



## Capture–recapture models with heterogeneity to study survival senescence in the wild

Guillaume Péron, Pierre-André Crochet, Rémi Choquet, Roger Pradel, Jean-Dominique Lebreton and Olivier Gimenez

G. Péron (*guillaume.peron@cefe.cnrs.fr*), P.-A. Crochet, R. Choquet, R. Pradel, J.-D. Lebreton and O. Gimenez, Centre d'Ecologie Evolutive et Fonctionnelle UMR 5175, CNRS, 1919 Route de Mende, FR-34293 Montpellier Cedex 5, France.

Detecting senescence in wild populations and estimating its strength raise three challenges. First, in the presence of individual heterogeneity in survival probability, the proportion of high-survival individuals increases with age. This increase can mask a senescence-related decrease in survival probability when the probability is estimated at the population level. To accommodate individual heterogeneity we use a mixture model structure (discrete classes of individuals). Second, the study individuals can elude the observers in the field, and their detection rate can be heterogeneous. To account for detectability issues we use capture–mark–recapture (CMR) methodology, mixture models and data that provide information on individuals' detectability. Last, emigration to non-monitored sites can bias survival estimates, because it can occur at the end of the individuals' histories and mimic earlier death. To model emigration we use Markovian transitions to and from an unobservable state. These different model structures are merged together using hidden Markov chain CMR models, or multievent models. Simulation studies illustrate that reliable evidence for survival senescence can be obtained using highly heterogeneous data from non site-faithful individuals. We then design a tailored application for a dataset from a colony of black-headed gull *Chroicocephalus ridibundus*. Survival probabilities do not appear individually variable, but evidence for survival senescence becomes significant only when accounting for other sources of heterogeneity. This result suggests that not accounting for heterogeneity leads to flawed inference and/or that emigration heterogeneity mimics survival heterogeneity and biases senescence estimates.

Senescence, the decline in fitness components with age due to internal physiological deterioration (Medawar 1952), has been evidenced in several life-history traits in a variety of wild vertebrates (Loison et al. 1999, Crespin et al. 2006, Nussey et al. 2006) and is thought to be the rule for most (or all) vertebrate species (Finch 1990, Jones et al. 2008). Yet, several demographic studies of wild populations found no decline in individuals' performance with age (Nichols et al. 1997, Miller 2001, Pistorius and Bester 2002, Congdon et al. 2003), fuelling a debate over the validity of their results.

Among the potential flaws of demographic studies of senescence, the non-modelled effect of individual heterogeneity is recurrent (Vaupel and Yashin 1985, Cam et al. 2002, Zens and Peart 2003, van de Pol and Verhulst 2006). Individual heterogeneity can be defined as the occurrence of systematic variation among individuals in demographic parameters. It can originate from genetic differences or differences in the conditions experienced during development (Fox et al. 2006), variation in individual strategies or quality (e.g. covariation between reproductive effort and survival: Hamel et al. 2008), sex-bias in dispersal or behaviour or the interplay of behavioural differences and study design (e.g. social status- or body condition-dependent detectability: Whitehead and Wimmer 2005, Regehr et al. 2007,

Crespin et al. 2008). Individual heterogeneity can lead to population-level patterns that are not always representative of the actual relationship at the individual level (Vaupel and Yashin 1985). In the case of senescence studies, since the proportion of individuals with high survival probability will tend to increase with age, the age-specific population average of survival probability, which is used in most studies of survival senescence, might not decrease or might even increase with age (Vaupel and Yashin 1985; Fig. 1).

Heterogeneity can be modelled with known individual covariates: Regehr et al. (2007) modelled a gender effect combined with an effect of the observation method; Fox et al. (2006) modelled an effect of the year of birth and family structure. Yet, the precise cause of heterogeneity is often not identified, or not measured. Two ways to accommodate individual variation of unknown origin have been proposed: continuous random effects (Cam et al. 2002, Royle 2008) and mixture models that consider discrete classes of heterogeneity (Pledger et al. 2003, Pradel 2009).

In addition, it is well known that analysing data on marked individuals with models that do not formally estimate detection probabilities together with other demographic parameters (Lebreton et al. 1992) can bias estimations of demographic parameters, including the rate of senescence

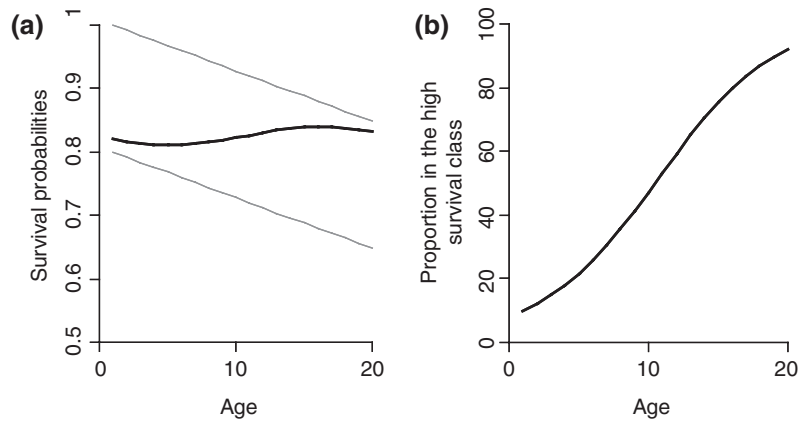


Figure 1. Example heterogeneous population where the average survival rate is not representative of the true ageing processes. This example population consists of 10% of ‘initially robust’ individuals and 90% of ‘initially weak’ individuals, each subpopulation experiencing the same slow decrease in survival with age. (a) age variation in the true survival rate for the two subpopulations (grey lines) and in the average survival rate for the population (black line). (b) age variation in the proportion of robust individuals.

(Gimenez et al. 2008). Individual heterogeneity in detection rate also biases survival estimates downwards if not accounted for (Pradel et al. 1997, Prévot-Julliard et al. 1998) and it violates the fundamental assumption of parameter homogeneity in CMR models (Lebreton et al. 1992), which can lead to flawed inference (Burnham and Anderson 2002).

