Dead or gone? Bayesian inference on age-specific mortality for the dispersing sex

Julia A. Barthold with Fernando Colchero, Andrew J. Loveridge, David W. Macdonald, Craig Packer (in alphabetical order)

Email: julia.barthold@zoo.ox.ac.uk • Twitter: @jabarthold

Background

As apex predators, flagship species, and desired trophies, African lions (*Panthera leo*) are paramount for the conservation of their ecosystems. Yet global population size is plummeting and many small populations face extinction. Demographic models can inform population management measures. Aiming to construct one that included both males and females, I realised that age-specific mortality of lion males is not known. We therefore set:

The Goal

In short To estimate age-specific mortality for males and females for two lion populations.

In detail To compare mortality between the sexes and between an "undisturbed" population that is hardly impacted by humans and a "disturbed" one that is hunted for trophies in order to study how human impact alters age trajectories of mortality, and whether males respond differently to human impact than females.

The Problem

In short Mortality of the dispersing sex can usually not be estimated.

In detail Estimating mortality of at least one of the sexes is commonly hindered by incomplete data on dispersing individuals. In lions, males leave their natal area around the age of maturity. If they disperse out of study areas, they are usually lost for data collection. Therefore, if males disappear around the age of maturity, their fate remains uncertain; they may have died or left. These uncertain male records prevent the estimation of male mortality using existing methods.

The Solution

In short A Bayesian model to infer mortality for data that include uncertain records due to the possibility of dispersal.

In detail We developed a Bayesian model that imputes dispersal state (i.e., died or left) for uncertain records as a latent state variable jointly with the coefficients of a parametric mortality model in a hierarchical framework. It also imputes sex as a latent state for individuals who died before their sex could be determined. The model takes first and last seen ages as input. Different types of records by sex, imputed dispersal state, and migration history make different contributions to the likelihoods of the mortality and a dispersal model. We fitted the model using an MCMC algorithm to data from Serengeti (TZ) and Hwange (ZW) lions.

More detail Download full methods from bit.ly/1yOgSeE or scan

The Finding

In short A simulation study validated our model (Fig. 1). Surprisingly, the disturbed lion population had only slightly higher adult mortality than the undisturbed one. It also had much lower cub mortality (Fig. 2).

In detail

Model validation

The model reliably recovered the mortality rates for simulated data of different sample sizes, with and without records of unknown sex, and of varying proportions of observed deaths (Fig. 1).

