

## Looking for a needle in a haystack: inference about individual fitness components in a heterogeneous population

**Emmanuelle Cam, Olivier Gimenez, Russell Alpizar-Jara, Lise M. Aubry, Matthieu Authier, Evan G. Cooch, David N. Koons, William A. Link, Jean-Yves Monnat, James D. Nichols, Jay J. Rotella, Jeffrey A. Royle and Roger Pradel**

*E. Cam (emmanuelle.cam@univ-tlse3.fr), O. Gimenez, M. Authier and R. Pradel, Centre d'Écologie Évolutive et Fonctionnelle, UMR 5175 1919 Route de Mende, FR-34293 Montpellier Cedex 5, France. EC also at: Évolution et Diversité Biologique, UMR 5174, Univ. Paul Sabatier - Bâtiment IVR1 pièce 254, 118 route de Narbonne, FR-31 062 Toulouse cedex 9, France. MA also at: Centre d'Études Biologiques de Chizé, UPR 1934 CNRS, FR-79360 Villiers en Bois, France. – R. Alpizar-Jara, Centro de Investigação em Matemática e Aplicações (CIMA-U.E.), Dept of Mathematics, Univ. of Évora, Rua Romão Ramalho, no. 59, PT-7000-Évora, Portugal. – L. M. Aubry and D. N. Koons, Dept of Wildland Resources, 5230 Old Main Hill, Utah State University, Logan, UT 84322-5230, USA. – E. G. Cooch, Dept of Natural Resources, Fernow Hall, Cornell Univ., Ithaca, NY 14853, USA. – W. A. Link, J. D. Nichols and J. A. Royle, USGS Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, MD 20708-4017, USA. – J.-Y. Monnat, Penn Ar Run Izella, FR-29770 Goulien, France. – J. J. Rotella, Dept of Ecology, Montana State Univ., PO Box 173460, Bozeman, MT 59717-3460, USA.*

Studies of wild vertebrates have provided evidence of substantial differences in lifetime reproduction among individuals and the sequences of life history 'states' during life (breeding, nonbreeding, etc.). Such differences may reflect 'fixed' differences in fitness components among individuals determined before, or at the onset of reproductive life. Many retrospective life history studies have translated this idea by assuming a 'latent' unobserved heterogeneity resulting in a fixed hierarchy among individuals in fitness components. Alternatively, fixed differences among individuals are not necessarily needed to account for observed levels of individual heterogeneity in life histories. Individuals with identical fitness traits may stochastically experience different outcomes for breeding and survival through life that lead to a diversity of 'state' sequences with some individuals living longer and being more productive than others, by chance alone. The question is whether individuals differ in their underlying fitness components in ways that cannot be explained by observable 'states' such as age, previous breeding success, etc. Here, we compare statistical models that represent these opposing hypotheses, and mixtures of them, using data from kittiwakes. We constructed models that accounted for observed covariates, individual random effects (unobserved heterogeneity), first-order Markovian transitions between observed states, or combinations of these features. We show that individual sequences of states are better accounted for by models incorporating unobserved heterogeneity than by models including first-order Markov processes alone, or a combination of both. If we had not considered individual heterogeneity, models including Markovian transitions would have been the best performing ones. We also show that inference about age-related changes in fitness components is sensitive to incorporation of underlying individual heterogeneity in models. Our approach provides insight into the sources of individual heterogeneity in life histories, and can be applied to other data sets to examine the ubiquity of our results across the tree of life.

When focusing on natural selection at the level of individual organisms, fitness in its most general sense is success in contributing descendants to the next generation (Fairbairn and Reeve 2001). Under some assumptions (Brommer et al. 2002), realized lifetime fitness can be assessed using lifetime reproduction (i.e. total number of offspring, or of recruited offspring, produced during life, Grafen 1988). Reviews of lifetime reproduction in wild vertebrates have provided evidence of substantial differences among individuals, with a small proportion of individuals producing many offspring, and a large proportion of individuals raising only a few or none (Clutton-Brock 1988, Newton

1989). This pattern is apparently consistent with the hypothesis of heterogeneity among individuals in their 'underlying' ability to produce a large number of offspring. Under this hypothesis, differences in total reproductive output among individuals result from differences in individual fitness components: a hierarchy that is determined before, or at the onset of reproductive life, that doesn't change afterwards, and that translates into a higher proportion of unsuccessful breeding events during the life of some individuals than others, or into reproductive lifespans of unequal length. Historically, the authors who first invoked this hypothesis did not necessarily attempt to explain such

differences (Curio 1983). Echoing this, subsequent efforts to estimate individual fitness components have relied on the hypothesis of ‘underlying’ or ‘latent’ unmeasured individual features (Link et al. 2002a). Recently, the relevance of underlying individual differences to variation among individual life histories and skewness in lifetime reproduction has been questioned (Tuljapurkar et al. 2009, Steiner et al. 2010, Orzack et al. 2011, Steiner and Tuljapurkar 2012).

The debate focuses on the sources of variation in fitness components (‘vital rates’) that are sufficient to generate observed distributions of life histories in wild animal populations (Caswell 2011), i.e. the observed diversity of sequences of life history states (e.g. unsuccessful breeding in the first breeding attempt, nonbreeding in the following year, successful breeding two years later, ..., death). In vertebrates, lifetime reproduction is classically viewed as depending on the age-specific probabilities that an individual survives, that survivors attempt to breed, that breeders raise offspring to independence (probability of breeding successfully), and that successful breeders raise a given number of offspring. Tuljapurkar et al. (2009) and Steiner et al. (2010) suggested that only some sources of variation in vital rates are responsible for the diversity of life histories and the observed skewness of lifetime reproduction across individuals. They suggested defining strata in the population within which individuals are assumed to be homogeneous with respect to fitness components. Strata can correspond to fixed individual features or groups (e.g. sex, Steiner et al. 2010, Orzack et al. 2011), features that change deterministically throughout life (e.g. with age), or stochastically such as life history states (Cam et al. 1998). The crux of the matter is to take Markovian transitions among life history states into account (Tuljapurkar et al. 2009, Steiner et al. 2010), a common approach in longitudinal studies of wild vertebrates (Lebreton et al. 2009). Biologically, first-order Markov models represent the hypothesis that the probability of an individual being in state  $s$  (e.g. alive or dead, breeding or not in survivors) at time  $t + 1$  depends only on state  $r$  at time  $t$  (e.g. whether the individual bred in the previous year, or not, Cam et al. 1998), not on genetic or ontogenetic differences that individuals carry with them over life. Even if individuals within each stratum are homogeneous with respect to the vital rates (e.g. all the females of age  $x$  in year  $t$  that bred successfully in the previous year will survive with probability 0.90), life history variation will result from stochastic sampling associated with the demographic process itself (this female will survive, but not that one, and the actual number of survivors within a stratum will follow the binomial distribution). Stochasticity thus generates heterogeneity, or differences among life paths and lifetime reproduction, such that some individuals are simply lucky enough to have substantially longer and more productive lives than others. The diversity of life paths consistent with first-order Markov processes has been called ‘dynamic heterogeneity’, which could be used as a ‘neutral model’ for generating variation in the absence of fixed underlying differences in fitness components (Tuljapurkar et al. 2009, Steiner et al. 2010, Steiner and Tuljapurkar 2012).

These ideas contrast with the hypothesis that individuals differ in their underlying fitness components in ways

that go beyond direct observation of current or previous life history state, cumulative breeding success, age, year and group membership such as sex (Cam et al. 2002, Link and Barker 2009, Bergeron et al. 2010, Hawkes 2010, Aubry et al. 2011). In fact, the current debate has two aspects. 1) Absence of individual variation with dynamic states during an individual’s life contrasts with the hypothesis of ‘fixed’ differences among individuals (Cam et al. 2002, Link and Barker 2009, Bergeron et al. 2010, Hawkes 2010, Aubry et al. 2011). However, the debate is more subtle than just ‘fixed’ versus ‘dynamic’ heterogeneity. Indeed, Steiner et al. (2010) and Orzack et al. (2011) acknowledged the relevance of some fixed differences among individuals, like sex-specific differences in survival for example, or differences according to age of first breeding. 2) The debate also involves observable vs unobservable sources of variation in life history traits among individuals. The dynamics of ‘observable states’ contrasts with the ‘underlying’, ‘latent’ nature of differences among individuals invoked in several papers (Cam et al. 2002, Link et al. 2002a, Hawkes 2010, Aubry et al. 2011). That is, the debate focuses on the relative contribution of observable state dynamics and ‘fixed latent’ biological heterogeneity among individuals to the diversity of observed life histories.

In reality, both ‘fixed’ underlying differences among individual fitness components and the stochastic outcome of state-specific probabilities governing state transition, survival and reproductive success are expected to generate a diversity of life histories in populations, but the underlying processes are clearly different, and the degree to which they contribute to life history diversity may differ as well (Tuljapurkar et al. 2009, Caswell 2011, Steiner and Tuljapurkar 2012). Life history states such as breeding activity and success can be directly observed by investigators, and the way vital rates vary according to state or covariates (e.g. sex) can be estimated without assuming unobservable differences among individuals, but ‘latent variation’ must necessarily be inferred using a statistical model (Link et al. 2002a, Link and Barker 2009). In some situations investigators temporarily fail to contact animals during sampling sessions, and true state cannot be directly observed. This observation error generates missing data and can be handled via capture–mark–recapture models incorporating the probability of detecting individuals (Lebreton et al. 2009), but must be distinguished from ‘underlying’ unobserved variation among individuals in vital rates (Aalen et al. 2008a).

The hypothesis of underlying differences among individuals is relevant to situations where all sources of variation in vital rates have not been identified. Although not a given, this assumption is commonly made in life history research using long-term data (Royle 2008, van de Pol and Wright 2009, Hawkes 2010, Gimenez and Choquet 2010, Jones et al. 2010). Permanent underlying differences among individuals may arise from genetic differences, phenotypic plasticity across gradients of ontogenetic developmental conditions that have long-lasting effects on survival or reproduction as adults, or both (Cam et al. 2003). Moreover, consistent differences among habitats selected by individuals over life may also contribute to individual variation in demographic performance (Kendall et al. 2011).