Last, temporary emigration to non-monitored sites, such that individuals missing during several capture occasions might still be alive elsewhere (Burnham 1993, Fujiwara and Caswell 2002), evidently impacts on the estimation of ‘true’ (as opposed to ‘local’) survival probability. If emigration probability is subject to individual heterogeneity, patterns similar to heterogeneity in survival might appear in the data. Consequently, heterogeneity in temporary emigration can affect the detection of survival senescence as well. In short, we identified three features of population studies (individual heterogeneity, imperfect detectability, temporary emigration out of the study area) which, when not included in the population models, can bias estimates of senescence and/or cause a lack of fit that leads to flawed inference.

Multievent models have been introduced as a unified framework by Pradel (2005). They extend multistate models, in which individuals move between states or die and can at each occasion be detected or not, by considering that the state of an individual is imperfectly determined when it is observed. Their structure rests on the more general framework of hidden Markov chain models (McDonald and Zucchini 1997). The introduction of this model structure in the field of CMR data analysis was initially motivated by capture heterogeneity as reviewed by Pledger (2000). The use of hidden Markov chains in cases not related to individual heterogeneity was to our knowledge initiated by Nichols et al. (2004). These authors modelled a situation where males and females had different survival probabilities, but could not always be separated in the field because of reduced sexual dimorphism. Since then, following the development of the software E-SURGE (Choquet et al. 2009a) numerous applications have been proposed (Pradel 2009). Among others, the implementation of memory models (where demographic parameters depend on the states occupied during the two preceding time steps) is made more straightforward through

the use of hidden states (Rouan et al. 2009), and epidemiology models can be fitted to data where health status determination is uncertain or incomplete (Conn and Cooch 2009).

In this paper we show how this recently developed modelling framework can be used to overcome the types of heterogeneity that typically plague senescence studies. Although these developments are quite general and can be applied to other study situations and taxa, we specifically tailor our example to a study of black-headed gulls *Chroicocephalus ridibundus*. This is a species for which survival senescence is likely to occur (Pugesek et al. 1995, Cam et al. 2002 in related species), but a population in which strong heterogeneities are expected (see Study site and population, in the method section). After presenting the dataset and the features that suggested the need for this new development in CMR models, we present this development and provide simulations that illustrate their performance.

## Methods

### Study site and population

Black-headed gulls *Chroicocephalus ridibundus* are long-lived Charadriiform birds (maximum longevity recorded in our study area is 30 years) and breed colonially, often on vegetated ponds. The data come from a long-term monitoring program of black-headed gulls breeding in La Ronze (noted LR) pond, a large (more than 4000 pairs in recent years) colony located in the Forez basin, at Craintilleux, central France (45°35'N, 4°14'E). In this population, detection is known to vary between individuals because nests are built within vegetation or at its edge and because a large proportion of the re-sightings are made on the nests (Prévot-Julliard et al. 1998). Note that some re-sightings are made on other perches so that all birds are potentially detectable even if their nests are not visible. Additionally, preliminary results in the same population indicated that dispersal rates were individually variable, in particular because of differences between males and females (Grosbois 2001), as is commonplace in birds (Greenwood 1980). Yet,

for both detection and emigration, we only had very partial information on the characteristics of the birds in the field. Detectability cannot be assessed for nests which are not visible of course and can't be evaluated for birds seen on other perches. Furthermore, black-headed gulls are only weakly sexually dimorphic, which precludes sexing of most birds in the field. It was thus clear that accounting for unknown or unmeasured sources of variation between individuals would be very useful if we were to assess survival senescence in this population.

Previous work in the same population indicated that time effects on survival probabilities were reduced or absent (Prévoit-Julliard et al. 1998, Grosbois 2001). We were thus confident that, despite most observations of old individuals occurred at the end of the time series, unaccounted time effects could not confound age effects.

## Data collection

Chicks were ringed before fledging with stainless steel rings. The use of stainless rings is particularly appropriate for the study of senescence since they almost do not wear with age (in all occasions when a known-age adult was physically recaptured at age >14, the code was perfectly readable and the ring could not be removed even with the use of pliers).

Observations of ringed adults were conducted using a floating blind from which metal ring codes could be read with a telescope (Lebreton 1987). At each observation of a ringed adult, we recorded whether it attended a nest or was feeding chicks, which would confirm its status as breeder. Now, an individual nesting in an accessible location had a high probability to be confirmed as breeder, whereas an individual breeding in a less accessible part of the colony was more often observed on roosts or other perches and was more likely to end up with an unconfirmed breeding status. The confirmation of breeding status thus potentially yielded information on the detection probability.

We analyzed the survival of 1556 stainless-ringed adults, for a period of 28 years (from 1978 to 2006), starting from their first re-observation in LR, which occurred from age two to age 23 (mean 5.5, SD 4.1). The large range of age at first resighting is partly due to the fact that black-headed gulls are not present on the colony before starting to reproduce between two and six year-old (Clobert et al. 1994), partly due to the non-exhaustiveness of detection and partly due to the fact that some birds start breeding on other colonies and disperse toward LR after several breeding attempts (Péron et al. unpubl.). The data were coded with one digit per year: '0' (not observed), '1' (confirmed breeder) or '2' (non-confirmed breeder).

## Model description

### *Multievent models for the study of black-headed gull senescence*

As introduced above, our approach was based upon multi-event CMR models (Pradel 2005). The observer records 'events' (here 'not seen', 'confirmed breeder', 'non-confirmed breeder') that carry uncertain information on the state that

the individual occupies at the current sampling occasion (see below and Supplementary material Appendix 1 for state description). The relationship between states and events is thus probabilistic (Pradel 2005).