Fig. 2 Mortality rates and life expectancies of both sexes for an undisturbed (Serengeti) and a disturbed (Hwange) lion population. Polygons represent 95% credible intervals with white lines indicating mean values. Sample sizes varied from ~4100 in Serengeti to ~500 in Hwange
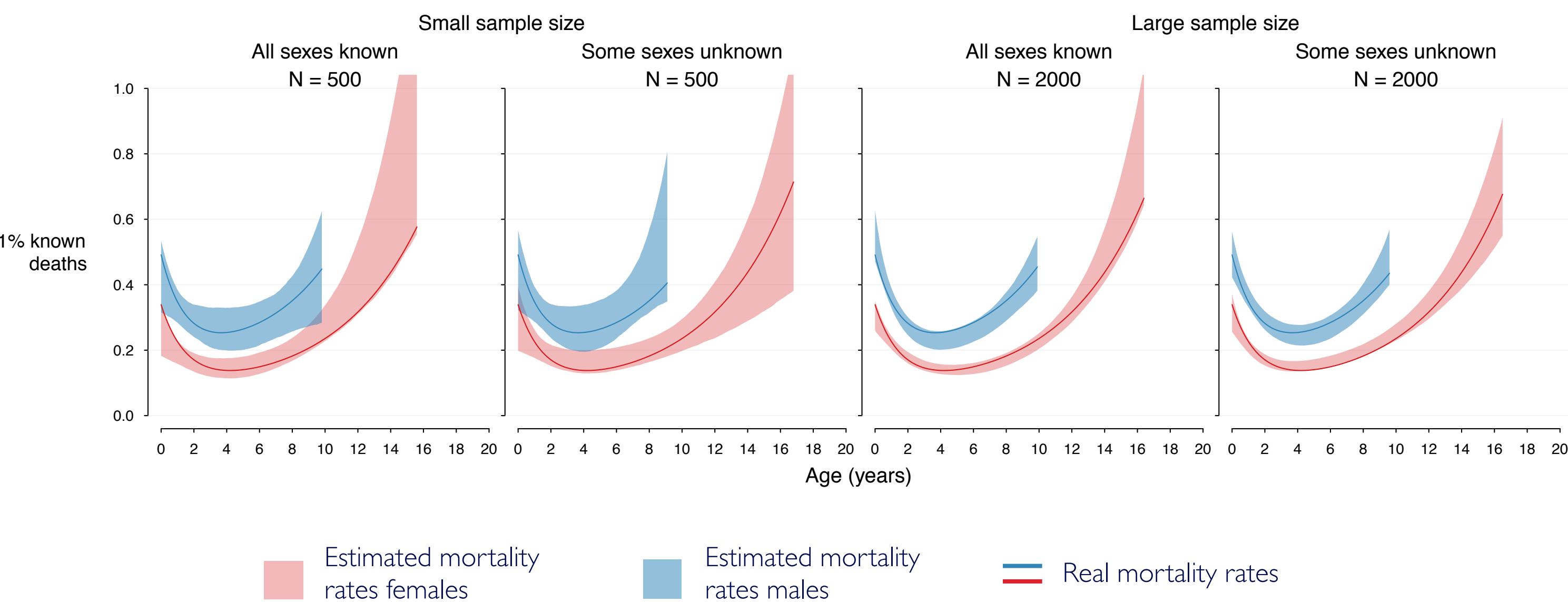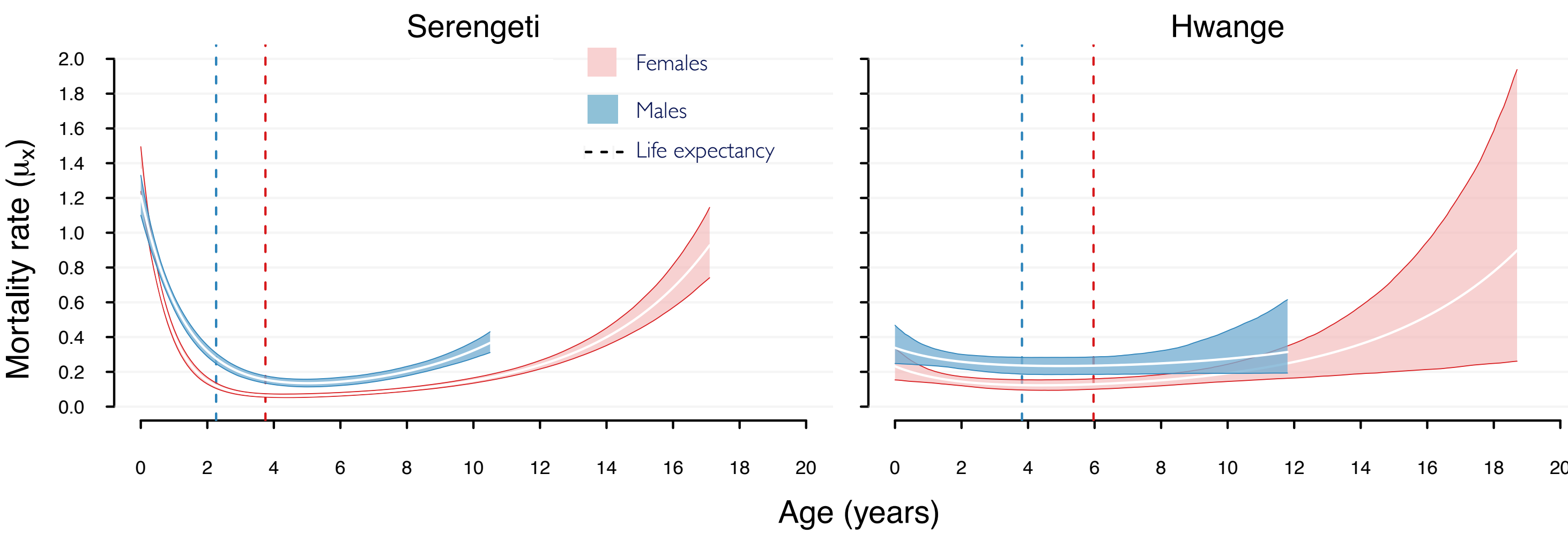
Fig. 1 Results of the simulation study. Estimated mortality rates for males and females vs. real mortality rates (used to simulate the data). Polygons represent 95% credible intervals of age-specific mortality rates. Results are given for four simulations when 1% of deaths were observed. We varied the size of the population and whether or not the sex of 30% of individuals dying younger than 1 year of age remained undetermined.