In fields that commonly use time-to-event or failure-time models (human demography, epidemiology), unobserved variation across individuals is addressed by estimating a ‘frailty’ term (Kannisto 1991, Banerjee et al. 2003, Jones et al. 2010). In ecology, mixed models with ‘random effects’ are used to estimate distributions of individual heterogeneity in fitness components not attributable to observable traits (van de Pol and Wright 2009). In longitudinal studies, the simplest models of this class are random intercept models, where each individual is attributed a latent ‘baseline’ values for vital rates (Lewis et al. 2006, Royle 2008, Aubry et al. 2011, Gimenez and Choquet 2010, Marzolin et al. 2011), the other covariates (sex, age, etc.) adding their effects to the baseline individual values. This corresponds to the concept of ‘fixed heterogeneity’ (Link and Barker 2009, Bergeron et al. 2010). In practice, random effects (latent variables) in the statistical sense are used to account for unobserved fixed biological attributes of individuals (Link and Barker 2009). This contrasts with fixed effects in the statistical sense, which are used to define the observable states behind dynamic heterogeneity (sensu Tuljapurkar et al. 2009).

Observed variation across individuals can be modeled using covariates and factors, their time-varying analogues, or with dynamic state models (e.g. Markov models). Unobserved variation can be modeled with random effects (Gelman and Hill 2007); and combinations of the different approaches are used as well (Cam et al. 2002, Aubry et al. 2011, Orzack et al. 2011). The appropriate way to model variation in fitness components has been vigorously debated among those studying humans and captive animals (Service et al. 1998, Service 2000a, b, 2004, Drapeau et al. 2000), but the topic has received much less attention in ecology (but see Hawkes 2010, Steiner et al. 2010, Knape et al. 2011, Orzack et al. 2011). Although the distinction between approaches designed for inference on fitness at the individual level seems to be innocuous and largely technical, there are major consequences in terms of 1) conceptual views of fitness (Link and Barker 2009; see Discussion), 2) inference about population dynamics (Kendall et al. 2011, Stover et al. 2012), 3) bias in marginal estimates depicting phenotypic patterns of variation in vital rates over life, especially senescence (Vaupel et al. 1979, Vaupel and Yashin 1985, Zens and Peart 2003), and interpretation of such patterns.

Here, we address whether the diversity of life history sequences across individuals in a kittiwake *Rissa tridactyla* population is better accounted for by 1) statistical models including observable covariates exclusively: age, year, age of first breeding, 2) their analogs including breeding state in the previous year (Markovian ‘trait dynamics’; Cam et al. 1998, where heterogeneity reflects stochastic outcomes of Markovian state transitions sensu Tuljapurkar et al. 2009, Steiner et al. 2010, Steiner and Tuljapurkar 2012), or 3) the same models as in 1) and 2) but that also incorporate unobserved individual heterogeneity via random effects. Under the hypothesis of fixed differences in baseline vital rates among individuals, we expect models with individual random effects to perform better than models without them (Link et al. 2002b). We also address whether the probability of being in state  $s$  at time  $t + 1$

depends on state  $r$  at time  $t$  (either because of tradeoffs between fitness components, or short-term positive correlations between reproductive states). Under the hypothesis that the sequence of life history states in the data can be accounted for by a first-order Markov process (as opposed to the absence of a Markov process), we expected models including previous breeding state to perform better than models without it (Steiner et al. 2010). Our hypotheses are not mutually exclusive: models accounting for both features may perform better than either single-feature model. Depending on results concerning the individual random effects, our third objective is to elucidate the influence of unobserved fixed heterogeneity on conclusions about age-related changes in fitness components by comparing parameter estimates (especially the slopes for age effects) between models with and without individual random effects (Aalen et al. 2008a).

## Methods

Data from kittiwakes *Rissa tridactyla* were collected in Brittany (France) from 1987 to 2008. Birds were marked as chicks with unique combinations of colored plastic bands, plus a metal leg band. We excluded the very first years of the study (1979–1986) for two reasons: 1) the criteria to assign breeding state to individuals were progressively developed at the beginning, 2) and we wanted to include first-time breeders of different ages in most of the years retained for analysis, which was impossible for the earliest years. Data from the pre-breeding period were excluded. Recapture probability is virtually equal to one once birds recruit to the breeding population (Cam et al. 1998, 2005), which allows identification of the first breeding event and determination of age at first breeding (‘afr’). This also allows identification of the year of death (or of permanent emigration out of the study area; Cam et al. 2005). Breeding states were defined as follows. Breeding activity (breeding/nonbreeding) was recorded every year; non-breeders are individuals known to be alive in year  $t$ , to have bred in the past, but not breeding in year  $t$ . Individuals bred successfully if they raised at least one chick to independence; breeding success was defined for birds that attempted to breed (i.e. nonbreeding was not considered as breeding failure). We considered the states ‘success’ and ‘failure’ in breeders, but not separate states for the ‘number of chicks fledged’ in successful breeders. ‘Fixed’ differences among individuals imply ‘consistency’ in breeding state throughout life (e.g. a successful individual has a higher probability of breeding successfully in the following year(s) than an unsuccessful bird, and these respective probabilities are always higher for some birds than others, year after year). Assessing the hypothesis of fixed differences in the number of chicks fledged due to intrinsic individual features will require additional work, and probably the selection of a subset of data because of the large number of stochastic external drivers of failure from laying to fledging (up to 70 days, Cam et al. 2003). Last, to avoid issues with small sample size when stratifying the data set according to observable covariates, we defined five classes of age at first breeding (‘afr’): 2 and 3 years old, 4, 5 or 6 years

old, and 7 and older (Cam et al. 2005). Two-year old first time breeders were too few to form a separate group, as well as first time breeders older than 7 years old. Similarly, we pooled together individuals aged 18 years old or more. We acknowledge that such pooling may have consequences for results, but our priority was to test a set of a priori hypotheses, not to account for all possible variations in late-life fitness components.

## Approach to modeling

### 1) Complete data set

We modeled survival ( $\varphi$ ), breeding ( $\beta$ ) and reproductive success ( $\gamma$ ) probabilities as functions of age (linear or quadratic effects), age at first reproduction ('afr', a 5-level factor), year (random effect), and the interaction of 'afr' and age. We included breeding state in the previous year ('pbs', a 3-level factor: nonbreeder, failed breeder, successful breeder) and correlated individual random effects ( $\alpha$  terms in Eq. 1) in the most general model. Incorporation of previous breeding state in the model necessarily implied that only observations from breeding occasion 2 and onward for the individual are considered as outcomes of random variables. Previous breeding activity on occasion 1 for each individual was necessarily 'breeder', with modalities 'success' or 'failure'. We compared models where the probability of being in state  $s$  at time  $t+1$  depends on previous state  $r$  at time  $t$  with models where it does not (without Markovian state dependence). Year (22 levels) was treated as a random effect, because we had no a priori reason to expect any specific pattern or trend in demographic parameters over time, but wanted to control for the temporal variability in each vital rate driven by environmental conditions. Data from 1971 individuals were used (7893 alive/dead events, 6425 breeding/nonbreeding events, and 5541 success/failure events).

$$\begin{aligned} \text{logit}(\varphi) = & \mu_{\varphi} + f_{\varphi}(\text{age}) + g_{\varphi}(\text{afr}) + h_{\varphi}(\text{age} \times \text{afr}) \\ & + k_{\varphi}(\text{year}) + \eta_{\varphi}(\text{pbs}) + \alpha_{\varphi}(\text{individual}) \end{aligned} \quad (1)$$

and

$$\begin{aligned} \text{logit}(\beta) = & \mu_{\beta} + f_{\beta}(\text{age}) + g_{\beta}(\text{afr}) + h_{\beta}(\text{age} \times \text{afr}) \\ & + k_{\beta}(\text{year}) + \eta_{\beta}(\text{pbs}) + \alpha_{\beta}(\text{individual}) \end{aligned}$$

and

$$\begin{aligned} \text{logit}(\gamma) = & \mu_{\gamma} + f_{\gamma}(\text{age}) + g_{\gamma}(\text{afr}) + h_{\gamma}(\text{age} \times \text{afr}) \\ & + k_{\gamma}(\text{year}) + \eta_{\gamma}(\text{pbs}) + \alpha_{\gamma}(\text{individual}) \end{aligned}$$

Here,  $f_{\varphi}(\cdot)$ ,  $f_{\beta}(\cdot)$  and  $f_{\gamma}(\cdot)$  are functions of age (linear or quadratic),  $g_{\varphi}(\cdot)$ ,  $g_{\beta}(\cdot)$  and  $g_{\gamma}(\cdot)$  are functions of the 'afr' (factor),  $\eta_{\varphi}(\cdot)$ ,  $\eta_{\beta}(\cdot)$  and  $\eta_{\gamma}(\cdot)$  are functions of the previous breeding state,  $h_{\varphi}(\cdot)$ ,  $h_{\beta}(\cdot)$  and  $h_{\gamma}(\cdot)$  stand for the interaction between 'afr' and age,  $k_{\varphi}(\cdot)$ ,  $k_{\beta}(\cdot)$  and  $k_{\gamma}(\cdot)$  are year effects (independent normal random effects), and  $\alpha_{\varphi}(\cdot)$ ,  $\alpha_{\beta}(\cdot)$  and  $\alpha_{\gamma}(\cdot)$  are individual random effects. We used a trivariate normal distribution with mean 0 and a variance-covariance matrix  $\Sigma$  for individual random effects,

$$\Sigma = \begin{pmatrix} \sigma_{\varphi}^2 & \rho_{\varphi,\beta}\sigma_{\varphi}\sigma_{\beta} & \rho_{\varphi,\gamma}\sigma_{\varphi}\sigma_{\gamma} \\ \rho_{\varphi,\beta}\sigma_{\varphi}\sigma_{\beta} & \sigma_{\beta}^2 & \rho_{\beta,\gamma}\sigma_{\beta}\sigma_{\gamma} \\ \rho_{\varphi,\gamma}\sigma_{\varphi}\sigma_{\gamma} & \rho_{\beta,\gamma}\sigma_{\beta}\sigma_{\gamma} & \sigma_{\gamma}^2 \end{pmatrix}$$

where  $\sigma_{\varphi}^2$ ,  $\sigma_{\beta}^2$ , and  $\sigma_{\gamma}^2$  are the variances of the individual random effects on survival, breeding and success probability, respectively.  $\rho$  stands for the correlation between two random effects specified as subscripts. That is, we considered correlated individual random effects (e.g. individuals with a high underlying survival probability might also have a high probability of breeding, and if they do breed, a high probability of producing viable offspring). We only considered random-intercept models (i.e. fixed hierarchy among individuals); random intercept and slope models (Gelman and Hill 2007, van de Pol and Wright 2009) are beyond the scope of this paper.