All models were fully described by first considering the vector of probabilities of initial presence in the various states (II-vector), then linking states at successive sampling occasions by the matrix of survival/transition probabilities ( $\Phi$ -matrix), exactly like in multistate models, while the events were linked to states by the matrix of event probabilities (B-matrix). For convenience we separated  $\Phi$  in two steps (S-matrix for survival probabilities and  $\Psi$ -matrix for dispersal probabilities), and B (P-matrix for detection probabilities and R-matrix for probabilities to confirm breeding status when detected). Full details on the model structure and examples of these matrices are presented in Supplementary material Appendix 1.

### *Individual heterogeneity*

Discrete classes of individuals were built to accommodate heterogeneity, each class being associated with a distinct value of the parameter(s) (Pradel 2009); these classes were the actual states of the multievent model. For a simple example, in a model with a two-class heterogeneity structure for survival probability and no possibility to emigrate (model  $\{\phi(h_i), p(\cdot)\}$  of Pledger et al. 2003), there will be a state 'low survival' and a state 'high survival'. In this model, the probability of the five first events in the first example history is:

$$p(11010) = \pi_1^H s_1^H p_2^H s_2^H (1-p_3)^H s_3^H p_4^H \left[ s_4^H (1-p_5) + (1-s_4^H) \right] + (1-\pi_1^H) s_1^L p_2^L s_2^L (1-p_3)^L s_3^L p_4^L \left[ s_4^L (1-p_5) + (1-s_4^L) \right]$$

Superscripts H and L refer to high and low survival classes respectively; subscripts refer to time-dependence.  $\pi$ ,  $s$ ,  $p$  stand respectively for the probabilities of initial state, survival and detection. The situation is similar to having two possible paths at first observation: one low-survival and one high-survival path (Fig. 2). The 'low survival' path has the greatest probability in 'short' histories like '110100000000' and the lowest in 'long' histories like '1000001000001' (Fig. 2).

In a more complex model, there are a larger number of possibilities when individuals are not encountered. In the presence of several kinds of heterogeneity (survival, emigration and detection), an individual can be in the low or the high-value class for each type of heterogeneity. There is up to eight 'classes of heterogeneity' (Supplementary material Appendix 1 part 4) in the models, which greatly complicates the computation of history probabilities. The need for an algorithm to calculate history probabilities should therefore be apparent. We used program E-SURGE 1.1.1 (Choquet et al. 2009a) to obtain maximum likelihood estimates of the parameters and perform model selection. A more rigorous and general development of the likelihood using matrix notation is presented in Supplementary material Appendix 1.

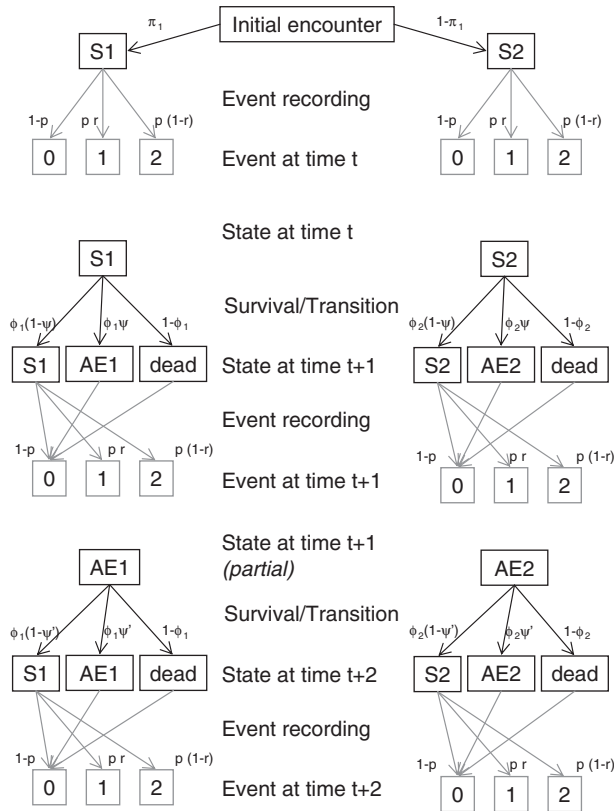


Figure 2. A tree diagram describing the hidden Markov chain probability structure for a model with a two-class individual heterogeneity structure on survival probabilities, the possibility to temporarily emigrate, and a detailed observation structure for the breeding status. Black boxes indicate the five possible states ( $S1$ : alive in the study site with high survival probability;  $S2$ : alive in the study site with low survival probability;  $AE1$ : alive outside the study site with high survival probability;  $AE2$ : alive outside the study site with low survival probability;  $dead$ ), while grey boxes represent the three possible observations following initial release (0: not seen, 1: seen and breeding status confirmed; 2: seen but breeding status not confirmed). The probability for observing a particular encounter history is obtained by summing the probability of all possible paths leading to a given encounter history (for the sake of clarity at time  $t+1$  only the states ‘ $AE$ ’ are represented; paths from states ‘ $S$ ’ are the same as at time  $t$  and the state ‘ $dead$ ’ is absorbing, i.e. there is no path out of it). The probability of a given path can be obtained by multiplying the probabilities appearing alongside its component arrows. These probabilities consist of functions of  $\pi$ , the initial state probabilities;  $\phi$ , apparent survival probabilities;  $\psi$ , state transition probabilities;  $p$ , detection probabilities; and  $r$ , the probabilities to confirm breeding status. A more formal matrix description of the same model is provided in Table 1.

In the following, survival is time-independent and detection is time-dependent, based on results of Prévot-Julliard et al. (1998) and Grosbois (2001).

#### Temporary emigration

We modelled temporary emigration as Markovian or state-dependent transitions to and from a site where detection probability is zero (Fujiwara and Caswell 2002, Schaub et al. 2004, Fig. 2, Supplementary material Appendix 1 part 2).

Colonies that were not searched for marked individuals were grouped in a single non-observable ‘site’ denoted ‘alive elsewhere’ (AE hereafter). This site was included in the usual multisite CMR models formulation (Arnason 1972, 1973, Schwarz et al. 1993) with the only difference that the detection rate was zero. Once in the state AE, individuals lost their classification as low/high detection and low/high emigration. Therefore, upon returning to the study site, they could become more/less detectable, or more/less site-faithful, than what they were before emigrating (see Supplementary material Appendix 1 part 2 for justification).