Application

For the undisturbed population from Serengeti National Park, mortality of both sexes was bathtub-shaped with high initial cub mortality, low mortality of prime-aged adults, and an age-dependent increase in mortality during older ages (Fig. 2). For the disturbed population from Hwange National Park, mortality varied less over the life course (Fig. 2). Compared to Serengeti lions, Hwange lions had lower cub mortality and slightly higher prime-aged adult mortality, resulting in higher life expectancies for both sexes for Hwange lions when compared to Serengeti lions.

In both populations, males had higher mortality rates than females across most ages, resulting in lower life expectancies for males than for females (Fig. 2).

A decomposition into age-dependent and age-independent sources of mortality showed that Hwange lions had higher age-independent mortality than Serengeti lions (Fig. 3), with Hwange males showing the highest values (Fig. 3). This likely reflects anthropogenic mortality and in particular trophy hunting in the Hwange population.
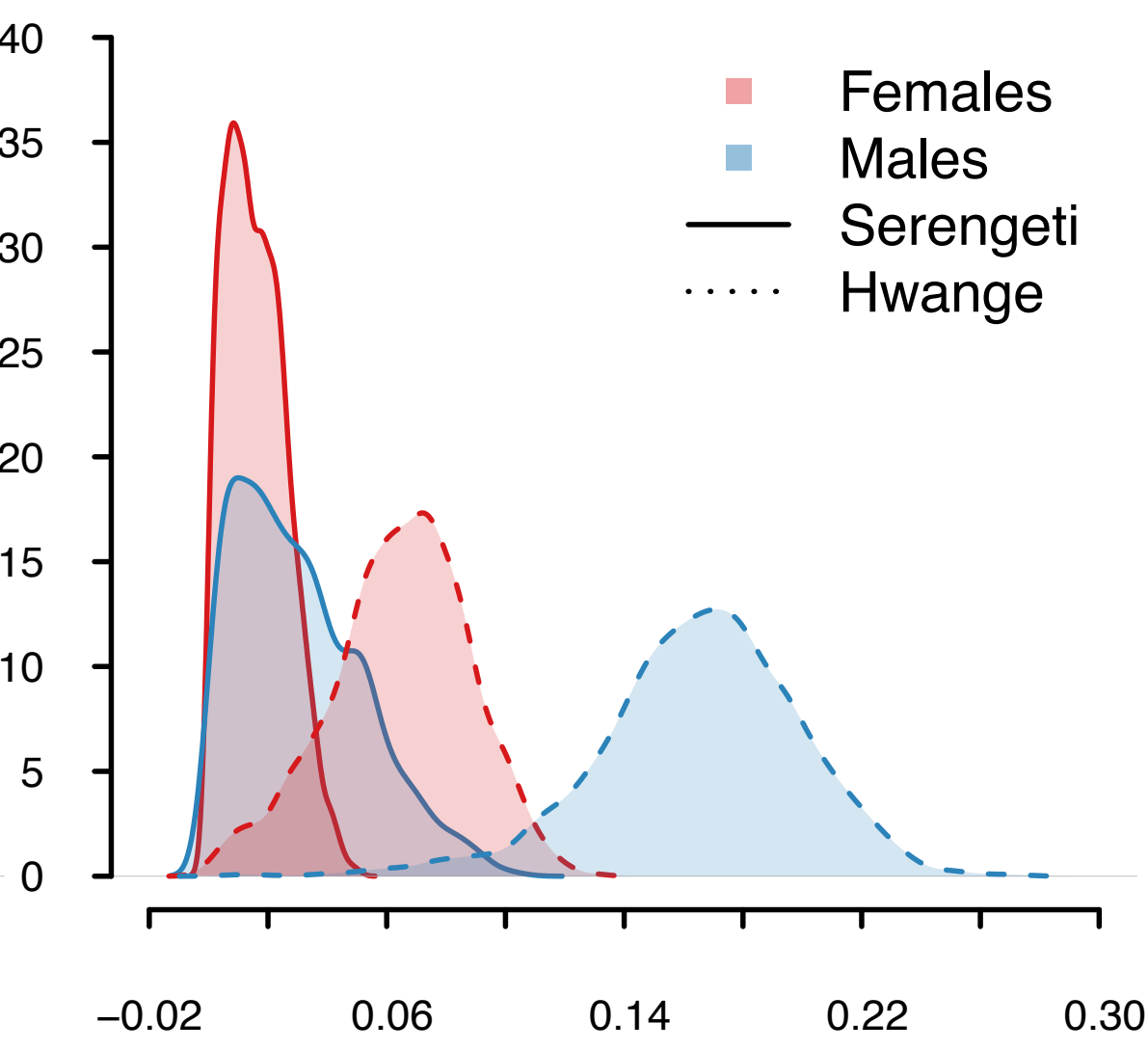
Fig. 3 Posterior density distributions of the age-independent mortality hazard (*c* parameter of the Siler Model).