### 2) Males and females

We addressed our two main hypotheses (individual random effects and effects of previous breeding state) using data from males and females separately. Birds were sexed through behavior and sex wasn't necessarily known before recruitment. Extra-pair paternity has been shown to be low in the study population (Helfenstein et al. 2004). Data from occasions where sex was unknown were excluded. Sex was not assigned retrospectively to avoid overestimating survival of one sex over the other if survival is sex-specific before age at sexing. We retained data from all location-years. In this population, we have no reason to assume that there is a relationship between mortality causes in members of a pair (e.g. a particularly high degree of age similarity in members of pairs, or pairs reunited during winter migration, etc.). In addition, nonbreeding is usually associated with divorce and the previous mate often breeds in the year following divorce. Conversely, both members of the pair necessarily share identical breeding success as long as they form a pair and breed. This lack of independence of observations for one of the random variables may lead to overestimation of the corresponding individual effect because consistency in success (if any) would appear to concern two individuals when only one observation unit (the pair) is concerned. Data from 885 females (3592 alive/dead events, 2950 breeding/nonbreeding events, 2492 success/failure events) and 1041 males (4206 alive/dead events, 3422 breeding/nonbreeding events, and 3011 success/failure events) were used. Note that for males the full model raised estimation issues that led us to start with a model without the interaction between age and age of first breeding for breeding probability (i.e. without  $\text{age} \times \text{afr}$  in Eq. 1).

### 3) Subset of data from 'favorable' local environmental conditions

We expected breeding state to depend on environmental conditions. Specifically, some colonies repeatedly experienced massive predation on eggs by corvids (*Corvus corax*, *Corvus corone*), inevitably leading to breeding failure. Massive dispersal to other colonies within the study area was observed in subsequent years (Cam et al. 2004).

This may create consistency in failure during the individual's life, e.g. repeated predation at the egg stage in the same location, then a dispersal event and the associated costs of dispersal. Because of identified causes of breeding failure and temporal autocorrelation in some of these causes (Danchin et al. 1998), individual consistency in the probability of breeding successfully may reflect environmental factors. Concerning breeding probability, some nonbreeders establish in a specific location but many do not (Cam et al. 2004). Consequently, they are not subjected to local environmental factors specific to breeding locations (feeding areas are likely the same for individuals from all the colonies considered here; Danchin et al. 1998), and differences in the level of consistency in breeding probability during their life are unlikely to reflect consistency in external conditions experienced by individuals. Local conditions (e.g. repeated predation on eggs) are more likely to create consistency in reproductive success than in the probability of deciding to breed during the individual's life. We therefore performed an analysis using data collected under 'favorable' local conditions excluding cases of obvious external cause of breeding failure between 1987 and 2008. Models with individual random effects may be selected using the complete data set, but not using the smaller one if individual random effects exclusively reflect external influences on reproduction and survival (as opposed to latent individual potential). Because of difficulties in quantifying massive predation events at the egg stage accurately in this retrospective study, results of this subset of analyses are reported in the Supplementary material Appendix A6.

### Missing data

Data were assumed to be missing at random (Lu and Copas 2004) when breeding success was unknown in a given occasion for an individual. The 'missing at random' assumption means that gaps in longitudinal data are positioned at random with regards to age for example (e.g. data selection did not create left- or right-censoring with a higher probability than other patterns). In the same vein, in analyses excluding obvious cases of massive predation at the egg stage (Supplementary material Appendix A6), observations used for analyses did not correspond to consecutive years in some individual histories. High predation at the egg stage in the location led to exclusion of some location-years. We assumed that the data selection process did not create any specific exclusion pattern.

### Estimation and model selection

We used a Bayesian approach to estimate model parameters. Analyses were conducted using Markov chain Monte Carlo (MCMC) simulations in the software program OpenBUGS (Lunn et al. 2009; see Supplementary material Appendix A1 for more details about the Methods; BUGS codes can be found in the Supplementary material Appendix A7). We ran three chains with different sets of initial values. Convergence was checked using the Brooks–Gelman–Rubin diagnostic  $\hat{R}$  for each model (Gelman and

Rubin 1992). We used the R (R Development Core Team) package CODA (Plummer et al. 2006) to analyze results.

Model selection was performed using the deviance information criterion 'DIC' (Spiegelhalter et al. 2002, Barnett et al. 2010). Smaller DIC values indicate that a model better approximates the biological processes that generated the data relative to other models. We also provided DIC weights (Stauffer 2008) to facilitate assessment of the importance of DIC differences among models in the set considered. Weights always range between 0 and 1; models with smaller DIC values have larger DIC weights. Weights provide an index of 'relative plausibility' of the models in the set considered (Cooch and White 2012). However, information criteria may tend to favor complex models with random effects (Plummer 2008). Consequently, we also used an alternative approach, based on 'inclusion variables' (O'Hara and Sillanpää 2009), to estimate the posterior probability of including individual random effects (see Supplementary material Appendix A1 for additional information). Inclusion variables are indicator variables that have a specified prior probability ( $p$  throughout this paper) and are associated with random effects. A prior probability of 0.50 corresponds to a situation where investigators do not favor any hypotheses a priori: an individual random effect may, or may not be relevant for the data set considered, with equal probability. Given prior probabilities, posterior probabilities of inclusion variables are interpreted in terms of degree of support for the hypothesis that the random effect is relevant to describe the variation in the data. Technical details for the analysis of correlated random effects using the inclusion variable approach are presented in Supplementary material Appendix A1.

Estimating model parameters in the way specified above (number of samples and chains, etc.) was time-consuming: from 2 to 8 days per model on average (sometimes more) on a PC running Windows, 3.06 GHz, 4 GB Ram using OpenBUGS. Because of the very large number of possible models and the time needed to estimate the parameters of each model, it was unrealistic to build all possible simplifications of the most general model. Consequently, we performed model selection in a sequential manner. Our main interest was in hypotheses concerning the role of previous breeding activity and success, individual random effects and the correlations among them. We first built models to address these hypotheses, and selected the model with the lowest DIC from that step and proceeded while keeping its structure. We also addressed hypotheses about covariates such as age and age of first breeding. We worked using the same covariates for the 3 random variables simultaneously, except in a few cases that were treated as post hoc hypotheses (i.e. not considered for DIC weights). More specifically, we addressed whether the functions of age were the same as in Link et al. (2002b); the current data set contains an additional six years compared to those used in the previous analysis. In addition, we considered models with and without individual random effects at different steps of the selection process: at the beginning and also using simpler models selected using DIC. The need for these random effects was assessed using DIC and inclusion variables.

## Sensitivity analyses

The Bayesian approach is characterized by the use of ‘prior’ distributions for parameters, which summarize what investigators know about parameters before data are collected, or what they believe. For example, for the slope of the age effect on survival, one may consider a normal prior distribution centered on 0, with a very large variance. Such a prior distribution reflects the fact that investigators do not favor positive or negative values for the linear and quadratic slopes, and consider that extreme values are less likely than values closer to 0. Because inference about fixed underlying heterogeneity among individuals is central to our study, we conducted sensitivity analyses to prior distributions for the individual random effects (Supplementary material Appendix A1, A3). These analyses were performed using the lowest-DIC models.

## Results

### 1) Complete data set

Based on DIC, models containing individual random effects were selected (Model 1 vs 2 and 1 vs 4, Table 1). The model with correlated individual random effects (Model 1) received larger support than the one with independent random effects (Model 5). In addition, when using a prior probability of inclusion of all the random effects simultaneously (and correlations) in the general model (Model 1) of 0.50, our results supported inclusion of individual random effects and their correlations (Supplementary material Appendix A2). For survival and success probability, results were sensitive to prior probabilities of inclusion of the individual effect in the general model (Supplementary material Appendix A2). Mean posterior probabilities for the variables including the individual random effect on survival or reproductive success probability decreased only when strongly penalizing these variables a priori ( $p = 0.125$  or  $0.008$ ). We are not aware of universally accepted rules concerning posterior probability thresholds above which the effects associated with inclusion variables would be systematically retained in models. Since all the mean posterior probability estimates of inclusion variables ranged between 0.79 and 1.00 when using a prior inclusion probability of 0.50 for the complete variance–covariance matrix  $\Sigma$  of individual random effects, we considered that this provided support for inclusion of the individual random effects in models.

The model with the lowest DIC (and largest DIC weight: 0.67) included age, age of first breeding (‘afr’) and the interaction between them (Model 9, Table 1). We did not find evidence that the covariate ‘pbs’ (breeding state in the previous year) was needed in the model to account for the process that gave rise to our data (Model 3 vs 1). Based on DIC, the year random effect was selected (Model 11 vs 3). The model without a quadratic effect of age had a smaller DIC than the model including it (Model 9 vs 3, Table 1), but DIC values were close (18830.60 vs 18832.00). Here again, there is no universally accepted rule concerning what constitutes a noteworthy difference in information criteria (Plummer 2008, Arnold 2010), but the DIC weight of

the model without the quadratic effect of age was larger (0.67) than the weight of the model including this effect (0.33). Our data did not allow definitive conclusions concerning the form of the relationship between fitness components and age. We chose the simplest model because the support for the quadratic effects of age was not large.

We used inclusion variables with the model with the largest DIC weight (Model 9, Table 1) to re-address the hypothesis of differences in fitness components among individuals (Supplementary material Appendix A2). Posterior probabilities of including individual random effects were large and results were insensitive to prior probability of including individual effects.