#### Modelling age-effects on survival when individuals enter the dataset at various ages

The straightforward implementation of age effects in the sense of CMR models (Lebreton et al. 1992) corresponds to the effect of time elapsed since first occurrence in the dataset (hereafter TFC to match a previous acronym: Crespin et al. 2006). In our case gulls were marked as chicks and were thus of known age, but they entered the dataset as adults in the colony at a varying age (Clobert et al. 1994). TFC did thereby not correspond to true age. To model the effect of true age, we had to constrain survival to vary with time across as many groups as there were ages at first occurrence in the dataset (hereafter ‘group approach’; described in details in Supplementary material Appendix 1 part 2). This procedure was computer-time-hungry (around 24 h were needed to fit such a model using an Intel Pentium 4HT, 2.6 GHz ( $3.25 \times 800$ ) processor with 512 Mb of system memory, vs less than 1 h for a TFC model) and was thus impractical for model selection which required running many models sequentially.

Yet using TFC as a proxy for true age in a similar study design does not prevent the detection of survival senescence as shown by Crespin et al. (2006). A test of power (Crespin et al. 2006) indicated that sample size rather than the use of TFC versus true age is the most critical factor preventing the detection of senescence. The main drawback of using TFC is that individuals of various true ages are mixed in a same TFC-class, thus creating noise and increasing the standard error on the estimation of the strength of senescence. Model selection using TFC was therefore considered conservative for what concerns the detection of senescence. We thus used TFC instead of age to select for the best model, and then confirmed our results by running the preferred model with true age instead of TFC.

To represent a biologically sound relationship between age (or TFC) and survival we used a constrained piecewise relationship; we modelled a separate survival probability at age 1, a plateau lasting until 14-years old and a log-linear decrease in survival with age afterwards (see Supplementary material Appendix 1 part 2 for justification).

#### Modelling observations that provide information on detectability

As described in the ‘data collection’ section, the confirmation of breeding status conveyed information on the detectability of individuals. Models accommodating such data structure were introduced by Nichols et al. (2004; the individual

status was the gender, documented by behaviour) and Conn and Cooch (2009; the individual status was the state of health, documented by visible symptoms). In the present paper, the class of detection heterogeneity was documented by the breeding status (see Supplementary material Appendix 1 for practical implementation).

### Model selection

There were 16 models in our candidate set representing every combination of presence/absence of the four considered effects: age-effect on survival, heterogeneity in survival, detection and emigration probabilities. The most general model we considered included heterogeneity in survival, emigration and detection probabilities as well as age-effect and had 11 states (model denoted  $\{S_{H,+a}; D_H; E_H\}$ ; Supplementary material Appendix 1 part 4). Subscripts H and 0 referred to models with and without heterogeneity in survival (denoted S), detection (denoted D) and emigration (denoted E) probabilities, while subscripts +a and +ā referred to models with and without age-effect on survival. The lowest AIC-model (with a two AIC-points difference) was preferred (Akaike's information criterion; Burnham and Anderson 2002). However, all models in which a given effect was included could be considered as 'redundant'. Therefore, the importance value of each effect was computed as the sum of the AIC-weights of the models including the considered effect, and it was interpreted as the probability that the effect was present in the data (Burnham and Anderson 2002).

### Goodness-of-fit

Goodness-of-fit (GOF) tests were performed on raw capture/non capture histories (formed of '0' and '1' thus discarding the details about age at first detection and breeding status; note that all the data concerned a single site). One of the main consequences of heterogeneity in detection probability is an excess (when compared to homogeneous datasets) of encounter histories with consecutive 'captures' (i.e. runs of '1') and consecutive 'non-captures' (i.e. runs of '0'). Such histories indicate the presence of highly and poorly detectable individuals, respectively. Some of the runs of '0' will occur at the end of the capture history. As a consequence detection heterogeneity tends to induce both 'transience' (i.e. lower chance of recapture of first-encountered individuals than already encountered ones; Pradel et al. 1997) and 'trap-happiness' (i.e. higher probability to encounter at time  $t+1$  the individuals encountered at time  $t$  than the individuals not encountered at time  $t$  but known to be alive because of previous and future recaptures; Pradel 1993).

One-sided directional test statistics are the signed square roots of the  $\chi^2$ -statistics for the corresponding tests: Test3.SR for transience and Test2.CT for trap-dependence (Pradel 1993, Pradel et al. 1997, 2005; practical implementation detail in Choquet et al. 2005). They are the most relevant statistics for the detection of transience and trap-happiness

respectively (Pradel et al. 2005) and are inflated by individual heterogeneity.

We used techniques for partitioning  $\chi^2$ -variables (Rao 1973, pp. 185 and following) to approximate a GOF test for a time-dependent model with capture heterogeneity as follow: (1) we computed the overall GOF  $\chi^2$ -statistics for transience and trap-dependence, from Test3 and Test2 respectively (Pradel 1993, Pradel et al. 1997, Choquet et al. 2005) (2) we computed the directional statistics, from Test3.SR and Test2.CT respectively (3) we removed from the overall statistics the corresponding squared directional statistics (which are asymptotically distributed as  $\chi^2_1$ ), and we obtained non-directional components with one degree of freedom less. These components corresponded to a GOF test for a model where sources of transience and trap-happiness (here, individual heterogeneity) were accounted for. If this corrected test still proved statistically significant, we used an overdispersion coefficient  $\hat{c}$ , computed as the ratio between the  $\chi^2$ -statistic and the degree of freedom, in the model selection procedure (Burnham and Anderson 2002). All the GOF test components were computed using U-CARE (Choquet et al. 2009b).