The Significance

In short Our model provides the possibility to estimate age-specific mortality for the dispersing sex for many species. Our estimates for male lions can be used in future models of lion population dynamics.

In detail Our model estimates sex-specific mortality for species with natal dispersal, but can be adapted to account for dispersal of both sexes or higher order dispersal. Our finding that life expectancies were higher in the disturbed population than in the undisturbed one may be due to the disturbed population being kept at low densities through human impact, resulting in low cub mortality. This calls for future investigations whether density effects on cub mortality need to be incorporated into models that estimate sustainable male off-take quota.

Died or left? Bayesian inference on male and female survival for data with uncertain male fates due to male natal dispersal

Julia A. Barthold[1,2,3], Craig Packer [4], Fernando Colchero[3,5]

[1]Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK
[2]Max Planck Institute for Demographic Research, 18057 Rostock, Germany
[3]Max Planck Odense Center on the Biodemography of Aging, University of Southern Denmark, 5230 Odense M, Denmark
[4]Department of Ecology, Evolution and Behavior, University of Minnesota, 1987 Upper Buford Circle, Saint Paul, Minnesota 55108, USA
[5]Department of Mathematics and Computer Science, University of Southern Denmark, 5230 Odense M, Denmark

Corresponding author: Julia A. Barthold
**Email:** julia.barthold@zoo.ox.ac.uk

Methods

We focus on species in which males disperse only once at around the age of maturity ("natal dispersal"). To isolate the effect of uncertain male records on estimates of survival from other effects, we focus on data that are mostly free of other issues. We therefore assume that individuals are re-sighted with certainty if they are alive and in the study area, that individuals are only observed at one location, and that survival probabilities only depend on age and sex and not on any other covariates. We further assume that survival probabilities in- and outside of the study area are equal and that individuals born outside of the study area disperse into the study area with equal probabilities as individuals born in the study area disperse out of it. We also assume that ages of individuals whose birth was not observed (left-truncated records) can be estimated with sufficient certainty by a trained observer to allow us to not include time of birth as a latent state in the model and to model ages at death as a continuous variable. However, since the data available to us for the empirical application contained individuals that died before sexing was possible, we did construct the model to accommodate this type of record, treating the sex of unsexed individuals as another latent state.

Data structure

The life history data used to estimate age- and sex-specific mortality included records for native-borns and immigrants. Native-borns were born in the study population, defined as all individually recognisable and constantly monitored individuals. Immigrants entered the study population some time after their birth either due to migration, or being alive at the start of the study (Figure 1.1). The recorded types of departure from the population included death, censoring due to being alive at the end of the study, or uncertain fate (death or censoring through dispersal). Uncertain fates through dispersal were only caused by dispersals from the study population to an external population, and not by dispersals within the study population. Here, we refer to this out-migration from the study population when we use the term "dispersal".

