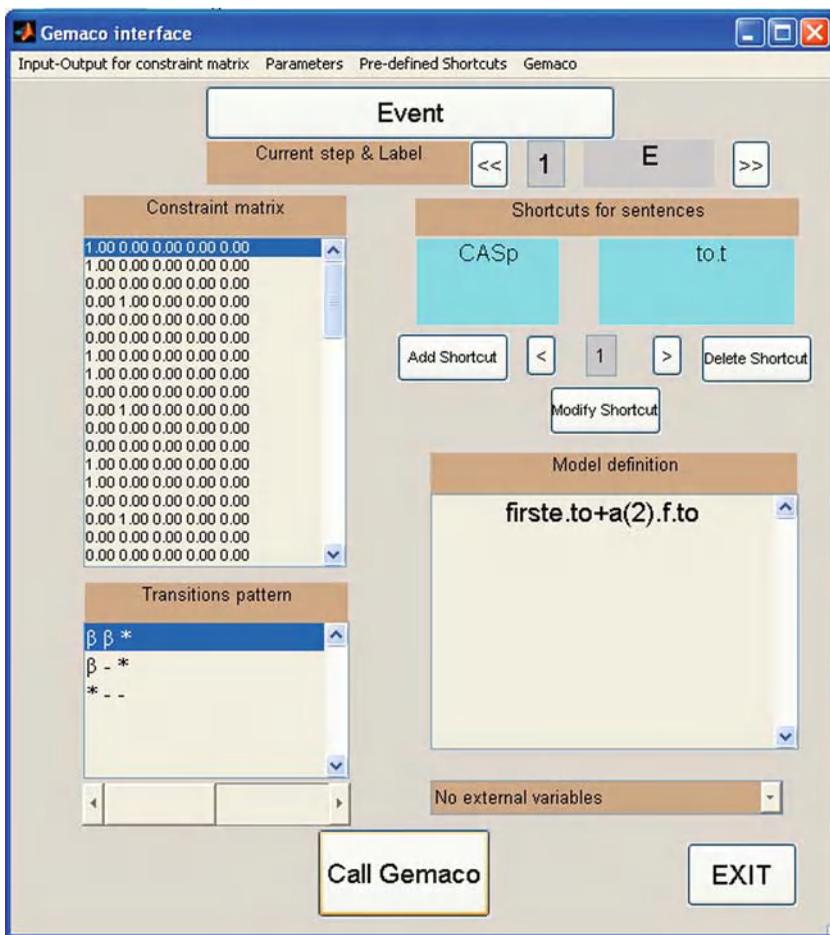- we examine and interpret the results.

(c)

Fig. 3 (continued)

4.2 Second Example: A Version of the Arnason–Schwarz Model Incorporating Site Fidelity

Now we consider a version of the Arnason–Schwarz model in which the probability of transition conditional on survival is further subdivided into a probability of leaving the site (the complement of site fidelity), and a probability of moving to each other site conditional on leaving (Grosbois and Tavecchia 2003). For this example, we need a DES(1,3,1). With 3 geographical sites and assuming that if an

animal is seen, its site is known without error, the set of events (i.e., the results of observations) is

$$\Omega = \{\text{'not seen', 'seen at 1', 'seen at 2', 'seen at 3'}\}$$

In this case, we will see that the intermediate states involved in the sequence of life processes are not the same as the basic set of states in the model. Thus the elementary matrices are not necessarily square.