### Simulation study

To judge whether the results provided by the proposed methodology were reliable in a complex but known case (three kinds of heterogeneity and a decrease in survival probability with age), we carried out Monte Carlo simulation studies (Supplementary material Appendix 2). Very briefly, these exercises illustrated that the multievent framework allowed (1) detecting simultaneously and using AIC all three kinds of heterogeneities when present, and (2) obtaining reliable evidence and precise estimates for survival senescence by accounting for these heterogeneities.

More precisely the simulation study indicated (1) that the magnitude of the age-effect on survival was very precisely retrieved when the heterogeneity structure in the model exactly matched the simulated structure, and (2) that bringing additional information on detectability of individuals was sufficient (and necessary) to separate detectability and emigration heterogeneities. In our case, such information was provided by the confirmation of breeding status. The main drawbacks of the method were the non-reliability of emigration probability estimates (the presence of emigration heterogeneity was retrieved, but the actual values of the parameters were not), and the fact that, most probably because the data on them were sparse, estimates for parameters associated to low-survival individuals were often inaccurate.

## Results

### Black-headed gull dataset: goodness-of-fit

As expected, the directional tests for transience and trap-happiness were statistically significant, which is a cue for heterogeneity in detection (Table 1). The corrected Test3.

Table 1. Components of an approximate goodness-of-fit (GOF) test for a model with heterogeneity, obtained by removing from the components of the standard GOF test for the time-dependent model the squared directional test statistics (see methods). The overall GOF test shows no sign of lack-of-fit for a model correcting transience and trap-happiness. DF is the degree of freedom.  $\hat{c}$  is the overdispersion coefficient computed as the ratio between the chi-squared statistic and the degree of freedom.

	Test3: transience		Test2: trap dependence		Total $\chi^2$ - statistic
	$\chi^2$ -statistic	Squared directional statistic	$\chi^2$ - statistic	Squared directional statistic	
Time-dependent model	64.19	38.67	25.88	9.21	164.39
DF	22	1	22	1	131
p-level	< 0.0001	< 0.0001	0.26	0.002	0.02
$\hat{c}$	2.92		1.18		1.25
Time-dependent model with heterogeneity of detection	25.52		16.67		116.52
DF	21		21		129
p-level	0.23		0.73		0.78
$\hat{c}$	1.22		0.79		0.90

SR and Test2.CT were statistically non significant, and the overall corrected GOF test indicated that there was no need to account for any overdispersion (Table 1). This test indicated that a time-dependent model with heterogeneity in detection probability fitted the data. Thus, accounting for other sources of heterogeneity could only improve the fit.

### Modelling age-dependence and testing for senescence

#### Model selection using TFC

There was strong support for a model with individual heterogeneity in both detection and emigration probabilities, along with TFC-effect on survival (Table 2: model  $\{S_{0+a};D_H;E_H\}$ ). This model was nearly four AIC-points lower and three times more likely than the next model (as indicated by the ratio of AIC-weights). The importance value (computed as the sum of

the AIC-weights of the models in which the considered effects occurred) of the TFC-effect on survival probability was 0.83, which we interpret as a high probability for a decrease in survival with age. These results supported the existence of survival senescence in the population, although the slope of the TFC-effect was statistically not different from zero:  $-0.16$  (95% CI:  $-0.49$ ;  $0.17$ ); see next section.

Importance values of heterogeneity in survival, detection and emigration were 0.09, 0.94, and 0.78, respectively. We interpret these values as high probability that two-class heterogeneity structure was present in detection and emigration probabilities, and low probability that such heterogeneity was present in survival probability. As discussed later, we do not exclude that the discrete-class heterogeneity models were unable to detect small, continuous individual variation in survival probability.

Table 2. Model selection. The 16 candidate models vary in the presence/absence (Y/N) of heterogeneity and of the age-effect on survival. For each model the number of parameters (np), deviance (Dev), AIC and AIC-weight are given. Subscripts H and 0 referred to models with and without heterogeneity in survival (denoted S), detection (denoted D) and emigration (denoted E) probabilities, while subscripts a and  $\bar{a}$  referred to models with and without age-effect on survival. The models are sorted by AIC.

Model	Heterogeneity in:			Age-effect	np	Dev	AIC	AIC weight
	Survival	Detection	Emigration					
$\{S_{0+a};D_H;E_H\}$	N	Y	Y	Y	43	7781.88	7867.88	0.761
$\{S_{0+\bar{a}};D_H;E_0\}$	N	Y	N	N	36	7799.44	7871.44	0.128
$\{S_{H+a};D_H;E_0\}$	Y	Y	N	Y	47	7779.50	7873.50	0.046
$\{S_{H+\bar{a}};D_0;E_0\}$	Y	N	N	N	35	7803.62	7873.62	0.043
$\{S_{0+a};D_0;E_H\}$	N	N	Y	Y	37	7802.12	7876.12	0.012
$\{S_{H+a};D_0;E_0\}$	Y	N	N	Y	38	7801.99	7877.99	0.005
$\{S_{0+\bar{a}};D_H;E_H\}$	N	Y	Y	N	40	7800.28	7880.28	0.002
$\{S_{0+a};D_H;E_0\}$	N	Y	N	Y	41	7797.91	7879.91	0.002
$\{S_{H+a};D_0;E_H\}$	Y	N	Y	Y	42	7797.25	7881.25	0.001
$\{S_{H+\bar{a}};D_H;E_H\}$	Y	Y	Y	Y	53	7777.12	7883.12	0.000
$\{S_{H+\bar{a}};D_H;E_0\}$	Y	Y	N	N	43	7802.21	7888.21	0.000
$\{S_{0+\bar{a}};D_0;E_0\}$	N	N	N	N	31	7856.10	7918.10	0.000
$\{S_{0+a};D_0;E_0\}$	N	N	N	Y	34	7820.42	7888.42	0.000
$\{S_{H+\bar{a}};D_0;E_H\}$	Y	N	Y	N	37	7852.13	7926.13	0.000
$\{S_{H+\bar{a}};D_H;E_H\}$	Y	Y	Y	N	49	7834.15	7932.15	0.000
$\{S_{0+\bar{a}};D_0;E_H\}$	N	N	Y	N	34	7857.33	7925.33	0.000