We also addressed some post hoc hypotheses about the quadratic effect of age on each fitness component separately. Model without age<sup>2</sup> on survival probability: DIC = 18814.1 (effective number of parameters = 1477.8); model without age<sup>2</sup> on breeding probability: DIC = 18853.7 (effective number of parameters = 1388.7); model without age<sup>2</sup> on success probability: DIC = 18841.1 (effective number of parameters = 1378.0). These DIC values were compared to 18830.6 (Model 9, Table 1;  $\Delta\text{DIC} = \text{DIC}_j - \text{DIC}_i$ ):  $\Delta\text{DIC} = 16.5$ ,  $-23.1$ , and  $-10.5$ , respectively. These results provided evidence that the quadratic effect of age wasn’t necessary to account for variation in survival in our data, but that it was for breeding and success probability. We used the model with DIC = 18814.1 and assessed the year effect again; model without year effect: DIC = 19326.0 (effective number of parameters = 1297.9;  $\Delta\text{DIC}: 18814.1 - 19326.0 = -511.9$ ). We thus retained ‘year’ in subsequent analyses.

Concerning sensitivity analyses, results for survival probability were sensitive to the parameters of the prior distribution placed on the variance–covariance matrix of individual random effects, but sensitivities were small for breeding and success probability (Supplementary material Appendix A3).

### Estimates from the model that received the largest support

Estimates made under the model without the quadratic effect of age on survival are reported in Table 2; they were obtained by performing the analyses a second time while discarding 9 samples out of 10 to reduce autocorrelation in the Markov chains. Survival probability decreased with age in all groups after recruitment (initial survival differed according to recruitment age), whereas probabilities of breeding and reproductive success increased with age (Fig. 1, Supplementary material Appendix A4). The rate of increase in breeding probability was initially substantial but slowed down with age. Estimates of fitness components and 95% credible intervals of the coefficients corresponding to the different ages of first breeding suggested that the 5 classes are not necessarily different (Table 2); post hoc contrasts would be needed to precisely assess such differences. In addition, the estimated covariance between survival and individual breeding probability on the one hand, and between breeding and success probability at the individual level on the other hand, were positive (their 95% credible interval excluded 0; Table 2). The covariance

Table 1. Model selection for survival, breeding, and success probability: inference about age, age of first breeding, year, previous breeding activity and success, and random individual effects. Complete data set.

Model	Trait	Age	Age <sup>2</sup>	Afr	Afr × Age	Pbs	Year RE	Ind RE	Corr	DIC	ENP	w
1*	φ	yes	yes	yes	yes	yes	yes	yes	yes	18871.0	937.20	0.00
	β	yes	yes	yes	yes	yes	yes	yes	yes			
	γ	yes	yes	yes	yes	yes	yes	yes	yes			
2	φ	yes	yes	yes	yes	yes	yes	no	NA	19017.4	95.00	0.00
	β	yes	yes	yes	yes	yes	yes	no	NA			
	γ	yes	yes	yes	yes	yes	yes	no	NA			
3	φ	yes	yes	yes	yes	no	yes	yes	yes	<b>18832.0</b>	1362.8	0.33
	β	yes	yes	yes	yes	no	yes	yes	yes			
	γ	yes	yes	yes	yes	no	yes	yes	yes			
4	φ	yes	yes	yes	yes	no	yes	no	NA	19673.1	89.0	0.00
	β	yes	yes	yes	yes	no	yes	no	NA			
	γ	yes	yes	yes	yes	no	yes	no	NA			
5	φ	yes	yes	yes	yes	yes	yes	yes	no	18920.2	720.6	0.00
	β	yes	yes	yes	yes	yes	yes	yes	no			
	γ	yes	yes	yes	yes	yes	yes	yes	no			
6	φ	yes	yes	yes	no	no	yes	yes	yes	18850.0	1313.0	0.00
	β	yes	yes	yes	no	no	yes	yes	yes			
	γ	yes	yes	yes	no	no	yes	yes	yes			
7	φ	yes	yes	no	no	no	yes	yes	yes	18852.0	1267.0	0.00
	β	yes	yes	no	no	no	yes	yes	yes			
	γ	yes	yes	no	no	no	yes	yes	yes			
8	φ	yes	yes	yes	yes	no	yes	yes	no	18960.5	1273.5	0.00
	β	yes	yes	yes	yes	no	yes	yes	no			
	γ	yes	yes	yes	yes	no	yes	yes	no			
9*	φ	yes	no	yes	yes	no	yes	yes	yes	<b>18830.6</b>	1520.8	0.67
	β	yes	no	yes	yes	no	yes	yes	yes			
	γ	yes	no	yes	yes	no	yes	yes	yes			
10	φ	yes	no	yes	no	no	yes	yes	yes	18882.9	1428.1	0.00
	β	yes	no	yes	no	no	yes	yes	yes			
	γ	yes	no	yes	no	no	yes	yes	yes			
11	φ	yes	yes	yes	no	no	no	yes	yes	19329.9	1290.2	0.00
	β	yes	yes	yes	no	no	no	yes	yes			
	γ	yes	yes	yes	no	no	no	yes	yes			

RE: random effect. Ind RE: individual random effect. Corr: correlation between individual random effects. DIC: deviance information criterion. ENP: effective number of parameters. w: deviance information criterion weight. φ = survival probability. β = breeding probability. γ = probability of raising at least one chick to independence given that the individual attempted to breed. Age<sup>2</sup> = quadratic effect of age on survival, breeding, and success probability. Afr = age of first breeding. Pbs = previous breeding activity and success. In bold: lowest-DIC models. \*Models used with inclusion variables (Chen and Dunson 2003).

between survival and the probability of breeding successfully was the smallest, and its credible interval included 0. The variance of the individual random effect on survival probability was the most difficult parameter to estimate. The autocorrelation in the chains (Supplementary material Appendix A5), was larger than for all other parameters, but the estimate of the Brooks–Gelman–Rubin diagnostic  $\hat{R}$  stabilized around 1.

### Estimates from a model without individual random effects

We compared estimates of fixed effects from the model accounting for individual heterogeneity (Table 2) with those from a model without heterogeneity (all other things being equal; Table 3). There were notable differences between estimates of intercepts and effects of age on survival made under the two models. In the random-intercept model (Table 2), the mean individual had lower survival and breeding probabilities and a higher probability of breeding successfully than the population mean obtained without

accounting for the variance among individuals (Table 3). Differences in intercepts between the models with and without random effects were associated with differences in some estimates of changes in survival probability with age (e.g. a higher initial survival probability was associated with a lower rate of change in survival with age). Importantly, the linear component of the relationship between age and survival was negative when individual variation was accounted for (with random effects, Table 2), whereas it was positive when such variation was ignored (Table 3).

### 2) Males

Based on DIC, results from males concerning our main hypotheses (Markovian structure of the model and individual random effects) were consistent with those obtained using the entire data set (Table 4). Concerning inclusion variables, with a prior probability of inclusion of correlated random effects of 0.50, results provided evidence of individual random effects on breeding probability, but provided smaller support for the random effect on the

Table 2. Estimates of coefficients and variance-covariance matrix of individual random effects of the best performing model (complete data set).

Covariate	Life history trait	Mean	Standard deviation	Median	Quantiles		
					0.025	0.975	
Intercept	$\phi$	1.3700	0.1507	1.3810	1.0400	1.6349	
	$\beta$	2.8350	0.2268	2.832	2.3940	3.2870	
	$\gamma$	0.1873	0.1409	0.1876	-0.0900	0.4643	
Age	$\phi$	-0.06600	0.0488	-0.0558	-0.1842	0.0012	
	$\beta$	0.3286	0.0400	0.3281	0.2518	0.4084	
	$\gamma$	0.1068	0.0227	0.1068	0.0624	0.1516	
Age <sup>2</sup>	$\beta$	-0.0257	0.0045	-0.0257	-0.0346	-0.0167	
	$\gamma$	-0.0257	0.0045	-0.0103	-0.0157	-0.0049	
Age of first breeding	4	$\phi$	0.0427	0.1017	0.0376	-0.1423	0.2593
	5	$\phi$	0.0241	0.1349	0.0134	-0.2116	0.3233
	6	$\phi$	0.0793	0.2047	0.06242	0.2759	0.5368
	>6	$\phi$	0.5858	0.4214	0.5587	-0.1657	1.4950
	4	$\beta$	0.0119	0.1724	0.0126	-0.3282	0.3466
	5	$\beta$	-0.0736	0.2039	-0.0747	-0.4709	0.3267
	6	$\beta$	-0.6190	0.2786	-0.6197	-1.1640	-0.0688
	>6	$\beta$	-0.8558	0.5433	-0.8637	-1.898	0.2312
	4	$\gamma$	-0.1033	0.1029	-0.1033	-0.3046	0.0991
	5	$\gamma$	-0.1987	0.1231	-0.1988	-0.4403	0.0432
	6	$\gamma$	0.0067	0.1837	0.0054	-0.3506	0.0432
	>6	$\gamma$	-0.6584	0.3571	-0.6570	-1.3620	0.3683
Interaction	4	$\phi$	-0.0266	0.0267	0.0263	-0.0800	0.0247
	5	$\phi$	-0.0419	0.0354	-0.0415	-0.1138	0.0257
	6	$\phi$	0.0051	0.0509	0.0051	-0.0949	0.1053
	>6	$\phi$	-0.0529	0.0816	-0.0534	-0.2127	0.1077
	4	$\beta$	-0.1166	0.04943	0.1160	-0.2154	0.0019
	5	$\beta$	-0.2180	0.0616	-0.2175	-0.3401	-0.0986
	6	$\beta$	-0.2109	0.0786	-0.2107	-0.3658	-0.0588
	>6	$\beta$	-0.0528	0.1277	-0.0532	-0.3022	0.1997
	4	$\gamma$	-0.0372	0.0268	-0.0372	-0.0899	0.0152
	5	$\gamma$	-0.0843	0.0359	-0.0844	-0.1548	-0.0141
	6	$\gamma$	0.0220	0.0555	0.0219	-0.0869	0.1315
	>6	$\gamma$	0.1254	0.0851	0.1249	-0.0398	0.2947
		SD $\phi$	0.7121	0.2639	0.6456	0.3666	1.3570
		SD $\beta$	1.4700	0.0927	1.4690	1.2930	1.6570
		SD $\gamma$	0.8997	0.0597	0.8989	0.7858	1.019
	Cov ( $\phi,\beta$ )	0.3994	0.1356	0.3891	0.1607	0.6955	
	Cov ( $\phi,\gamma$ )	0.1376	0.08435	0.1336	-0.0166	0.3149	
	Cov ( $\beta,\gamma$ )	0.8752	0.1224	0.8719	0.6439	1.1240	
Year effects	SD $\phi$	0.42894	0.0814	0.4191	0.2985	0.6162	
	SD $\beta$	0.6692	0.1215	0.6549	0.4737	0.6162	
	SD $\gamma$	0.4448	0.0829	0.4350	0.3110	0.6354	

$\phi$  = survival probability.  $\beta$  = breeding probability.  $\gamma$  = probability of raising at least one chick to independence given that the individual attempted to breed. Age<sup>2</sup> = quadratic effect of age on breeding and success probability. SD = standard deviation. Cov (X, Y) = covariance between variables X and Y. Burn-in = 5000, 1 every 6 samples retained for inferences, total number of samples = 50000.

probability of breeding successfully and survival (Supplementary material Appendix A2).