Model variables and functions

We used a Bayesian approach that allowed us to estimate not only parameters for mortality and dispersal but also the different latent states such as dispersal state and sex. We defined the random variables X for ages at death and Y for ages at natal dispersal. We also defined
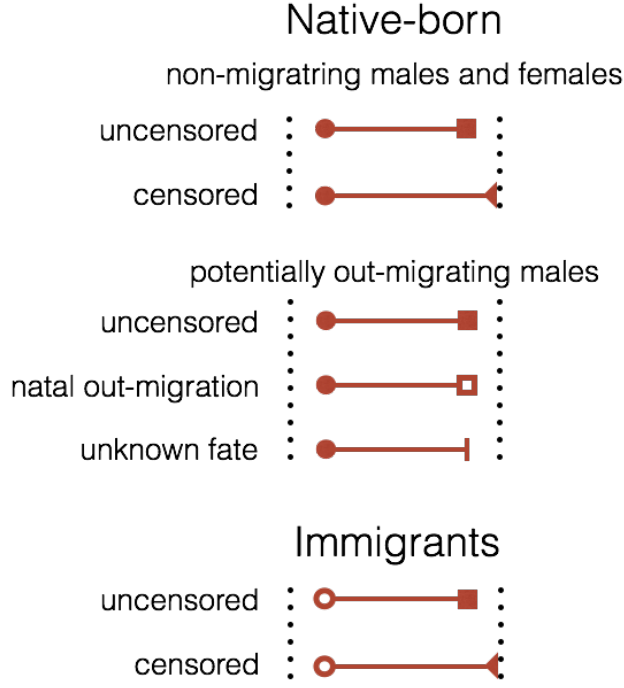
Figure 1.1. Example of types of records in the lion datasets. Circles represent times of entry, where the entry type for filled circles corresponds to known times of birth and open circles are entries after birth (i.e. immigration or birth before the study started). Squares are departure times where filled squares are known times of death and open squares are dispersal. Filled triangles indicate individuals known to be alive at the end of the study and vertical bars indicate that the type of departure from the study population is uncertain (i.e. either death or dispersal).

a dispersal state D which assigned 1 if an individual i, born at b_i and last detected at t_i^L, dispersed in its last detection age, $x_i^L = t_i^L - b_i$, and 0 if otherwise. We treated D as a latent variable for all individuals with uncertain fate. We also defined a variable S that assigned 1 if an individual was female and 0 otherwise.

We used a parametric model to infer age-specific mortality. This is expressed for continuous age x as the mortality rate, or hazard of death,

$$\mu(x|\boldsymbol{\theta}) = \lim_{\Delta x \to 0} \frac{\Pr(x \le X < x + \Delta x \mid x \le X, \boldsymbol{\theta})}{\Delta x}, \tag{1.1}$$

where $\boldsymbol{\theta}$ is a vector of mortality parameters to be estimated. The estimated mortality rate can be used to calculate the probability to survive from birth to age x, or survivor function,

$$S(x|\boldsymbol{\theta}) = Pr(X \ge x) = \exp\left[-\int_0^x \mu(z|\boldsymbol{\theta})\mathrm{d}z\right], \tag{1.2a}$$

the probability that death occurs before age x, or the cumulative density function (CDF),

$$F(x|\boldsymbol{\theta}) = Pr(X < x) = 1 - S(x|\boldsymbol{\theta}), \tag{1.2b}$$

and the probability density function (PDF) for age at death

$$f(x|\boldsymbol{\theta}) = \frac{d}{dx}F(x|\boldsymbol{\theta}) = S(x|\boldsymbol{\theta})\mu(x|\boldsymbol{\theta}). \tag{1.2c}$$

To capture the bathtub-shaped mortality rate typical of large mammals, we used the Siler

model (Siler 1979) in the form

$$\mu(x|\boldsymbol{\theta}) = e^{a_0 - a_1 x} + c + e^{b_0 + b_1 x}, \tag{1.3}$$

where $\boldsymbol{\theta} = \{a_0, a_1, c, b_0, b_1\}$, with $a_0, b_0 \in \mathbb{R}$ and $a_1, c, b_1 > 0$. The Siler model is a competing risk model constituted of three additive mortality hazards. For modelling the ages at dispersal, we used a gamma distribution and defined the PDF as $g(y|\boldsymbol{\gamma})$ with CDF $G(y|\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is a vector of parameters to be estimated. A summary of all the functions, parameters, indicators and variables is provided in Table 1.1.