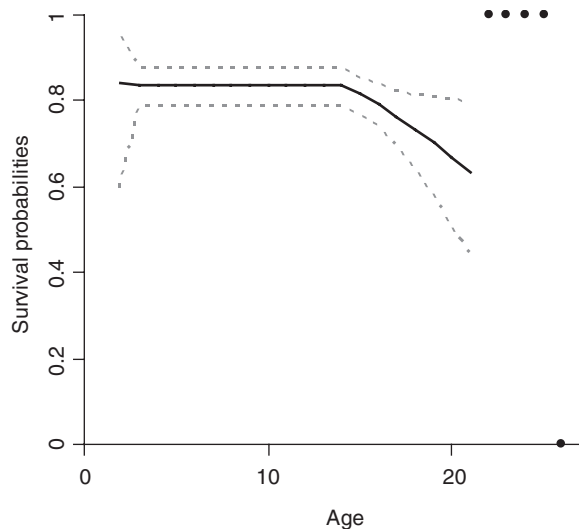


Figure 3. Survival probabilities for the black-headed gulls as a function of true age. Parameters estimates are from model  $\{S_{0+a}; D_H; E_H\}$  which includes heterogeneity in both detection and emigration probabilities, and a piecewise constrained relationship between age and survival probabilities. Dotted lines correspond to the 95% asymptotic CI. Black dots are boundary estimates, which come with no standard errors. The survival probability for the first age class (age 2) was estimated separately and fell on the line for older age classes.

The second best model (Table 2: model  $\{S_{0+a}; D_H; E_0\}$ ), as opposed to the best model, did not account for heterogeneity in emigration probability and, most importantly, did not include any variation in survival probability with TFC. This result means that, when not accounting for heterogeneity in emigration probabilities, the selection procedure discarded TFC-effect on survival, in other words the detection of survival senescence was prevented.

#### Parameter estimates in the true age formulation

Parameter estimates are from the preferred model  $\{S_{0+a}; D_H; E_H\}$ , ran using true age instead of TFC. Prime age survival (between 2 and 14 years old) was 0.84 (95% CI: 0.79; 0.88) and the slope of the decrease in survival after age 14 was  $-0.16$  ( $-0.30$ ;  $-0.02$ ) on a logit scale (Fig. 3). Thus, when comparing true age- (this model) and TFC-models (previous section), we observed that the use of TFC increased the error on the estimated slope of the decrease in survival with age, as expected, but did not modify the value of the estimate.

The estimated temporary emigration probabilities to state AE were 0.17 (0.03; 0.54) in stayers and 0.67 (0.28; 0.91) in movers. The estimated detection probabilities (averaged over time) were 0.08 (0.05; 0.10) and 0.48 (0.41; 0.55) in the low- and high-detectability classes. The estimated proportion of sightings on the nest or with chicks was 0.53 (0.44; 0.63) for the low-detectability class and 0.61 (0.56; 0.66) for the high-detectability class.

## Discussion

### Detecting senescence in wild populations

We used multievent framework to combine three pre-existing types of CMR models that were potentially required

to fit the gull data: heterogeneity models with discrete classes (Pledger et al. 2003, Pradel 2009), temporary emigration models (Fujiwara and Caswell 2002, Schaub et al. 2004), and models accommodating partial information on individual status (Nichols et al. 2004, Conn and Cooch 2009). The framework was flexible enough to combine these model structures which had only been used separately until now.

Survival senescence in our population was only detected after accounting for heterogeneity in temporary emigration in the models (Table 2), which illustrates the interest of complex models mixing several possible causes of heterogeneity. This result was perhaps not unexpected since temporary emigration events occurring at the end of an individual's life, when not followed by a detection event, could mimic earlier death. Thereby, emigration heterogeneity might have created patterns in the data similar to heterogeneity in survival, and prevented the detection of survival senescence when not accounted for. However, our simulation studies only weakly supported this purported effect of non modelled heterogeneity in temporary emigration on the detection of survival senescence. The result that senescence was discarded when not accounting for emigration heterogeneity might therefore originate from a lack-of-fit impairing the model selection when models did not include the appropriate heterogeneity structure.

Definitive emigration has evidently an even greater impact on survival estimates than temporary emigration, but it can only be modelled if recoveries data (i.e. rings recovered on birds shot or found dead throughout the year) are available (Burnham 1993).

In senescence studies, we thus recommend that (1) emigration is modelled when field observations indicate its presence and (2) either individual heterogeneity is explicitly considered or evidence for its absence is provided by GOF tests or biological considerations. In particular, empirical support for the absence of senescence in animals (Nichols et al. 1997, Miller 2001, Pistorius and Bester 2002, Congdon et al. 2003) should be considered with caution until the results are verified with analyses accounting for heterogeneity.

### Sources of heterogeneity in the gull dataset: emigration, detectability

Although there was a strong support for the existence of low- and high- emigration classes in our population, biological explanations are not straightforward. Emigration heterogeneity might include the skipping of breeding attempts and the effect of early nest failures. These two phenomena result in an absence of the individuals from the colonies during field-work (thereby mimicking temporary emigration), and they are influenced by individual quality (Calladine and Harris 1997), a well known source of heterogeneity in demographic parameters (Hamel et al. 2008). Moreover, true temporary emigration, reproduction failures and reproduction skipping do not occur at the same frequency which might create individual heterogeneity when they are modelled using a same transition probability. Alternatively, individual heterogeneity in true temporary emigration rate could stem from sex-biased dispersal (Greenwood 1980) or from heterogeneity in individual quality and/or the conditions experienced during early life or previous

reproduction attempts, both of which are known to affect habitat choice and dispersal behaviours (Switzer 1997, Clobert et al. 2009). Last, and perhaps more speculatively, heterogeneity in dispersal behaviour can result from genetically determined differences in ‘personalities’ or other behavioural syndromes (Cote and Clobert 2007, Clobert et al. 2009). Tradeoffs between investment in exploration-dispersal and in other traits (Wolf et al. 2007) or frequency-dependent selection acting through environmental stochasticity (Dingemanse et al. 2004) can maintain stable polymorphism in dispersal tendencies.