## Females

The DIC values of models with individual random effects were lower than those of models without heterogeneity (Table 4), which is consistent with the results obtained using the full data set. As for males, inclusion variables (prior probability of 0.50 for the entire variance-covariance matrix of random effects) provided evidence of the need for individual random effects on breeding probability (Supplementary material Appendix A2). Based on mean posterior probability of indicator variables, there was greater

support for inclusion of the individual random effects on reproductive success and survival probabilities than in males. The results were nevertheless slightly different in females for the fixed effects corresponding to one of our main hypotheses (Table 4). The models with and without previous breeding state had very similar DIC values (8633.3 versus 8634.0) suggesting that our data did not allow definitive conclusions concerning Markovian processes in this sex.

## 3) Subset of data from 'favorable' local environmental conditions

Results differed when data from massive predation periods were excluded (Supplementary material Appendix A6).



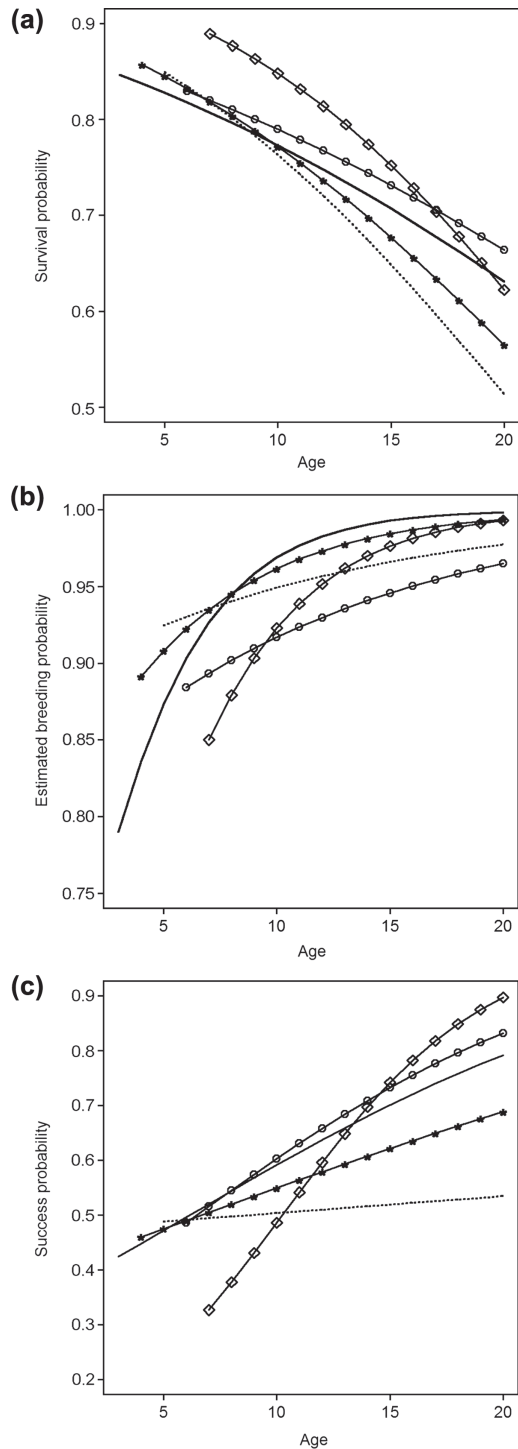


Figure 1. (a) Relationship between age, age of first breeding and survival probability. Line without symbol: age of first breeding  $\leq 3$ . Stars: age of first breeding = 4. Dashed line: age of first breeding = 5. Circles: age of first breeding = 6. Squares: age of first breeding  $\geq 7$ . (b) Relationship between age, age of first breeding and breeding probability. Line without symbol: age of first breeding  $\leq 3$ . Stars: age of first breeding = 4. Dashed line: age of first breeding = 5. Circles: age of first breeding = 6. Squares: age of first breeding  $\geq 7$ . (c) Relationship between age, age of first breeding and probability of reproductive success. Line without symbol: age of first breeding  $\leq 3$ . Stars: age of first breeding = 4. Dashed line: age of first breeding = 5. Circles: age of first breeding = 6. Squares: age of first breeding  $\geq 7$ .

Even with prior values of 0.50 for inclusion variables, posterior probabilities indicated that the hypothesis of heterogeneity in individual success probability received small support. Based on DIC, the correlations between individual random effects were not selected in this case.

## Discussion

### Which model best accounts for observed variation in individual life histories?

#### Individual random effects

Our results provided evidence that there are differences in baseline survival, breeding and success probability among individuals in this population when using the complete data set (model selection based on deviance information criterion; Spiegelhalter et al. 2002). There was a positive estimated covariance between survival and breeding probability at the individual level (the higher the breeding probability, the higher the survival probability), as well as between breeding and reproductive success probability. With indicator variables to assess posterior probabilities that individual random effects help describe variation in the data, we obtained consistent results when considering a prior probability of inclusion of the variance–covariance matrix of individual random effect of 0.50 (i.e. when the hypotheses of homogeneity and heterogeneity had equal prior probabilities). Previous longitudinal studies of vital rates in wild vertebrates have usually considered prior probabilities of inclusion of 0.50 (Royle 2008). Results differed when data from massive predation on eggs periods were excluded. The hypothesis of heterogeneity in success probability among individuals received small support, and correlations between individual random effects were not selected. This suggests that part of the heterogeneity in the probability of breeding successfully among individuals in the complete data set results from contrasted environmental conditions experienced by individuals and differences in behavioral responses according to the conditions (e.g. dispersal and divorce Cam et al. 2004, Supplementary material Appendix A6).

#### Previous breeding state (first-order Markov process)

With the complete data set, we did not find evidence that survival probability from time  $t$  to  $t + 1$ , or breeding and reproductive success probability at time  $t + 1$  depended on breeding activity and success in year  $t$ . Results were more ambiguous when analyzing data from females separately: models with and without previous breeding state obtained virtually equal support, which suggested that our data do not allow definitive conclusions for Markovian processes in females. This may seem at odds with previous work on this data set using multistate capture–mark–recapture models (Lebreton et al. 2009) based on two reproductive states (Cam et al. 1998). Specifically, in previous work, there was unambiguous evidence of an influence of previous state on both survival and future breeding probability (Cam et al. 1998). The multistate model in question also accounted for yearly variation in both survival and breeding probability, but differences in baseline vital

Table 3. Estimates of coefficients of a modified version of the best performing model: without individual random effects (complete data set).

Covariate	Life history trait	Mean	Standard deviation	Median	Quantiles		
					0.025	0.975	
Intercept	$\phi$	1.5250	0.1062	1.5250	1.3160	1.7360	
	$\beta$	2.4930	0.1648	2.4910	2.1720	2.8190	
	$\gamma$	0.3342	0.1055	0.3340	0.1277	0.5434	
Age	$\phi$	0.0033	0.0176	0.0031	-0.0310	0.0382	
	$\beta$	0.3252	0.0315	0.3245	0.2653	0.3889	
	$\gamma$	0.0943	0.0176	0.0943	0.0600	0.1292	
Age <sup>2</sup>	$\beta$	-0.0218	0.0038	-0.0219	-0.0293	-0.0431	
	$\gamma$	-0.0105	0.0024	-0.0105	-0.0152	-0.0058	
	$\phi$	-0.0223	0.0731	-0.0213	-0.1409	0.2671	
Age of first breeding	4	$\phi$	-0.1057	0.0892	-0.1057	-0.2804	0.0689
	6	$\phi$	-0.1159	0.1392	-0.1169	-0.3847	0.1597
	> 6	$\phi$	0.2560	0.3237	0.2492	-0.3608	0.9121
	4	$\beta$	-0.0551	0.115	-0.0537	-0.2850	0.1676
	5	$\beta$	-0.2564	0.1341	-0.2564	-0.5199	0.0068
	6	$\beta$	-0.6726	0.1885	-0.6735	-1.0400	-0.2995
	> 6	$\beta$	-0.9553	0.3862	-0.9644	-1.6860	-0.1698
	4	$\gamma$	-0.0943	0.0692	-0.0944	-0.2299	0.04182
	5	$\gamma$	-0.2153	0.0856	-0.2154	-0.3830	-0.0471
	6	$\gamma$	-0.0606	0.1369	-0.0612	-0.3291	0.2099
	> 6	$\gamma$	-0.5568	0.2844	-0.5567	-1.1140	-0.0013
	Interaction	4	$\phi$	-0.0177	0.0228	-0.0175	-0.0626
5		$\phi$	-0.0245	0.0295	-0.0246	-0.0821	0.0034
6		$\phi$	0.0123	0.0453	0.0119	-0.0756	0.1022
> 6		$\phi$	-0.0504	0.0738	-0.0511	-0.1922	0.0949
4		$\beta$	-0.0870	0.04947	-0.0834	-0.1686	-0.0093
5		$\beta$	-0.1621	0.0490	-0.1616	-0.2597	-0.0675
6		$\beta$	-0.1658	0.0652	-0.1658	-0.2887	-0.0438
> 6		$\beta$	-0.0558	0.0987	-0.0569	-0.2463	0.1404
4		$\gamma$	-0.0201	0.0214	-0.0200	0.0622	0.0217
5		$\gamma$	-0.0601	0.0288	-0.0599	-0.1167	-0.0038
6		$\gamma$	0.0350	0.0464	0.0348	-0.0549	0.1266
> 6		$\gamma$	0.1152	0.0719	0.1146	-0.0237	0.2580
Year effects	SD $\phi$	0.4055	0.0744	0.2851	0.3969	0.5754	
	SD $\beta$	0.5688	0.1029	0.5565	0.4042	0.8043	
	SD $\gamma$	0.3872	0.0721	0.3786	0.2705	0.5518	

$\phi$  = survival probability.  $\beta$  = breeding probability.  $\gamma$  = probability of raising at least one chick to independence given that the individual attempted to breed. Age<sup>2</sup> = quadratic effect of age on breeding and success probability. SD = standard deviation.

rates among individuals were not considered. Elsewhere, age-related variation in both survival and breeding probability has been addressed in the kittiwake using random effects models (Cam et al. 2002, Link et al. 2002a, b, Wintrebert et al. 2005), but the Markovian process was not considered. In contrast to Markovian state dynamics, these papers considered current breeding state, but not state in the previous year, and fitness components were not state-specific. Thus, when considering previous results, it is important to recognize that none of those papers considered all hypotheses simultaneously as was done in the work presented here. Importantly, if we had not considered individual random effects at all, such as in Cam et al. (1998) and Steiner et al. (2010), the model including previous breeding state would have been selected with the complete data set (DIC = 19017.4 vs 19673.1).