Table 1.1. Description of random variables, observed variables and indicators.

Modeled random variables	
X	Random variable for ages at death, where x is any age element
Y	Random variable for ages at natal dispersal with elements y
D	Binary random variable for disperser or non-disperser
S	Binary random variable for sex
Observed variables and indicators	
$\mathbf{x}^F$	Vector of ages at first detection
$\mathbf{x}^L$	Vector of ages at last detection
$\mathbf{m}$	Indicator vector for immigrants ($m_i = 1$ if immigrant)
Updated indicators	
$\mathbf{d}$	Indicator vector for dispersers ($d_i = 1$ if disperser and $d_i = 0$ otherwise)
$\mathbf{s}$	Indicator vector for sex ($s_i = 1$ if female and $s_i = 0$ otherwise)
$\boldsymbol{\omega}$	Indicator for potential dispersers
Parameters	
$\boldsymbol{\theta}$	Vector of mortality parameters
$\boldsymbol{\gamma}$	Vector of natal dispersal parameters
Functions	
Mortality	
$\mu(x \mid \boldsymbol{\theta})$	Mortality (Siler model)
$S(x \mid \boldsymbol{\theta})$	Survival
$F(x \mid \boldsymbol{\theta})$	CDF for age at death ($F(x) = 1 - S(x)$)
$f(x \mid \boldsymbol{\theta})$	PDF for age at death
Dispersal	
$g(x \mid \boldsymbol{\gamma})$	PDF for age at natal dispersal (gamma)
$G(x \mid \boldsymbol{\gamma})$	CDF for age at natal dispersal

Likelihood and posterior

To construct the likelihood for our Bayesian model, we assigned a different probability to each type of record in Figure 1.1. The likelihood for females and non-migrating native-born males was

$$p(D, x^F, x^L; \boldsymbol{\theta}) = \begin{cases} \Pr(X = x^L \mid X > x^F) & \text{if uncensored} \\ \Pr(X > x^L \mid X > x^F) & \text{if censored,} \end{cases} \tag{1.4a}$$

where x^L corresponds to the age at last detection and x^F is the age at first detection (i.e., $x^F = 0$ for individuals born in the study area and $x^F > 0$ for immigrants or individuals born

before the study started), while for dispersers the likelihood was constructed as

$$p(D, x^F, x^L; \boldsymbol{\theta}) = \begin{cases} \Pr(X = x^L, Y > x^L \mid X > x^F) & \text{if uncensored} \\ \Pr(X > x^L, Y > x^L \mid X > x^F) & \text{if censored} \\ \Pr(X > x^L, Y = x^L \mid X > x^F) & \text{if dispersed.} \end{cases} \quad (1.4b)$$

As we will show below, the dispersal state D for individuals with uncertain fate was treated as a latent variable that needed to be estimated. The censored and uncensored probabilities for dispersers were used to determine how likely a potential dispersers (i.e., departure type of uncertain fate and age older than minimum age at dispersal) out-migrated (e.g., last expression in equation 1.4b) or died at the age of last detection x_L (e.g., first expression in equation 1.4b). With this, we could construct the full Bayesian model as

$$p(\mathbf{d}_u, \mathbf{s}_u \boldsymbol{\theta}, \boldsymbol{\gamma} \mid \mathbf{d}_k, \mathbf{s}_k, \mathbf{x}^F, \mathbf{x}^L) \quad \propto \quad \underbrace{p(\mathbf{d}, \mathbf{s}, \mathbf{x}^F, \mathbf{x}^L \mid \boldsymbol{\theta}, \boldsymbol{\gamma})}_{\text{likelihood}}$$
$$\times \quad \underbrace{p(\mathbf{d})p(\mathbf{s})}_{\text{priors for states}}$$
$$\times \quad \underbrace{p(\boldsymbol{\theta})p(\boldsymbol{\gamma})}_{\text{priors for parameters}}, \quad (1.5)$$

where $\mathbf{d}$ was the vector of dispersal states (i.e., $d_i = 1$ if individual i dispersed and $d_i = 0$ otherwise) and $\mathbf{s}$ was the indicator vector for sex ($s_i = 1$ if female and $s_i = 0$ if male). Each of these vectors had two subsets represented by the subscripts u for unknown and k for known.