Heterogeneity in detection probability was probably related to the fact that high vegetation density hindered the detection of birds breeding far inside the vegetation. These are likely to have strongly contributed to the class with a low detection probability and a low proportion of sightings on a nest. Such heterogeneity in detection induced by habitat heterogeneity in the study site is supposedly common, and we recommend it be explicitly incorporated.

### Individual heterogeneity of unknown origin: modelling approach

As in most modelling exercises, our study relied on some untested assumptions. Modelling individual heterogeneity using discrete classes was a priori appropriate in our study: birds could breed inside vegetation versus outside vegetation, their movements could occur within versus outside of the colony. However, we do not claim that the discrete-class model represented individual variation in survival probabilities better than a continuous individual random effect could (Royle 2008, but see Pledger 2005). Concerning the number of heterogeneity classes, Pledger (2005) indicated that, as a theoretic and approximated representation of individual heterogeneity, the two-class models were more parsimonious than models with more classes, excepted in the presence of strong multimodality in the true distribution of the parameters of interest. Standard quantitative tools such as AIC generally fail to separate different forms of heterogeneity models (data not shown; Pledger 2005). We suggest that information on the study system from the field might be the most reliable cue when deciding which form of individual heterogeneity to incorporate in CMR models.

### Performance and identifiability issues

The simulation studies illustrated that reliable evidence for survival senescence could be obtained in the simultaneous presence of three kinds of heterogeneity. When the data included events whose frequency depended on detectability, the three sources of heterogeneity were separately identifiable, although in most cases only the structure and not the actual values of emigration probabilities were retrieved. The magnitude of the age-effect was very precisely retrieved when the heterogeneity structure in the model exactly matched the way data was simulated. In the gull application we did not detect any parameter redundancy issue using the rank of the models’ numeric derivative matrix (Rouan et al. 2009, Appendix A, Choquet et al. 2009a, pp. 56–57). Overall

our results are therefore encouraging for the application of multievent models to the study of population dynamics of species with complex life-histories, weak or variable site-fidelity, or inhabiting very heterogeneous habitats.

*Acknowledgements* – This research was partly supported by a grant from the ‘Jeunes Chercheuses at Jeunes Chercheurs’ program of the French ANR (ANR-08-JCJC-0028-01).

## References

- Arnason, A. N. 1972. Parameter estimates from mark–recapture experiments on two populations subject to migration and death. – *Res. Popul. Ecol.* 13: 97–113.
- Arnason, A. N. 1973. The estimation of population size, migration rates and survival in a stratified population. – *Res. Popul. Ecol.* 15: 1–8.
- Burnham, K. P. 1993. A theory for combined analysis of ring recovery and recapture data. – In: Lebreton, J. D. and North, P. H. (eds), *Marked individuals in the study of bird population*. Birkhäuser, pp. 199–213.
- Burnham, K. P. and Anderson, D. R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.) – Springer.
- Calladine, J. and Harris, M. P. 1997. Intermittent breeding in the herring gull *Larus argentatus* and the lesser black-backed gull *Larus fuscus*. – *Ibis* 139: 259–263.
- Cam, E. et al. 2002. Individual covariation in life–history traits: seeing the trees despite the forest. – *Am. Nat.* 159: 96–105.
- Choquet, R. et al. 2005. U-CARE 2.2 user’s manual – CEFE, Montpellier, France.
- Choquet, R. et al. 2009a. Program E-SURGE: a software application for fitting multievent models. – In: Thomson, D. L. et al. (eds), *Modelling demographic processes in marked populations*. Springer, pp. 845–865.
- Choquet, R. et al. 2009b. U-CARE: Utilities for performing goodness of fit tests and manipulating capture–recapture data. – *Ecography* 32: 1071–1074.
- Clobert, J. et al. 1994. The estimation of age–specific breeding probabilities from recaptures or resightings in vertebrate populations. *Longitudinal models*. – *Biometrics* 50: 375–87.
- Clobert, J. et al. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. – *Ecol. Lett.* 12: 197–209.
- Congdon, J. D. et al. 2003. Testing hypotheses of aging in long-lived painted turtles (*Chrysemys picta*). – *Exp. Gerontol.* 38: 765–72.
- Conn, P. B. and Cooch, E. G. 2009. Multistate capture–recapture under imperfect state observation: an application to disease models. – *J. Appl. Ecol.* 46: 486–492.
- Cote, J. and Clobert, J. 2007. Social personalities influence natal dispersal in a lizard. – *Proc. R. Soc. Lond. B* 274: 383–390.
- Crespin, L. et al. 2006. Increased adult mortality and reduced breeding success with age in a population of common guillemot *Uria aalge* using marked birds of unknown age. – *J. Avian Biol.* 37: 273–282.
- Crespin, L. et al. 2008. Is heterogeneity of catchability in capture–recapture studies a mere sampling artefact or a biologically relevant feature of the population? – *Popul. Ecol.* 50: 247–256.
- Dingemanse, N. J. et al. 2004. Fitness consequences of avian personalities in a fluctuating environment. – *Proc. R. Soc. Lond. B* 271: 847–852.
- Finch, C. E. 1990. *Longevity, senescence, and the genome*. – Univ. of Chicago Press.