Moreover, in the current analyses, incorporation of heterogeneity among individuals in a first-order Markov model did not result in detection of any negative relationship between current and future reproduction, or between previous reproduction and survival. In females, where

models with and without previous breeding state received virtually equal support, nonbreeders had the lowest probability of breeding in the following year compared to failed and successful breeders. Results were similar for success probability (results not shown). Conversely, nonbreeders and failed breeders had equal survival probability, which was lower than that of successful breeders. Assuming that there are short-term tradeoffs between fitness components within individuals, they remained undetected at the phenotypic level in this observational study.

#### **Inference about age effects**

As in Cam et al. (2002) and Link et al. (2002b), our current results confirmed that inference about variation in survival with age differs according to whether the model is individual-specific, or not (see also Knape et al. 2011), and that the intercept of the mixed model (mean individual) differs from that of the fixed-effects model (population mean). There is a major difference between inference based on the individual-specific mixed model and the fixed effect model: a decline in survival with age with the former,

Table 4. Model selection for survival, breeding, and success probability in males and females. Inferences about previous breeding state and individual effects.

Model	Trait	Age	Age <sup>2</sup>	Afr	Afr × Age	Pbs	Year RE	Ind RE	Corr	DIC	ENP	w
<b>Males</b>												
1*	φ	yes	yes	yes	yes	yes	yes	yes	yes	9954.2	608.6	0.00
	β	yes	yes	yes	yes	yes	yes	yes	yes			
	γ	yes	yes	yes	yes	yes	yes	yes	yes			
2	φ	yes	yes	yes	yes	yes	yes	no	NA	10036.5	87.6	0.00
	β	yes	yes	yes	yes	yes	yes	no	NA			
	γ	yes	yes	yes	yes	yes	yes	no	NA			
3	φ	yes	yes	yes	yes	no	yes	yes	yes	<b>9929.3</b>	782.8	1.00
	β	yes	yes	yes	yes	no	yes	yes	yes			
	γ	yes	yes	yes	yes	no	yes	yes	yes			
4	φ	yes	yes	yes	yes	yes	yes	yes	no	9977.7	609.8	0.00
	β	yes	yes	yes	yes	yes	yes	yes	no			
	γ	yes	yes	yes	yes	yes	yes	yes	no			
<b>Females</b>												
1*	φ	yes	yes	yes	yes	yes	yes	yes	yes	<b>8633.3</b>	543.1	0.59
	β	yes	yes	yes	yes	yes	yes	yes	yes			
	γ	yes	yes	yes	yes	yes	yes	yes	yes			
2	φ	yes	yes	yes	yes	yes	yes	no	NA	8720.5	86.2	0.00
	β	yes	yes	yes	yes	yes	yes	no	NA			
	γ	yes	yes	yes	yes	yes	yes	no	NA			
3	φ	yes	yes	yes	yes	no	yes	yes	yes	8634.0	724.2	0.41
	β	yes	yes	yes	yes	no	yes	yes	yes			
	γ	yes	yes	yes	yes	no	yes	yes	yes			
4	φ	yes	yes	yes	yes	yes	yes	yes	no	8666.0	507.9	0.00
	β	yes	yes	yes	yes	yes	yes	yes	no			
	γ	yes	yes	yes	yes	yes	yes	yes	no			

RE: random effect. Ind RE: individual random effect. Corr: correlation between individual random effects. DIC: deviance information criterion. ENP: Effective number of parameters. w: deviance information criterion weight. φ = survival probability. β = breeding probability. γ = probability of raising at least one chick to independence given that the individual attempted to breed. Age<sup>2</sup> = quadratic effect of age on survival, breeding probability, and success probability. Afr = age of first breeding. Pbs = previous breeding activity and success. In bold: lowest-DIC models. \*Models used with inclusion variables (Chen and Dunson 2003).

but no such decline with the latter. This is consistent with hypotheses considered in human demography (Vaupel et al. 1979, Aalen et al. 2008a). More complex models may better account for the finer details in age-specific patterns of variation in fitness components, but we leave this topic for future work.

### Where does individual variation in life histories come from?

#### Different views of heterogeneity

The question of consequences of heterogeneity in wild animal populations for life history and population dynamics studies is currently receiving much attention (Bergeron et al. 2010, Wilson and Nussey 2010, Caswell 2011, Kendall et al. 2011, Stover et al. 2012), but the controversies about the relevance of different models of heterogeneity will be fruitful only if the differences among views are clearly identified. Four conceptual views of heterogeneity can be distinguished. 1) Because survival, breeding, and success events were accounted for by a first-order Markov process, Tuljapurkar et al. (2009) and Steiner et al. (2010) called the resulting heterogeneity among life histories ‘dynamic heterogeneity’: heterogeneity caused by the stochastic movement of individuals among homogeneous strata in a population stratified according to a small number of observable criteria. 2) This observable ‘dynamic heterogeneity’ should be distinguished from ‘dynamic

frailty’ of the statistical, medical and economical literature (Yue and Chan 1997, Pennel and Dunson 2005), which corresponds to models with individual-specific parameters (unobserved heterogeneity) that change over life. 3) Alternatively, variation among individual life histories can be accounted for by models incorporating unobserved heterogeneity among individuals where the baseline individual vital rate doesn’t change after birth or recruitment (depending on the type of data used): fixed heterogeneity (Link and Barker 2009). This approach is analogous to classical ‘frailty’ continuous-time mortality models in one respect: that the hierarchy in unobserved mortality risk among individuals is fixed (Aalen et al. 2008a, b). 4) Last, random-intercept and slope models have been considered in life history studies (van de Pol and Wright 2009, Martin et al. 2011); they incorporate individual-specific fixed underlying baseline vital rates and fixed individual-specific rate of change in vital rates over life. These models may lead to changes in the hierarchy among individual vital rates in survivors when individuals get older, but they still imply fixed individual underlying features (slope and intercept). In this paper, we considered 1), 3) and a combination of the two views.

#### Contrasted approaches to draw inferences about ‘fixed, underlying’ heterogeneity in vital rates

Steiner et al. (2010) hypothesized that fixed underlying heterogeneity (i.e. individual random effects) will be detected

even in the absence of differences among individuals if the Markovian state dynamics are not taken into account (Cam et al. 2002, Link et al. 2002b, Wintrebert et al. 2005, Aubry et al. 2011), i.e. there will be 'spurious support for the existence of latent traits' (Steiner et al. 2010). This implies that the information in the data accounted for by individual random effects in models is confounded with the information accounted for by a first-order fixed-effect Markov model of life history states. Here we addressed the combination of two features: the individual random effects and the first-order Markovian structure of transitions between observable states in consecutive years. If differences in baseline fitness components among individuals play an important part in determining the sequences of life history states in our dataset, and if previous breeding state does too, our full model with random intercepts and first-order Markov processes should have received better support. Rather, our results provided evidence that the sequences of life history states in individual histories of kittiwakes are better accounted for by random-intercept models where intercepts are individual-specific compared to first-order Markov models or a combination of the two features. We moreover did not find strong evidence that accounting for the Markovian dynamics is necessary for this data set.

Data from the same study population of kittiwakes breeding at Cap Sizun, France, have now been used in two very different ways to draw inferences about the hypothesis of unobserved heterogeneity among individuals. As in their other work (Tuljapurkar et al. 2009, Orzack et al. 2011, Steiner and Tuljapurkar 2012), Steiner et al. (2010) developed support for the ideas of dynamic heterogeneity and neutral generation of life history variation using numerical simulations. In contrast, we chose to estimate and compare models accounting for different biological hypotheses using an inferential statistical framework designed to assess the degree of consistency of the models with our data (Burnham and Anderson 2002). Our motivation was twofold: 1) completely different underlying processes may lead to identical simulated patterns, a well-known problem when testing the neutral theory in community ecology for example (Chave 2004), 2) progress in statistical theory and methods have led to the development of ways to quantify differences between model parameterizations that reflect neutral versus non-neutral models (Jabot and Chave 2009, 2011, Beaumont et al. 2010, Clark 2010). Historical limitations of estimation methods have hampered assessment of complex hypotheses involving differences among individuals and species in community ecology (Jabot and Chave 2009, 2011). Recent statistical techniques (Beaumont et al. 2010) have been used to estimate individual- and species-level models in community ecology, in an inference framework designed to assess several traditionally 'alternative' ecological hypotheses (trait-based competition among species and neutral theory; Jabot and Chave 2009). Here, because the likelihood of models with individual effects can be defined and such models estimated, we chose an approach based on an information criterion (Burnham and Anderson 2002) to draw inferences about heterogeneity among individual vital rates. Information criteria do not allow inferences in terms of 'true model': they quantify the relative

ability of different models to account for the processes that gave rise to a particular dataset (Burnham and Anderson 2002), but all the models considered may be very distant from 'truth'. Here, individual-specific models received larger support than others.