4.2.1 Defining the Sets of Intermediate States

In the classical separate formulation of the Arnason–Schwarz model, the set of possible states for an individual is the same for both elementary matrices (survival and transition conditional on survival). However, as we are going to see, there may be a different set of states at each of the elementary steps. It can be helpful to use the graphical formulation of periodic matrix models (Caswell 2001) to construct the elementary matrices and their states sets. Figure 4 shows the formulation of the Grosbois and Tavecchia (2003) model, where the states of each elementary matrix are denoted as numbered nodes on a row. Each step in the life process is represented by a subsequent row, and the possible transitions are denoted by arrows. The initial set of states at time $k + 1$ is reconstituted at the bottom of the graph.
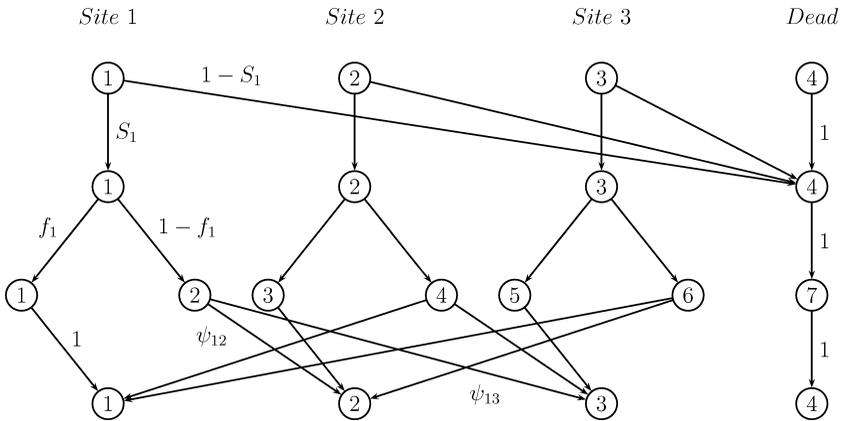
Fig. 4 Graphical representation of the Grosbois and Tavecchia (2003) model. Shown are the transitions for survival, fidelity given survival, and destination given movement. The elementary transition probabilities are shown on the pathways originating from site 1 and dead (at the top); the transition probabilities from the other sites 2 and 3 follow the same pattern. The first row-stochastic matrix $\mathbf{\Phi}^{(1)}$ maps row 1 onto row 2 of the diagram (survival step S). The matrix $\mathbf{\Phi}^{(2)}$ maps row 2 onto row 3 (fidelity given survival, f), and the matrix $\mathbf{\Phi}^{(3)}$ maps row 3 onto row 4 (destination given movement, ψ). In row 3, for practical reason, the numbers 1–7 do not refer to the original states. For instance, the intermediate state 2 corresponds to an individual that has moved from site 1 and the intermediate state 3 corresponds to an individual that has remained in site 2

In the Grosbois and Tavecchia (2003) site fidelity model, the sets of states are

$$\mathbb{E}^{(0)} = \{\text{'site 1', 'site 2', 'site 3', '}\dagger\text{'}\}$$
$$\mathbb{E}^{(1)} = \{\text{'site 1', 'site 2', 'site 3', '}\dagger\text{'}\}$$
$$\mathbb{E}^{(2)} = \{\text{'staying in 1', 'leaving 1', 'staying in 2', 'leaving 2',}$$
$$\text{'staying in 3', 'leaving 3', '}\dagger\text{'}\}$$

4.2.2 Matrices

The initial state matrix is

$$\mathbf{\Pi}^{(k)} = \begin{pmatrix} \pi_1^{(k)} & \pi_2^{(k)} & 1 - \pi_1^{(k)} - \pi_2^{(k)} & 0 \end{pmatrix}$$

The first elementary transition matrix (for survival) maps $\mathbb{E}^{(0)}$ onto $\mathbb{E}^{(1)}$, and hence is a 4×4 matrix:

$$\mathbf{\Phi}^{(k,(1))} = \begin{pmatrix} s_1^{(k)} & 0 & 0 & 1 - s_1^{(k)} \\ 0 & s_2^{(k)} & 0 & 1 - s_2^{(k)} \\ 0 & 0 & s_3^{(k)} & 1 - s_3^{(k)} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The second elementary matrix (for site fidelity given survival) maps from $\mathbb{E}^{(1)}$ to $\mathbb{E}^{(2)}$, and hence is a 4×7 matrix. Letting f_i be the probability of remaining in site i given survival, we have

$$\mathbf{\Phi}^{(k,(2))} = \begin{pmatrix} f_1^{(k)} & 1 - f_1^{(k)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & f_2^{(k)} & 1 - f_2^{(k)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & f_3^{(k)} & 1 - f_3^{(k)} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The third elementary matrix (for movement conditional on emigration, ψ) maps from $\mathbb{E}^{(2)}$ back to $\mathbb{E}^{(0)}$, and so is a 7×4 matrix:

$$\mathbf{\Phi}^{(k,(3))} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \psi_{12}^{(k)} & 1 - \psi_{12}^{(k)} & 0 \\ 0 & 1 & 0 & 0 \\ \psi_{21}^{(k)} & 0 & 1 - \psi_{21}^{(k)} & 0 \\ 0 & 0 & 1 & 0 \\ \psi_{31}^{(k)} & 1 - \psi_{31}^{(k)} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The event matrices $\boldsymbol{B}$ map from the set $\mathbb{E}^{(0)}$ of states to the set Ω of events, and thus are of dimension 4×4.

$$\boldsymbol{B}^{(a=1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{B}^{(k,a=2)} = \begin{pmatrix} 1 - p_1^{(k)} & p_1^{(k)} & 0 & 0 \\ 1 - p_2^{(k)} & 0 & p_2^{(k)} & 0 \\ 1 - p_3^{(k)} & 0 & 0 & p_3^{(k)} \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Note The choice of states for the intermediate transitions is not always unique. There may be more than one way to group individuals, and at present the only advice we can give is to determine from the structure of the model what information needs to be kept at any one step in order to define the probability of subsequent transitions.