- Fox, G. A. et al. 2006. Consequences of heterogeneity in survival probability in a population of Florida scrub-jays. – *J. Anim. Ecol.* 75: 921–927.
- Fujiwara, M. and Caswell, H. 2002. A general approach to temporary emigration in mark–recapture analysis. – *Ecology* 83: 3266–3275.
- Gimenez, O. et al. 2008. The risk of flawed inference in evolutionary studies when detectability is less than one. – *Am. Nat.* 172: 441–448.
- Greenwood, P. J. 1980. Mating systems, philopatry and dispersal in birds and mammals. – *Anim. Behav.* 28: 1140–1162.
- Grosbois, V. 2001. La dispersion : trait d'histoire de vie et paramètre démographique : étude empirique dans une population de mouette rieuse. PhD thesis – Univ. Montpellier II, Montpellier, France, in French.
- Hamel, S. et al. 2008. Individual variation in reproductive costs of reproduction: high-quality females always do better. – *J. Anim. Ecol.* 78: 143–151.
- Jones, O. R. et al. 2008. Senescence rates are determined by ranking on the fast-slow life-history continuum. – *Ecol. Lett.* 11: 664–673.
- Lebreton, J. D. 1987. Régulation par le recrutement chez la Mouette rieuse *Larus ridibundus*. – *Rev. Ecol. - Terre Vie Suppl.* 4: 173–187, in French.
- Lebreton, J. D. et al. 1992. Modelling survival and testing biological hypotheses using marked animals - a unified approach with case-studies. – *Ecol. Monogr.* 62: 67–118.
- Loison, A. et al. 1999. Age-specific survival in five populations of ungulates: evidence of senescence. – *Ecology* 80: 2539–2554.
- MacDonald, I. L. and Zucchini, W. 1997. Hidden Markov and other models for discrete-valued time series. – Chapman and Hall.
- Medawar, P. B. 1952. An unsolved problem of biology. – Lewis, London.
- Miller, J. K. 2001. Escaping senescence: demographic data from the three-toed box turtle (*Terrapene carolina triunguis*). – *Exp. Gerontol.* 36: 829–832.
- Nichols, J. D. et al. 1997. Test for senescent decline in annual survival probabilities of common pochard, *Aythya ferina*. – *Ecology* 78: 1009–1018.
- Nichols, J. D. et al. 2004. Estimation of sex-specific survival from capture–recapture data when sex is not always known. – *Ecology* 85: 3192–3201.
- Nussey, D. H. et al. 2006. The rate of senescence in maternal performance increases with early-life fecundity in red deer. – *Ecol. Lett.* 9: 1342–1350.
- Pistorius, P. A. and Bester, M. N. 2002. A longitudinal study of senescence in a pinniped. – *Can. J. Zool.* 80: 395–401.
- Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. – *Biometrics* 56: 434–442.
- Pledger, S. 2005. The performance of mixture models in heterogeneous closed population capture–recapture. – *Biometrics* 61: 868–873.
- Pledger, S. et al. 2003. Open capture–recapture models with heterogeneity: I. Cormack–Jolly–Seber model. – *Biometrics* 59: 786–794.
- Pradel, R. 1993. Flexibility in survival analysis from recapture data: handling trap-dependence. – In: Lebreton, J. D. and North, P. H. (eds), *Marked individuals in the study of bird population*. Birkhäuser, pp. 29–37.
- Pradel, R. 2005. Multievent: an extension of multistate capture–recapture models to uncertain states. – *Biometrics* 61: 442–447.
- Pradel, R. et al. 1997. Capture–recapture survival models taking account of transients. – *Biometrics* 53: 60–72.
- Pradel, R. et al. 2005. Principles and interests of GOF tests for multistate capture–recapture models. – *Anim. Biodiv. Conserv.* 28: 189–204.
- Pradel, R. 2009. The stakes of capture–recapture models with state uncertainty. – In: Thomson, D. L. et al. (eds), *Modeling demographic processes in marked populations*. Springer, pp. 781–795.
- Pradel, R. et al. 2007. Estimation of sex-specific survival with uncertainty in sex assessment. – *Can. J. Stat.* 36: 29–42.
- Prévot-Julliard, A. C. et al. 1998. Re-evaluation of adult survival of black-headed gulls (*Larus ridibundus*) in presence of recapture heterogeneity. – *Auk* 115: 85–95.
- Pugesek, B. H. et al. 1995. Mark-resighting analysis of a California gull population. – *J. Appl. Stat.* 22: 625–39.
- Rao, C. R. 1973. Linear statistical inference and its applications. – Wiley.
- Regehr, E. V. et al. 2007. Effects of earlier sea ice breakup on survival and population size of polar bear in western Hudson Bay. – *J. Wildlife Manage.* 71: 2673–2683.
- Rouan, L. et al. 2009. A general framework for modeling memory in capture–recapture data. – *J. Agric. Biol. Environ.*, in press.
- Royle, J. A. 2008. Modeling individual effects in the Cormack–Jolly–Seber model: a state-space formulation. – *Biometrics* 64: 364–370.
- Schaub, M. et al. 2004. Estimating survival and temporary emigration in the multistate capture–recapture framework. – *Ecology* 85: 2107–2113.
- Schwarz, C. J. et al. 1993. Estimating migration rates using tag-recovery data. – *Biometrics* 49: 177–193.
- Switzer, P. V. 1997. Past reproductive success affects future habitat selection. – *Behav. Ecol. Sociobiol.* 40: 307–312.
- van de Pol, M. and Verhulst, S. 2006. Age-dependant traits: a new statistical model to separate within- and between-individual traits. – *Am. Nat.* 167: 766–773.
- Vaupel, J. W. and Yashin, A. I. 1985. Heterogeneity's ruses: some surprising effects of selection on population dynamics. – *Am. Stat.* 39: 176–185.
- Whitehead, H. and Wimmer, T. 2005. Heterogeneity and the mark–recapture assessment of the Scotian Shelf population of northern bottlenose whales (*Hyperoodon ampullatus*). – *Can. J. Fish. Aquat. Sci.* 62: 2573–2585.
- Wolf, M. et al. 2007. Life-history tradeoffs favour the evolution of animal personalities. – *Nature* 447: 581–584.
- Zens, M. S. and Peart, D. R. 2003. Dealing with death data: individual hazards, mortality and bias. – *Trends Ecol. Evol.* 18: 366–373.

Supplementary material (available online as Appendix O17882 at <www.oikos.ekol.lu.se/appendix>). Appendix 1: Model description. Appendix 2: Simulation study.