#### ***Limitations of inferences based on mixed models***

The decision to model 'latent' (individual) effects is strongly influenced by habits and priorities specific to different research areas. It is extremely common in social and political sciences, public health, economics, education, psychometrics (Gelman and Hill 2007), where unobserved risk factors are routinely considered even if their 'cause' or 'origin' remains unknown. The motivation for this approach is the inability of investigators to measure all the relevant variables in empirical studies to assess the 'propensity' of individuals to express a given modality of a trait (e.g. mortality; Duffie et al. 2009), either for practical reasons (e.g. retrospective studies) or because current theoretical bases are not sufficiently advanced to consider a priori some factors as relevant. Surprisingly, 'latent' traits models are not so common in demographic studies of wild vertebrates despite the complexity of the traits addressed (e.g. the probability of recruiting to the breeding segment of a population at a given age). This may be due to the classical limitations of data sets from wild vertebrates (we may contact an individual only once or twice before recruitment). However, modeling such differences among individuals can be technically difficult. In human demography, not all classical hazard models accounting for 'frailty' can be estimated without restrictive assumptions (Yashin et al. 2001, Aalen et al. 2008a, b). In all research areas using models for individual latent rates, there are a priori hypotheses about the distribution of latent values that cannot necessarily be assessed because associated statistical theory hasn't been developed yet, or because methods are still novel and actively debated. Model selection with random effects is currently an active area of research (Burnham and White 2002, Spiegelhalter et al. 2002, Celeux et al. 2006, Plummer 2008, O'Hara and Sillanpää 2009). Several studies have shown that inferences about random effects based on the deviance information criterion are robust (Wilberg and Bence 2008, Barnett et al. 2010), but we lack tools to assess whether the conditions under which they are robust are met with a particular empirical data set. Development of statistical theory and techniques may lead us to update our inferences in the future.

#### ***Quantification of the degree of heterogeneity in simulation studies***

Recently, Caswell (2011) developed a new simulation approach that quantifies the extent to which variability in lifetime reproductive output exceeds the baseline level created when a set of identical vital rates is applied to a cohort of individuals. This is an important step forward because 1) it could be used to help identify inconsistencies among conclusions drawn in different studies based on the same data (Cam et al. 2002, Link et al. 2002a, b, Link and Barker 2009, Steiner et al. 2010, this study), and 2), it provides a framework for decomposing individual-level variation in life histories into the selectively neutral

and non-neutral components. To our knowledge, life history simulation studies have not yet used models that accommodate latent individual heterogeneity (e.g. with random effects). However, as shown here, estimates of fixed effects (e.g. age effects) may differ depending on whether or not individual random effects are considered. The differences between parameters that should be used to simulate data do not concern exclusively variance terms: coefficients associated with variations in the mean may also be relevant. The entries in transition matrices used in simulations are sensitive to the class of statistical models used to estimate vital rates. We see great potential for integrating models like ours into the simulation approach developed by Caswell (2011).

### **‘Latent fitness’ and fixed heterogeneity**

#### ***Individual-level models in animal demography: historical context***

The relevance of the simplest individual-specific models (random-intercept models), and of fixed underlying heterogeneity in general is currently debated in the evolutionary ecology literature because of difficulties in identifying measurable criteria associated with such differences in fitness (Wilson and Nussey 2010, Bergeron et al. 2010). Research aiming at quantifying these differences, assessing whether they are relevant to natural selection, and identifying traits whose evolution may depend on such differences is central to evolutionary ecology (Wilson and Nussey 2010). However, it is important to note that the first studies of life histories that subsequently inspired the development of individual-specific models did not speculate on the relevance of heterogeneity to the ‘consequences of natural selection’ (Fairbain and Reeve 2001). Several decades ago, Curio (1983) explicitly referred to the within-cohort phenotypic mortality selection hypothesis (the ‘process of natural selection’; Fairbain and Reeve 2001) to explain the increase in breeding performance with age observed in many bird species: ‘The first [explanation] assumes that young birds of ‘bad quality’ both breed worse and then die earlier than those of superior quality’. Simultaneously, in human demography, a conceptually similar process was suspected to explain some patterns of variation in mortality risk with age (Vaupel et al. 1979, Vaupel and Yashin 1985), within a different theoretical framework. The motivation to develop and implement individual-level statistical models in animal demography was the use of long-term longitudinal data sets in retrospective studies: measuring proxies of ‘quality’ or ‘frailty’ was not an option, and the authors were interested in 1) patterns of age-related change over life in fitness components at the individual level, and 2) biases in inferences about these patterns when ignoring heterogeneity.

#### ***Fitness as the realization of an individual-specific model***

More fundamental are the implications of inferences at the individual level for conceptual views of fitness. To estimate fitness components and lifetime fitness, Link et al. (2002a) and Link and Barker (2009) advocated an approach where the vital rates are defined even if the individual dies immediately after birth; the a priori, ‘latent’

fitness at birth of such individuals isn’t necessarily null, but the realization of a sequence of a Bernoulli random variables governed by survival probability may result in immediate death at birth or independence (fledging, weaning) even if the expected value of survival is  $> 0$ . The a priori fitness is a conceptual construct with biological bases (e.g. the potential associated with the genetic make-up of the individual, or ontogenetic development in studies starting at recruitment) and practical implications: it allows estimation and inference about the underlying distribution of individual fitness components and lifetime fitness in the population (Link and Barker 2009), a capability that has long been sought in quantitative genetics (Charlesworth 1994). The key point is that the observed life histories are considered as realizations of an individual-specific stochastic model whose form is explicitly defined and whose parameters can be estimated. Assuming that investigators manage to identify relevant measurable variables to replace all latent heterogeneity in fitness components, investigators will likely face the problem of large degrees of stratification of the data unless a few known variables account for a large proportion of the heterogeneity (Cooch et al. 2002). Otherwise, scalar abstractions (also a conceptual construct) of multiple phenotypic traits are necessary (Wilson and Nussey 2010). In individual-based models, the collection of individual values is explicitly modeled by simply estimating the ‘distribution’ of individual values (e.g. a variance among individuals). In the Bayesian framework, Link and Barker (2009) took a step further and assessed the population distribution of individual fitness values using hyperparameter estimates.

### **Conclusion**

The inferential statistical framework based on individual-level models makes it possible to compare hypotheses pertaining to the sources of individual-level variation in life histories and assess the corresponding evolutionary or ecological implications. This framework can also be used to address whether observable covariates account for differences in fitness among individuals in a satisfactory manner (proxies for ‘quality’; Bergeron et al. 2010, Wilson and Nussey 2010). It is moreover the basis of recent capture–mark–recapture developments to the ‘animal models’ used in quantitative genetics (Papaix et al. 2010), where the use of latent individual traits has a long history (Moorad and Promislow 2011). Teasing apart the sources of individual heterogeneity of fitness components could also have important implications for our understanding of population dynamics and viability (Kendall et al. 2011, Stover et al. 2012), as well as coexistence amongst species in diverse communities (Clark 2010).

*Acknowledgements* – We thank the Conseil Général du Finistère for allowing us to work in the nature reserve of Goulien Cap Sizun (Brittany, France, managed by Bretagne Vivante, the Société pour l’Étude et la Protection de la Nature en Bretagne) and in the Pointe du Raz. The Institut Universitaire Européen de la Mer (Plouzané, France) provided financial support for this long-term study. LMA was supported by NSF DEB 1019613. EC and RAJ were supported by the Programme d’Actions Universitaires Intégrés

Luso-Françaises (Laboratoire Évolution et Diversité Biologique, Université Paul Sabatier, France - Depto de Matemática, Universidade de Évora, Portugal). We are very grateful to Dan Nussey, Matthew Symonds and John Sauer for providing helpful comments on an earlier version of the manuscript. We thank all the people involved in data collection, especially Jacques Nisser and Morgane Huteau for their help marking birds, and all the processors of Euler, Diesel and Eagle1, the three computers that have been working on this paper for more than a year. This study was authorized by the Centre de Recherches sur la Biologie des Populations d'Oiseaux (CRBPO, Museum d'Histoire Naturelle, Paris, France). The permit to capture kittiwakes and mark them was granted by the CRBPO. All work was carried out in accordance with standard animal care protocols approved by the CRBPO.