MCMC and conditional posteriors

We used a Markov Chain Monte Carlo (MCMC) algorithm to fit the model in equation 1.5. For all implementations, we ran four parallel MCMC sequences with different randomly drawn starting values and set the number of iterations to 15,000 steps with a burn-in of 5,000 initial steps and a thinning factor of 20. We used a hierarchical framework that only needed the conditionals for posterior simulation by Gibbs sampling, specifically Metropolis-within-Gibbs (Gelfand & Smith 1990; Clark 2007). This means that, for this particular case, the algorithm divided the posterior for the joint distribution of unknowns into four sections: (a) estimation of mortality parameters, (b) estimation of dispersal parameters, (c) estimation of unknown dispersal state, and (d) estimation of unknown sexes. Here we present each section, specifying the conditional posterior and the acceptance probability for the Gibbs Sampler algorithm.

Section a: Posterior for mortality parameters

The conditional posterior to estimate the mortality parameters $\boldsymbol{\theta}$ required only the ages at last detection x_i^L and the dispersal states d_i. The posterior for a given individual i was

$$p(\boldsymbol{\theta} \mid x_i^L, d_i) \propto \begin{cases} \dfrac{f(x_i^L \mid \boldsymbol{\theta})}{S(x_i^F \mid \boldsymbol{\theta})} \; p(\boldsymbol{\theta}|\boldsymbol{\theta}_p) & \text{if } d_i = 0 \\[4mm] \dfrac{S(x_i^L \mid \boldsymbol{\theta})}{S(x_i^F \mid \boldsymbol{\theta})} \; p(\boldsymbol{\theta}|\boldsymbol{\theta}_p) & \text{if } d_i = 1 \,, \end{cases} \quad (1.6)$$

where $\boldsymbol{\theta}_p$ was a vector of prior hyper-parameters and x_i^F the age at first detection. If the individual was a native-born, then $x_i^F = 0$ and the denominator in both expressions was equal to 1. At every iteration and for a given parameter $\theta \in \boldsymbol{\theta}$ with conditional posterior $p(\theta \mid \ldots)$, the algorithm proposes a new parameter value for each element of θ' and accepts it with acceptance

probability

$$p(\theta, \theta') = \min \left\{ 1, \frac{\prod\limits_{i=1}^{n} p(\theta' \mid \ldots)}{\prod\limits_{i=1}^{n} p(\theta \mid \ldots)} \right\}. \tag{1.7}$$

Section b: Posterior for dispersal parameters

The conditional posterior to estimate the parameters $\boldsymbol{\gamma}$ for the distribution of ages at first dispersal for a given individual i was

$$p(\boldsymbol{\gamma} \mid x_i^F - \alpha, x_i^L - \alpha, d_i, \omega_i, m_i) \propto \begin{cases} g(x_i^L - \alpha \mid \boldsymbol{\gamma}) \; p(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_p) & \text{if } \omega_i = 1, \; m_i = 0 \; \& \; d_i = 1 \\ [1 - G(x_i^L - \alpha \mid \boldsymbol{\gamma})]; \; p(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_p) & \text{if } \omega_i = 1, \; m_i = 0 \; \& \; d_i = 0 \\ \dfrac{g(x_i^F - \alpha \mid \boldsymbol{\gamma})}{S(x_i^F \mid \boldsymbol{\theta})} \; p(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_p) & \text{if } m_i = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{1.8}$$

where $\boldsymbol{\gamma}_p$ was a vector of prior hyper-parameters for $\boldsymbol{\gamma}$, ω_i was an indicator that assigns 1 if an individual was a potential disperser (i.e., it belonged to the dispersing sex and disappeared at an age older than the minimum age at dispersal α), and m_i was an indicator for immigrants. We set the minimum age at dispersal to $\alpha = 1.75$ years for the simulated data and $\alpha = 1.5$ for the Serengeti data. The age α corresponded to the earliest age at which immigrants could be detected and potential dispersers could be last seen. For a parameter $\gamma \in \boldsymbol{\gamma}$ with conditional posterior density $p(\gamma \mid \ldots)$ The acceptance probability for a proposed parameter of γ' was

$$p(\gamma', \gamma) = \min \left\{ 1, \frac{\prod\limits_{i=1}^{n} p(\gamma' \mid \ldots)}{\prod\limits_{i=1}^{n} p(\gamma \mid \ldots)} \right\}. \tag{1.9}$$