4.3 Third Example: The Separate Formulation of the Memory Model

Finally, we consider a version of the memory model given in Pradel (2005) in which the survival-transition probability is further subdivided into the probability of survival and the probability of movement from each site to the others conditional on surviving. We need a DES(1,2,1). This formulation is the analogue of the separate formulation of the conditional Arnason–Schwarz model.

With the 2 geographical sites 1 and 2, $\mathbb{E}$ consists of 5 states {'11', '12', '21', '22', '†'}. State '11' denotes presence at site 1 at time $k-1$ and k, state '12' denotes presence at site 1 at time $k-1$ and presence at site 2 at time k, and so on. The set of states remains constant across the two life processes steps

$$\mathbb{E}^{(0)} = \mathbb{E}^{(1)} = \{\text{'11', '12', '21', '22', '†'}\}$$

Assuming that if an animal is seen and its current site is known without error, the set of events (i.e., the results of observations) is

$$\Omega = \{\text{'not seen', 'seen at 1', 'seen at 2'}\}$$

The initial state matrix is

$$\boldsymbol{\Pi}^{(k)} = \begin{pmatrix} \pi_{11}^{(k)} & \pi_{12}^{(k)} & \pi_{21}^{(k)} & \pi_{22}^{(k)} & 0 \end{pmatrix}$$

There are now two elementary transition matrices, one corresponding to survival(noted s) and one to movements (noted ψ) conditional on survival:

$$\mathbf{\Phi}^{(k,(1))} = \begin{pmatrix} s_{11}^{(k)} & 0 & 0 & 0 & 1 - s_{11}^{(k)} \\ 0 & s_{12}^{(k)} & 0 & 0 & 1 - s_{12}^{(k)} \\ 0 & 0 & s_{21}^{(k)} & 0 & 1 - s_{21}^{(k)} \\ 0 & 0 & 0 & s_{22}^{(k)} & 1 - s_{22}^{(k)} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{\Phi}^{(k,(2))} = \begin{pmatrix} \psi_{111}^{(k)} & 1 - \psi_{111}^{(k)} & 0 & 0 & 0 \\ 0 & 0 & 1 - \psi_{122}^{(k)} & \psi_{122}^{(k)} & 0 \\ \psi_{211}^{(k)} & 1 - \psi_{211}^{(k)} & 0 & 0 & 0 \\ 0 & 0 & 1 - \psi_{222}^{(k)} & \psi_{222}^{(k)} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

There are two elementary event matrices. The first one is fixed and the second is varying:

$$\mathbf{B}^{(a=1)} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{B}^{(k,a=2)} = \begin{pmatrix} 1 - p_1^{(k)} & p_1^{(k)} & 0 \\ 1 - p_2^{(k)} & 0 & p_2^{(k)} \\ 1 - p_1^{(k)} & p_1^{(k)} & 0 \\ 1 - p_2^{(k)} & 0 & p_2^{(k)} \\ 1 & 0 & 0 \end{pmatrix}$$

This model is an extension of the model presented in Pradel et al. (2008), but the model of Brownie et al. (1993) can also be decomposed in the same way.

5 Discussion and Perspectives

In this paper, we have presented E-SURGE, the first program to fit non-homogeneous hidden Markov chains models in CR, called multievent models. This program has an interface (named GEPAT) to build various parameterizations and goes far beyond the classical combined and separate parameterizations. We can consider any product of elementary transition matrices, any product of elementary encounter matrices and any product of elementary initial matrices, which allows great flexibility to

fit different kind of models. This approach permits incorporation of multievents models in a GLM framework and avoids dealing with non-linear constraints, which are difficult to manage. GEPAT coupled with GEMACO for model constraints gives an easy, flexible and powerfull way to construct models, that are useful to address biological questions.

Advanced features of E-SURGE are not described in this paper, yet we enumerate some of them. Essential but hidden are elaborate algorithms used to fit a model, compute its rank and detect redundant parameters. More visible are the results stored in a spreadsheet file, which allows an easy access to the estimated parameters of all fitted models for further calculations or for drawing graphs. Lastly, the post allocation of classes proposed in Pledger et al. (2003) is also available.

We are continuously exploring new features with working versions. We hope to have soon implemented the profile deviance (and more generally all the tools already implemented in M-SURGE), estimation of the Lifetime Reproductive Success (Rouan et al. 2008), the possibility to model individual covariates (with fixed or random effect) as a promising link with genetics (Gimenez et al. 2006) and the Stochastic EM algorithm (Celeux et al. 1995) to deal with local minima in a more efficient way.

6 Program Availability

Program E-SURGE can be downloaded freely from the WWW at
http://www.cefe.cnrs.fr/BIOM/logiciels.htm
Instructions for installation and a user guide are provided.