## References

- Aalen, O. O. et al. 2008a. Unobserved heterogeneity: the odd effects of frailty. – In: Aalen, O. O. et al. (eds), *Survival and event history analysis: a process point of view. Statistics for biology and health.* Springer Science + Business Media, pp. 231–270.
- Aalen, O. O. et al. 2008b. Diffusion and Lévy process models for dynamic frailty. – In: Aalen, O. O. et al. (eds), *Survival and event history analysis: a process point of view. Statistics for biology and health.* Springer Science + Business Media, pp. 425–455.
- Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike's information criterion. – *J. Wildlife Manage.* 74: 1175–1178.
- Aubry, L. M. et al. 2011. Drivers of age-specific survival in a long-lived seabird: contributions of observed and hidden sources of heterogeneity. – *J. Anim. Ecol.* 80: 375–383.
- Banerjee, S. et al. 2003. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. – *Biostatistics* 4: 123–142.
- Barnett, A. G. et al. 2010. Using information criteria to select correct variance-covariance structure for longitudinal data in ecology. – *Meth. Ecol. Evol.* 1: 15–24.
- Beaumont, M. et al. 2010. In defence of model-based inference in phylogeography. – *Mol. Ecol.* 19: 436–446.
- Bergeron, P. et al. 2010. Individual quality: tautology or biological reality? – *J. Anim. Ecol.* 80: 361–364.
- Brommer, J. E. et al. 2002. Reproductive timing and individual fitness. – *Ecol. Lett.* 5: 802–810.
- Burnham, K. P. and Anderson, D. R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. – Springer.
- Burnham, K. P. and White, G. C. 2002. Evaluation of some random effects methodology applicable to bird ringing data. – *J. Appl. Stat.* 29: 245–264.
- Cam, E. et al. 1998. Are nonbreeders prudent parents? The Kittiwake model. – *Ecology* 79: 2917–2930.
- Cam, E. et al. 2002. Individual covariation in life history traits: seeing the trees despite the forest. – *Am. Nat.* 159: 96–105.
- Cam, E. et al. 2003. Long-term consequences of early conditions in the kittiwake. – *J. Anim. Ecol.* 72: 411–424.
- Cam, E. et al. 2004. Dispersal and individual quality in a long lived species. – *Oikos* 106: 386–398.
- Cam, E. et al. 2005. Earlier recruitment or earlier death? On the assumption of homogeneous survival in recruitment studies. – *Ecol. Monogr.* 75: 419–434.
- Caswell, H. 2011. Beyond R-0: demographic models for variability of lifetime reproductive output. – *PLoS One* 6: e20809. doi: 10.1371/journal.pone.0020809
- Celeux, G. et al. 2006. Deviance information criteria for missing data models (with Discussion). – *Bayesian Anal.* 1: 651–706.
- Charlesworth, B. 1994. *Evolution in age-structured populations.* – Cambridge Univ. Press.
- Chave, J. 2004. Neutral theory and community ecology. – *Ecol. Lett.* 7: 241–253.
- Chen, Z. and Dunson, D. B. 2003. Random effects selection in linear mixed models. – *Biometrics* 59: 762–769.
- Clark, J. S. 2010. Individuals and the variation needed for high species diversity in forest trees. – *Science* 327: 1129–1132.
- Clutton-Brock, T. H. 1988. *Reproductive success. Studies of individual variation in contrasting breeding systems.* – Univ. of Chicago Press.
- Cooch, E. G. and White, G. C. 2012. Program Mark, a gentle introduction. < [www.phidot.org/software/mark/docs/book/](http://www.phidot.org/software/mark/docs/book/) > (accessed 23 April 2012).
- Cooch, E. G. et al. 2002. Occam's shadow: levels of analysis in evolutionary ecology – where to next? – *J. Appl. Stat.* 29: 19–48.
- Curio, E. 1983. Why do young birds reproduce less well? – *Ibis* 125: 400–404.
- Danchin, E. et al. 1998. Conspecific reproductive success and breeding habitat selection: implications for the evolution of coloniality. – *Ecology* 79: 2415–2428.
- Drapeau, M. D. et al. 2000. Testing the heterogeneity theory of late-life mortality plateaus by using cohorts of *Drosophila melanogaster*. – *Exp. Gerontol.* 35: 71–84.
- Duffie, D. et al. 2009. Frailty correlated default. – *J. Finance* 64: 2089–2123.
- Fairbain, D. J. and Reeve, J. P. 2001. Natural selection. – In: Fox, C. W. et al. (eds), *Evolutionary ecology.* Oxford Univ. Press, pp. 29–43.
- Gelman, A. and Rubin, D. B. 1992. Inference from iterative simulation using multiple sequences. – *Stat. Sci.* 7: 457–472.
- Gelman, A. and Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models.* – Cambridge Univ. Press.
- Gimenez, O. and Choquet, R. 2010. Individual heterogeneity in studies on marked animals using numerical integration: capture-recapture mixed models. – *Ecology* 91: 951–957.
- Grafen, A. 1988. On the uses of data on lifetime reproductive success. – In: Clutton-Brock, T. H. (ed.), *Reproductive success.* Univ. of Chicago Press, pp. 454–471.
- Hawkes, K. 2010. How grandmother effects plus individual variation in frailty shape fertility and mortality: guidance from human-chimpanzee comparisons. – *Proc. Natl Acad. Sci. USA* 107: 8977–8984.
- Helfenstein, F. et al. 2004. Low frequency of extra-pair paternity and high frequency of adoption in black-legged kittiwakes. – *Condor* 106: 149–155.
- Jabot, F. and Chave, J. 2009. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. – *Ecol. Lett.* 12: 239–248.
- Jabot, F. and Chave, J. 2011. Analyzing tropical forest tree species abundance distributions using a non-neutral model and through approximate Bayesian inference. – *Am. Nat.* 178: E37–47.
- Jones, J. H. et al. 2010. Phenotypic quality influences fertility in Gombe chimpanzees. – *J. Anim. Ecol.* 79: 1262–1269.
- Kannisto, V. 1991. Frailty and survival. – *Genus* 47: 101–118.
- Kendall, B. E. et al. 2011. Demographic heterogeneity, cohort selection and populations growth. – *Ecology* 92: 1985–1993.
- Knappe, J. H. et al. 2011. Individual heterogeneity and senescence in silvereyes on Heron Island. – *Ecology* 92: 813–820.
- Lebreton, J. D. et al. 2009. Modeling individual animal histories with multistate capture-recapture models. – *Adv. Ecol. Res.* 41: 87–173.
- Lewis, S. et al. 2006. Determinants of quality in a long-lived colonial species. – *J. Anim. Ecol.* 5: 1304–1312.
- Link, W. A. and Barker, R. J. 2009. *Bayesian Inference with ecological examples.* – Academic Press.

- Link, W. A. et al. 2002a. Model-based estimation of individual fitness. – *J. App. Stat.* 29: 207–224.
- Link, W. A. et al. 2002b. Of BUGS and birds. An introduction to Markov chain Monte Carlo. – *J. Wildlife Manage.* 66: 277–291.
- Lu, G. and Copas, J. D. 2004. Missing at random, likelihood ignorability and model completeness. – *Ann. Stat.* 32: 754–765.
- Lunn, D. et al. 2009. The BUGS project: evolution, critique and future directions (with discussion). – *Stat. Med.* 28: 3049–3082.
- Martin, J. G. A. et al. 2011. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. – *Meth. Ecol. Evol.* 2: 362–274.
- Marzolin, G. et al. 2011. Frailty in state-space models: application to actuarial senescence in the dipper. – *Ecology* 92: 562–567.
- Moorad, J. A. and Promislow, D. E. L. 2011. Evolutionary demography and quantitative genetics: age-specific survival as a threshold trait. – *Proc. R. Soc. B* 278: 144–151.
- Newton, I. 1989. Lifetime reproduction in birds. – Academic Press.
- O’Hara, R. B. and Sillanpää, M. J. 2009. A review of Bayesian variable selection methods: what, how and which. – *Bayesian Anal.* 4: 85–118.
- Orzack, S. H. et al. 2011. Static and dynamic expression of life history traits in the northern fulmar *Fulmarus glacialis*. – *Oikos* 120: 369–380.
- Papaix, J. et al. 2010. Combining capture-recapture data and pedigree information to assess heritability of demographic parameters in the wild. – *J. Evol. Biol.* 23: 2176–2384.
- Pennel, M. L. and Dunson, D. B. 2005. Bayesian semiparametric dynamic frailty models for multiple event time data. – *Biometrics* 62: 1044–1052.
- Plummer, M. 2008. Penalized loss functions for Bayesian model comparison. – *Biostatistics* 9: 523–539.
- Plummer, M. et al. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* 6: 7–11. <<http://CRAN.R-project.org/doc/Rnews/>>.
- Royle, J. A. 2008. Modeling individual effects in the Cormack–Jolly–Seber model: a state–space formulation. – *Biometrics* 64: 364–370.
- Service, P. M. et al. 1998. Experimental evolution of senescence: an analysis using a ‘heterogeneity’ mortality model. – *Evolution* 52: 1844–1850.
- Service, P. M. 2000a. Heterogeneity in individual mortality risk and its importance for evolutionary studies of senescence. – *Am. Nat.* 156: 1–13.
- Service, P. M. 2000b. Stress resistance, heterogeneity, and mortality plateaus: a comment on Drapeau et al. – *Exp. Gerontol.* 35: 1085–1087.
- Service, P. M. 2004. Demographic heterogeneity explains age-specific patterns of genetic variance in mortality rates. – *Exp. Gerontol.* 39: 25–30.
- Spiegelhalter, D. J. et al. 2002. Bayesian measures of model complexity and fit. – *J. R. Stat. Soc. B* 64: 583–639.
- Stauffer, H. B. 2008. Contemporary Bayesian and Frequentist research methods for natural resources scientists. – Wiley.
- Steiner, U. K. and Tuljapurkar, S. 2012. Neutral theory for life histories and individual variability in fitness components. – *Proc. Natl Acad. Sci. USA* 109: 4684–4689.
- Steiner, U. K. et al. 2010. Dynamic heterogeneity and life history variability in the kittiwake. – *J. Anim. Ecol.* 79: 436–444.
- Stover, J. P. et al. 2012. Demographic heterogeneity impacts density-dependent population dynamics. – *Theor. Ecol.* 5: 297–309.
- Tuljapurkar, S. et al. 2009. Dynamic heterogeneity in life histories. – *Ecol. Lett.* 11: 1–14.
- van de Pol, M. and Wright, J. 2009. A simple method for distinguishing within- versus between-subject effects using mixed models. – *Anim. Behav.* 77: 753–758.
- Vaupel, J. W. and Yashin, A. I. 1985. Heterogeneity’s ruses: some surprising effects of selection on population dynamics. – *Am. Stat.* 39: 176–185.
- Vaupel, J. W. et al. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. – *Demography* 16: 439–454.
- Wilberg, M. J. and Bence, J. R. 2008. Performance of deviance information criterion model selection in statistical catch-at-age analysis. – *Fish. Res.* 93: 212–221.
- Wilson, A. J. and Nussey, D. H. 2010. What is individual quality? An evolutionary perspective. – *Trends Ecol. Evol.* 25: 207–214.
- Wintrebert, C. M. A. et al. 2005. Joint modelling of breeding and survival in the kittiwake using frailty models. – *Ecol. Modell.* 181: 203–213.
- Yashin, A. I. et al. 2001. Hidden frailty: myths and reality. – *Dept of Statistics, Univ. of southern Denmark, (Odense Res. Rep. 34)*. <[www.demogr.mpg.de/en/projects\\_publications/publications\\_1904/other\\_working\\_papers/hidden\\_frailty\\_myths\\_and\\_reality\\_1993.htm](http://www.demogr.mpg.de/en/projects_publications/publications_1904/other_working_papers/hidden_frailty_myths_and_reality_1993.htm)> (accessed 8 May 2012).
- Yue, H. and Chan, K. S. 1997. A dynamic frailty model for multivariate survival data. – *Biometrics* 53: 785–793.
- 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 O20532 at <[www.oikosoffice.lu.se/appendix](http://www.oikosoffice.lu.se/appendix)>). Appendix A1. Estimation and model selection (Methods). Appendix A2. Inclusion variables (complete data set). Appendix A3. Sensitivity analyses (complete data set). Appendix A4. Relationship between age and vital rates: mean individual. Appendix A5. Posterior densities and autocorrelation (Model 9, Table 1). Appendix A6. Exclusion of location-years with massive predation on eggs. Appendix A7. BUGS codes (Model 1, Table 1).