Section c: Posterior for dispersal states

Dispersal state was evaluated on individuals that were potential dispersers (i.e. $\omega_i = 1$) and estimated the joint probabilities

$$p(d_i \mid x_i^L, d_i, \omega_i, m_i) \propto \begin{cases} f(x^L)(1 - G(x^L)) \; p(d_i \mid \boldsymbol{\theta}_p, \boldsymbol{\gamma}_p) & \text{if } \omega_i = 1, m_i = 0, d_i = 0 \\ S(x^L)g(x^L) \; p(d_i \mid \boldsymbol{\theta}_p, \boldsymbol{\gamma}_p) & \text{if } \omega_i = 1, m_i = 0, d_i = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1.10}$$

The first terms on the right hand side of equation 1.10 correspond to the likelihood function as defined in equations 1.4, while the second terms are the priors for dispersal state. For this section the acceptance probability for the Metropolis Sampling given the last seen ages, the dispersal states, the potential disperser states, and the immigration states was

$$p(d_i, d_i') = min \left\{ 1, \frac{\prod\limits_{i=1}^{n} p(d_i' \mid x_i^L, d_i, \omega_i, m_i)}{\prod\limits_{i=1}^{n} p(d_i \mid x_i^L, d_i, \omega_i, m_i)} \right\}. \tag{1.11}$$

Section d: Posterior for unknown sexes

Some individuals disappeared before the minimum age at dispersal without that their sex had been determined. The conditional posterior for the latent state of sex was

$$p(s_i \mid x_i^L, \boldsymbol{\theta}) \propto p(x_i^L, \boldsymbol{\theta} \mid s_i) \ p(s_i), \tag{1.12}$$

where the second term on the right-hand side is a prior for sex based on sex ratio at birth, or if the analysis was condition to survive to age x, then based on the sex ratio at age x (0.5 for the simulated data and 0.51 for the Serengeti data, respectively). The indicator for potential dispersers ω_i (see section c) was updated in each iteration. Individuals of undetermined sex and last seen ages older than the minimum age at dispersal were assigned 1 if imputed to be male and 0 if imputed to be female. The acceptance probability given the last seen ages and the mortality parameters was

$$p(s_i, s_i') = min \left\{ 1, \frac{\prod\limits_{i=1}^{n} p(s_i' \mid x_i^L, \boldsymbol{\theta})}{\prod\limits_{i=1}^{n} p(s_i \mid x_i^L, \boldsymbol{\theta})} \right\}. \tag{1.13}$$

Mortality and dispersal priors

We set the Siler parameters for the prior for females to $a_{0p} = -1.4$ ($\sigma = 0.5$), $a_{1p} = 0.65$ ($\sigma = 0.25$), $c_p = 0.07$ ($\sigma = 0.25$), $b_{0p} = -3.8$ ($\sigma = 0.5$), and $b_{1p} = 0.2$ ($\sigma = 0.25$), and for males to $a_{0p} = -1.2$ ($\sigma(a_{0p}) = 0.5$), $a_{1p} = 0.7$ ($\sigma(a_{1p}) = 0.25$), $c_p = 0.16$ ($\sigma(c_p) = 0.25$), $b_{0p} = -3.5$ ($\sigma(b_{0p}) = 0.5$), and $b_{1p} = 0.23$ ($\sigma(b_{1p}) = 0.25$). For dispersal, the Gamma parameters (shape and scale) for the prior were set to $\boldsymbol{\gamma}_p = \{8, 2\}$ with $\sigma(\boldsymbol{\gamma}_p) = \{2, 1\}$. Both the mortality and dispersal priors were uninformative.

Bibliography

Clark, J.S. (2007) *Models for ecological data.* Princeton University Press, Princeton, New Jersey.

Gelfand, A.E. & Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Siler, W. (1979) A competing-risk model for animal mortality. *Ecology*, **60**, 750–757.