References

Akaike H (1987) Factor analysis and AIC. Psychometrika 52:317–332.
Brownie C, Hines J, Nichols J, Pollock K, Hestbeck J (1993) Capture–recapture studies for multiple strata including non-markovian transitions. Biometrics 49:1173–1187.
Burnham K, Anderson D (2002) Model selection and inference: A practical information-theoretic approach, 2nd edition. Springer-Verlag, New York.
Cappé O, Moulines E, Rydn T (2005) Inference in Hidden Markov models (Springer series in Statistics). Springer, New York.
Caswell H (2001) Matrix population models 2nd edition., Sinauer Associates, Sunderland, MA.
Celeux G, Chauveau D, Diebolt J (1995) On stochastic versions of the EM algorithm. Technical Report 2514, INRIA Rhone-Alpes.
Choquet R (2008) Automatic generation of multistate capture–recapture models. The Canadian Journal of Statistics 36(1):43–57.
Choquet R, Reboulet AM, Pradel R, Gimenez O, Lebreton JD (2004) M-SURGE: new software specifically designed for multistate capture–recapture models. Animal Biodiversity and Conservation 27(1):207–215.

Choquet R, Reboulet A, Pradel R, Gimenez O, Lebreton JD (2005) M-SURGE: 1.8 user's manual. Technical report, CEFE UMR 5175, Mondpellier.

Fujiwara M, Caswell H (2002) Estimating population projection matrices from multi-stage mark-recapture data. Ecology 83(12):3257–3265.

Gimenez O, Covas R, Brown C, Anderson M, Brown M, Lenormand T (2006) Nonparametric estimation of natural selection on a quantitative trait using mark-recapture data. Evolution 60(3):460–466.

Grosbois V, Tavecchia G (2003) Modeling dispersal with capture–recapture data: Disentangling decisions of leaving and settlement. Ecology 84(5):1225–1236.

Hestbeck J, Nichols J, Malecki R (1991) Estimates of movement and site fidelity using mark-resight data of wintering Canada Geese. Ecology 72:523–533.

Hines JE (1994) Mssurviv user's manual, National Biological Service, Patuxent Wildlife Research Center, Laurel, MD 20708–4017.

Lebreton JD, Almeras T, Pradel R (1999) Competing events, mixtures of information and multi-stratum recapture models. Bird Study 46:39–46.

Lebreton J-D, Burnham K, Clobert J, Anderson D (1992) Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. Ecological Monographs 62:67–118.

Lebreton JD, Pradel R (2002) Multistate recapture models: modelling incomplete individual histories. Journal of Applied Statistics 29(1–4):353–369.

MacDonald IL, Zucchini W (2000) Hidden Markov and other models for discrete-valued time series. Monographs on Statistics and applied probability. Chapman and Hall, London.

Nichols JD, Kendall WL, Hines JE, Spendelow JA (2004) Estimation of sex-specific survival from capture-recapture data when sex is not always known. Ecology 85(12):3192–3201.

Pledger S, Pollock KH, Norris J (2003) Open capture–recapture models with heterogenety: I. Cormack–Jolly–Seber model. Biometrics 59:786–794.

Pradel R (2005) Multievent: An extension of multistate capture recapture models to uncertain states. Biometrics 61:442–447.

Pradel R (2008) The stakes of models with state uncertainty. In: Thomson DL, Cooch EG, Conroy MJ (eds.) Modeling Demographic Processes in Marked Populations. Environmental and Ecological Statistics, Springer, New York, 3:781–796.

Pradel R, Johnson AR, Viallefont A, Nager RG, Cezilly F (1997) Local recruitment in the Greater Flamingo: A new approach using capture-mark-recapture data. Ecology 78:1431–1445.

Pradel R, Maurin-Bernier L, Gimenez O, Genovart M, Choquet R, Oro D (2008) Estimation of sex-specific survival with uncertainty in sex assessment. The Canadian Journal of Statistics 36(1):29–42.

Rouan L, Choquet R, Pradel R (2008) A general framework for modeling memory in capture–recapture data. Jabes, In Press.

Rouan L, Gaillard JM, Guédon Y, Pradel R (2008) Estimation of lifetime reproductive success when reproductive status cannot always be assessed. In: Thomson DL, Cooch EG, Conroy MJ (eds.) Modeling Demographic Processes in Marked Populations. Environmental and Ecological Statistics, Springer, New York, 3:867–880.

Véran S, Gimenez O, Flint E, Kendall B, Doherty P, Lebreton JD (2007) Quantifying the impact of longline fisheries on adult survival in the black-footed Albatross. Journal of Applied Ecology 44(5):942–952

White G, Burnham K (1999) Program MARK: Survival estimation from populations of marked animals. Bird Study 46(suppl.):120–139.